

Eastern
Economy
Edition

FOURTH EDITION

INTRODUCTION TO
MEASUREMENTS AND
INSTRUMENTATION

ARUN K. GHOSH



Introduction to Measurements and Instrumentation

FOURTH EDITION

ARUN K GHOSH

Visiting Professor

Sir J.C. Bose School of Engineering, Hooghly

Formerly

Head, Instrumentation Centre, University of Kalyani

Principal, Murshidabad College of Engineering and Technology, Berhampore

Principal, Bengal College of Engineering and Technology, Durgapur

PHI Learning Private Limited

New Delhi-110001

2012

INTRODUCTION TO MEASUREMENTS AND INSTRUMENTATION, Fourth Edition

Arun K Ghosh

© 2012 by PHI Learning Private Limited, New Delhi. All rights reserved. No part of this book may be reproduced in any form, by mimeograph or any other means, without permission in writing from the publisher.

ISBN-978-81-203-4625-3

The export rights of this book are vested solely with the publisher.

Ninth Printing (Fourth Edition)

...

...

October, 2012

Published by Asoke K. Ghosh, PHI Learning Private Limited, M-97, Connaught Circus, New Delhi-110001 and Printed by Mohan Makhijani at Rekha Printers Private Limited, New Delhi-110020.

To
the memory of my elder brother
AMIYA

Contents

Foreword ix

Preface xi

Preface to the First Edition xiii

List of Abbreviations xv

| | |
|--|---------------|
| 1. INTRODUCTION | 1–3 |
| 1.1 Measurements | 1 |
| 1.2 Instruments | 2 |
| 2. STATIC CHARACTERISTICS OF INSTRUMENTS | 4–28 |
| 2.1 Desirable Characteristics | 4 |
| 2.2 Undesirable Characteristics | 9 |
| <i>Review Questions</i> | 24 |
| 3. ESTIMATION OF STATIC ERRORS AND RELIABILITY | 29–79 |
| 3.1 Definition of Parameters | 29 |
| 3.2 Limiting Error | 31 |
| 3.3 Statistical Treatment | 34 |
| 3.4 Error Estimates from the Normal (or Gaussian) Distribution | 41 |
| 3.5 Chi-Square Test | 49 |
| 3.6 Curve Fitting Methods | 51 |
| 3.7 Reliability Principles | 65 |
| <i>Review Questions</i> | 75 |
| 4. DYNAMIC CHARACTERISTICS OF INSTRUMENTS | 80–112 |
| 4.1 Transfer Function | 80 |
| 4.2 Standard Inputs to Study Time Domain Response | 82 |
| 4.3 Dynamic Characteristics | 84 |
| 4.4 Zero Order Instrument | 85 |

| | | | |
|-----------|---|-----|----------------|
| 4.5 | First Order Instrument | 86 | |
| 4.6 | Second Order Instrument | 96 | |
| | <i>Review Questions</i> | 108 | |
| 5. | TRANSDUCERS | | 113–169 |
| 5.1 | Classification of Transducers | 113 | |
| 5.2 | A Few Phenomena | 116 | |
| 5.3 | Selection of Transducers | 165 | |
| 5.4 | Smart Sensors and IEEE 1451 Standard | 166 | |
| | <i>Review Questions</i> | 168 | |
| 6. | DISPLACEMENT MEASUREMENT | | 170–237 |
| 6.1 | Pneumatic Transducers | 170 | |
| 6.2 | Electrical Transducers | 173 | |
| 6.3 | Optical Transducers | 200 | |
| 6.4 | Ultrasonic Transducer | 211 | |
| 6.5 | Magnetostrictive Transducer | 215 | |
| 6.6 | Digital Displacement Transducers | 216 | |
| 6.7 | Proximity Sensors | 220 | |
| | <i>Review Questions</i> | 232 | |
| 7. | STRAIN MEASUREMENT | | 238–279 |
| 7.1 | Stress-Strain Relations | 238 | |
| 7.2 | Resistance Strain Gauges | 241 | |
| 7.3 | Fibre-optic Strain Gauges | 264 | |
| | <i>Review Questions</i> | 275 | |
| 8. | PRESSURE MEASUREMENT | | 280–329 |
| 8.1 | Definitions | 280 | |
| 8.2 | Pressure Units and Their Conversions | 282 | |
| 8.3 | Comparison with Known Dead-weights | 283 | |
| 8.4 | Force-summing Devices | 291 | |
| 8.5 | Secondary Transducers | 296 | |
| 8.6 | Vacuum Measurement | 306 | |
| 8.7 | Accessories | 318 | |
| | <i>Review Questions</i> | 324 | |
| 9. | ACCELERATION, FORCE AND TORQUE MEASUREMENT | | 330–371 |
| 9.1 | Acceleration Measurement | 330 | |
| 9.2 | Force Measurement | 339 | |
| 9.3 | Industrial Weighing Systems | 346 | |
| 9.4 | Torque Measurement | 350 | |
| 9.5 | Tachometers | 359 | |
| | <i>Review Questions</i> | 369 | |

| | |
|--|----------------|
| 10. TEMPERATURE MEASUREMENT | 372–444 |
| 10.1 Temperature Scale | 372 |
| 10.2 Change in Dimensions | 376 |
| 10.3 Change in Electrical Properties | 380 |
| 10.4 Thermoelectricity | 399 |
| 10.5 Fibre-optic Sensors | 413 |
| 10.6 Quartz Thermometer | 419 |
| 10.7 Change in the Velocity of Sound Propagation | 420 |
| 10.8 Radiation Pyrometers | 421 |
| 10.9 Thermowells | 435 |
| <i>Review Questions</i> | 437 |
| 11. FLOW MEASUREMENT | 445–490 |
| 11.1 Reynolds Number and Flow Patterns | 446 |
| 11.2 Head-type Flowmeters | 447 |
| 11.3 Velocity Measurement-type Flowmeters | 459 |
| 11.4 Mass Flow Measurement Type Flowmeters | 473 |
| 11.5 Positive Displacement Flowmeters | 479 |
| 11.6 Open Channel Flowmeters | 482 |
| 11.7 Turndown and Rangeability of Flowmeters | 483 |
| <i>Review Questions</i> | 484 |
| 12. LEVEL MEASUREMENT | 491–521 |
| 12.1 Mechanical Level Indicators | 491 |
| 12.2 Optical Level Indicators | 503 |
| 12.3 Electrical Level Indicators | 505 |
| 12.4 Radiative, Other Than Optical, Methods | 510 |
| 12.5 Level Switches | 514 |
| <i>Review Questions</i> | 518 |
| 13. MISCELLANEOUS MEASUREMENTS | 522–602 |
| 13.1 Humidity and Moisture Measurement | 522 |
| 13.2 Density Measurement | 535 |
| 13.3 Conductivity Measurement | 544 |
| 13.4 Oxidation-Reduction Potential (ORP) | 551 |
| 13.5 pH Measurement | 562 |
| 13.6 Polarography | 571 |
| 13.7 Viscosity Measurement | 577 |
| 13.8 Consistency Measurement | 586 |
| 13.9 Turbidity Measurement | 588 |
| 13.10 Opacity Measurement | 595 |
| <i>Review Questions</i> | 600 |
| 14. ANALYTICAL INSTRUMENTATION | 603–724 |
| 14.1 Industrial Gas Analysis | 603 |
| 14.2 Chromatography | 616 |

| | | | |
|------------|---|-----|----------------|
| 14.3 | Mass Spectrometer | 632 | |
| 14.4 | Infrared Analyser | 653 | |
| 14.5 | Atomic Spectrometry | 666 | |
| 14.6 | UV-visible Absorption Spectrophotometer | 682 | |
| 14.7 | Nuclear Magnetic Resonance Spectroscopy | 684 | |
| 14.8 | Electron Spin Resonance Spectrometer | 690 | |
| 14.9 | X-ray Methods | 695 | |
| 14.10 | Radiation Detectors | 714 | |
| 14.11 | Sample Handling Systems | 715 | |
| | <i>Review Questions</i> | 719 | |
| 15. | HAZARDOUS AREAS AND INSTRUMENTATION | | 725–741 |
| 15.1 | Classification | 725 | |
| 15.2 | Explosion Protection of Electrical Apparatus | 728 | |
| 15.3 | Intrinsically Safe Instrumentation | 730 | |
| | <i>Review Questions</i> | 740 | |
| 16. | SIGNAL CONDITIONING | | 742–819 |
| 16.1 | Bridge Circuits | 742 | |
| 16.2 | Conditioning Processes | 754 | |
| 16.3 | Recovery of Signals | 793 | |
| 16.4 | Signal Conversion | 796 | |
| | <i>Review Questions</i> | 812 | |
| 17. | DISPLAY DEVICES AND RECORDING SYSTEMS | | 820–856 |
| 17.1 | Classification and Comparison | 820 | |
| 17.2 | Characteristics of Digital Display | 821 | |
| 17.3 | Digital Display Elements | 822 | |
| 17.4 | Recording | 835 | |
| 17.5 | Data Acquisition Systems | 847 | |
| 17.6 | Vitual Instrumentation | 851 | |
| | <i>Review Questions</i> | 854 | |
| | Appendix A Variance of Combinations | | 857 |
| | Appendix B Linear Time-invariant Systems | | 858–860 |
| | Appendix C Laplace Transform | | 861–864 |
| | Appendix D Statistical Tables | | 865–866 |
| | Appendix E Psychrometric Table | | 867 |
| | Appendix F Miscellaneous Data | | 868–873 |
| | Appendix G Solutions to Numerical Problems | | 874–907 |
| | Index | | 909–919 |

Foreword

Metrology is the science and technology of measurement. Since time immemorial, reliable measurement of various commodities and quantities has been important for trade and commerce as well as for agricultural and industrial activities. The present-day drive towards globalisation of the economy has made this to be a priority task both at national and international levels. Modern engineering practices require sufficiently precise and fast measurements. Science is breaking new ground in measuring the very tiny and the very big. Therefore, an introductory course on instrumentation principles, with an appreciation of the possible errors in the measurements, constitutes an important part of learning for both science and engineering students.

Although many voluminous treatises on this subject are available, Dr Ghosh's *Introduction to Instrumentation and Control* is a well-focussed textbook covering the physical principles rather than the engineering details, which can be taught in one semester of the undergraduate curriculum. The contents of the book cover most of the requirements of the students. Of course, each topic can become the subject of a detailed discussion. For example, the topic of signal conditioning is by itself a vast area of research work. Students specialising in various subjects will however find a common minimum amount of learning in this book.

Dr Ghosh's presentation is lucid and the style is not verbose. I am sure that the book will be welcomed by the student community and become a success in its area.

Prof ES Raja Gopal
Emeritus Scientist
Department of Physics
Indian Institute of Science, Bangalore
Formerly, Director
National Physical Laboratory
New Delhi

Preface

I am happy to present the Fourth Edition of the book within twelve years of its first publication.

In this edition, apart from minor additions and alterations here and there, I have included the following new topics:

Chapter 3: Linearisation and Spline interpolation

Chapter 5: Classification of transducers, Hall effect, Piezoresistivity, Surface acoustic waves, Optical effects. The Piezoelectricity section has been rewritten.

Chapter 6: Proximity sensors

Chapter 8: Hall effect and SAW transducers

Chapter 9: Proving ring, Prony brake, Industrial weighing systems, Tachometers

Chapter 10: ITS-90, SAW thermometer

Chapter 12: Glass gauge, Zero suppression and zero elevation, Level switches

Chapter 13: Rewritten the section on ISFET

Also, I have added a new chapter titled *Hazardous Areas and Instrumentation* (Chapter 15).

I shall be happier if the book in its present form is found more useful by them for whom it is written. I shall welcome any suggestion or comment on the book for its further improvement which, as it is said, is a continuous process.

Arun K Ghosh
arunkghosh_bect@yahoo.com

Preface to the First Edition

Although many wonderful treatises on instrumentation have been written, this textbook is designed to provide a comprehensive introduction to physical principles, rather than constructional details, that can be taught in one semester of our undergraduate engineering courses. The book is an outgrowth of my teaching experience satisfying these two criteria. I thank the University Grants Commission at the outset for sponsoring this endeavour.

The organisation of the book is as follows. In the introductory Chapter 1, I have stressed the fact that an instrument can be considered as a combination of a sensor/transducer, a signal conditioner, and a suitable display/recording device. After covering the background material concerning static and dynamic behaviours of instruments in Chapters 2 to 4, I have discussed transducers in general in Chapter 5. The text continues in Chapters 6 to 10 with a presentation of detailed treatment of transducers for the measurement of five representative non-electrical physical quantities, namely displacement, strain, pressure, temperature and flow. Some discussion on calibration of temperature and pressure transducers has been included in the relevant chapters.

Next comes the topic of conditioning signals. This topic is presented in three chapters. First, the bridge circuits are described in Chapter 11 since both dc and ac bridges are used very commonly in measurement systems. Secondly, many linear and nonlinear processes involved in signal conditioning are covered in Chapter 12. Thirdly, techniques for recovery of signals from noise as well as for conversion of signals to digital forms have been discussed in Chapter 13.

Display and recording devices constitute the last segment of an instrumentation system and act as the man-machine interface. While discussing these devices in Chapter 14, I have omitted many obsolete ones, for example, nixie tubes, and included only those which are in vogue.

Just as a study of static and dynamic characteristics of instruments is an integral component of instrumentation, I feel that some knowledge of process control and stability is equally indispensable to gain an insight into the subject. Therefore, the concluding Chapter 15 is devoted to this topic.

A few mathematical topics which the students should be aware of, but the inclusion of which in the body of the text makes it cumbersome, have been included in appendices. Also, an appendix on miscellaneous data, such as SI units, conversion of units and meaning of various prefixes, has been added for the sake of completeness.

I have excluded telemetry and a few very specialised instrumentation systems, such as nuclear and cryogenic, on the plea that even an elementary knowledge of these areas is hardly necessary in the Indian context. Nowadays, almost all the instrumentation systems have computer interfaces and microprocessor control. I have excluded them as well because they are subjects by themselves and one can do little justice to them by their cursory inclusion under display and recording devices.

I have included solved problems throughout the text with a view to bringing out the meaning of many discussions. Also, questions and problems culled from examination papers of some universities are presented at the end of chapters. Such an exercise was undertaken only with the objective of giving the student a feel of the ground realities about the type of questions set in various tests.

All my endeavours will be rewarded only if this book finds any use in the teaching and learning of instrumentation at the appropriate level. As a natural corollary to that, any criticism and suggestions for improvement will be highly appreciated.

I am indebted to Professor ES Raja Gopal, formerly Director, National Physical Laboratory, New Delhi (presently Emeritus Scientist, Department of Physics, Indian Institute of Science, Bangalore) for writing a Foreword for this book. I wish to thank my colleagues Sushanta Biswas and Sanjay Das for helping me in collecting question papers for this book. I am also grateful to my friends Professor S K Deb and Dr Ratnabali Banerjee who have been most helpful in making constructive suggestions. I am also very appreciative of the help of my wife Giti who offered constructive criticism and provided constant encouragement. Lastly, I thank my daughter Rumi who processed the entire LaTeX text for its conversion to MS-Word format.

Arun K Ghosh

List of Abbreviations

| | |
|---------|---|
| AAS | Atomic Absorption Spectrophotometry |
| ADC | Analogue to Digital Converter |
| AES | Auger Electron Spectroscopy |
| AFC | Automatic Frequency Control |
| AFS | Atomic Fluorescence Spectroscopy |
| aka | Also Known As |
| AM | Amplitude Modulation |
| ANSI | American National Standards Institute |
| APCI | Atmospheric Pressure Chemical Ionisation |
| ASTM | American Society for Testing and Materials |
| CCD | Charge Coupled Device |
| CCW | Counter ClockWise |
| CENELEC | Comité Européen de Normalisation Électrotechnique |
| CF-FAB | Continuous Flow Fast Atom Bombardment |
| CFC | ChloroFluoroCarbon |
| CGA | Colour Graphics Adapter |
| CGS | Centimetre Gramme Second |
| CI | Chemical Ionisation |
| CMA | Cylindrical Mirror Analyser |
| CMF | Coriolis Mass Flowmeter |
| CMRR | Common Mode Rejection Ratio |
| CPU | Central Processing Unit |
| CRT | Cathode Ray Tube |
| CTA | Constant Temperature Anemometer |
| CW | ClockWise/Continuous Wave |
| DAC | Digital to Analogue Converter |
| DAQ | Data AcQuisition system |

| | |
|------|--|
| DAS | Data Acquisition System |
| DCP | Direct Current Plasma |
| DCS | Distributed Control System |
| DME | Dropping Mercury Electrode |
| DSP | Digital Signal Processing |
| DTS | Distributed Temperature Sensing |
| DWG | Dead Weight Gauge |
| EC | Exclusion Chromatography |
| ECD | Electron Capture Detector |
| ECN | Effective Carbon Number |
| EDL | Electrodeless Discharge Lamp |
| EDS | Energy Dispersive Spectrometry |
| EDTA | Ethylene Diamine Tetra-acetic Acid |
| EDX | Energy Dispersive X-ray (analysis) |
| EFPI | Extrinsic Fabry-Pérot Interferometer |
| EGA | Enhanced Graphics Adapter |
| EI | Electron Impact (ionisation) |
| emf | ElectroMotive Force |
| EMI | ElectroMagnetic Interference |
| EMR | ElectroMagnetic Radiation |
| EMT | Electron Multiplier Tube |
| EPR | Electron Paramagnetic Resonance |
| ES | ElectroSpray |
| ESA | Electrostatic Sector Analyser |
| ESCA | Electron Spectroscopy for Chemical Analysis |
| ESD | ElectroStatic Discharge |
| ESR | Electron Spin Resonance |
| FAB | Fast Atom Bombardment |
| FBG | Fibre Bragg Grating |
| FES | Flame Emission Spectrophotometry |
| FET | Field Effect Transistor |
| FFT | Fast Fourier Transform |
| FID | Flame Ionisation Detector/Free Induction Decay |
| FM | Frequency Modulation |
| FMCW | Frequency Modulated Continuous Wave |
| FNU | Formazin Nephelometric Unit |
| FPMH | Failures Per Million Hours |
| FSD | Full Scale Deflection |
| FSOT | Fused Silica Open Tubular |

| | |
|--------|---|
| FT-NMR | Fourier Transform Nuclear Magnetic Resonance |
| FTIR | Fourier Transform InfraRed |
| FTU | Formazin Turbidity Unit |
| FWHM | Full Width at Half Maximum |
| GBC | Gas-Bonded phase Chromatography |
| GC | Gas Chromatograph |
| GCMS | Gas Chromatograph Mass Spectrometer |
| GLC | Gas-Liquid Chromatography |
| GM | Geiger-Muller |
| GPIB | General Purpose Instrumentation Bus |
| GSC | Gas-Solid Chromatography |
| HART | Highway Addressable Remote Transducer |
| HCFC | HydroChloroFluoroCarbon |
| HETP | Height Equivalent of Theoretical Plates |
| HFC | HydroFluoroCarbon |
| HPLC | High Performance (or Pressure) Liquid Chromatograph |
| HTG | Hydrostatic Tank Guaging |
| HVAC | Heating-Ventillation-AirConditioning |
| IC | Integrated Circuit |
| ICP | Inductively Coupled Plasma |
| ICR | Ion Cyclotron Resonance |
| IDT | InterDigital Transducer |
| IEC | International Electrotechnical Commission |
| IEC | Ion Exchange Chromatography |
| IEEE | Institute of Electrical and Electronics Engineers |
| IFPI | Intrinsic Fabry-Pérot Interferometer |
| IP | Ingress Protection |
| ISA | Instrumentation Systems and Automation (society) |
| ISE | Ion Selective Electrode |
| ISFET | Ion-Selective Field Effect Transistor |
| ISO | International Standards Organisation |
| ITS-90 | International Temperature Scale of 1990 |
| IUPAC | International Union of Pure and Applied Chemistry |
| JTU | Jackson Turbidity Unit |
| KE | Kinetic Energy |
| LBC | Liquid-Bonded phase Chromatography |
| LC | Liquid Chromatography |
| LCD | Liquid Crystal Display |
| LCMS | Liquid Chromatograph Mass Spectrometer |

| | |
|----------|--|
| LDR | Light Dependent Resistor |
| LED | Light Emitting Diode |
| LLC | Liquid-Liquid Chromatography |
| LPF | Low-Pass Filter |
| LSB | Least Significant Bit |
| LSC | Liquid-Solid Chromatography |
| LVDT | Linear Variable Differential Transformer |
| MALD | Matrix Assisted Laser Desorption |
| MC | Moisture Content |
| MCP | MicroChannel Plate |
| MDT | Mean Down Time |
| MEMS | Micro-ElectroMechanical Systems |
| MI | Mineral Insulated |
| MIE | Minimum Ignition Energy |
| MSB | Most Significant Bit |
| MSD | Mean Squared Deviation |
| MTBF | Mean Time Between Failures |
| MTTF | Mean Time To Failure |
| MTTR | Mean Time To Repair |
| NBF | Narrow Band Filter |
| NCAP | Network Capable Application Processor |
| NDIR | Non Dispersive InfraRed |
| NDT | Non-Destructive Testing |
| NEC | National Electrical Code, USA |
| NIR | Near InfraRed |
| NMR | Nuclear Magnetic Resonance |
| NPP | Normal Pulse Polarography |
| NRZ | Non-Return to Zero |
| NTC | Negative Temperature Coefficient |
| NTP | Normal Temperature and Pressure |
| NTU | Nephelometric Turbidity Unit |
| OD | Oven Dry |
| OES | Optical Emission Spectrophotometry |
| OP-FTIR | Open Path Fourier Transformed InfraRed |
| OP-TDLAS | Open Path Tunable Diode LASer |
| OP-UV | Open Path UltraViolet |
| OPC | Organic PhotoConductor |
| ORP | Oxidation Reduction Potential |
| OSI | Open System International |

| | |
|--------|---|
| OTDR | Optical Time Domain Reflectometry |
| OTFT | Organic Thin Film Transistor |
| PC | Personal Computer |
| PCMCIA | Personal Computer Memory Card International Association |
| PD | Plasma Desorption |
| PDP | Plasma Display Panel |
| PGIA | Programmable Gain Instrumentation Amplifier |
| PID | Proportional plus Integral plus Derivative |
| PIGA | Pendulous Integrating Gyro Accelerometer |
| PLC | Programmable Logic Controller |
| PLL | Phase-Locked Loop |
| PMF | Polarisation Measuring Fibre |
| PMMC | Permanent Magnet Moving Coil |
| PMT | PhotoMultiplier Tube |
| ppm | Parts Per Million |
| ppt | Parts Per Thousand |
| PRT | Platinum Resistance Thermometer |
| PSD | Position Sensitive Device/Phase Sensitive Detector |
| PSK | Phase Shift Keying |
| PTC | Positive Temperature Coefficient |
| PVDF | PolyVinylidene Fluoride |
| PZT | lead (Plumbum) Zirconium Titanate |
| QIT | Quadrupole Ion Trap |
| QPS | Quadrature Phase Shifted |
| RB | Return to Bias |
| RF | Radio-Frequency |
| RH | Relative Humidity |
| RI | Radio-frequency Interference |
| RIXS | Resonant Inelastic X-ray Scattering |
| rms | Root Mean Square |
| RSF | Relative Sensitivity Factor |
| rss | Root-Sum Square |
| RTD | Resistance Temperature Detector |
| RVDT | Rotary Variable Differential Transformer |
| RZ | Return to Zero |
| S/H | Sample and Hold |
| SAMA | Scientific Apparatus Makers' Association |
| SAW | Surface Acoustic Wave |
| SAXS | Small Angle X-ray Scattering |

| | |
|-------|--|
| SCADA | Supervisory Control And Data Acquisition |
| SCOT | Support Coated Open Tubular |
| SG | Specific Gravity |
| SI | Standard International |
| SONAR | SOund Navigation And Ranging |
| SPDT | Single Pole Double Throw |
| SPRT | Standard Platinum Resistance Thermometer |
| SQUID | Superconducting QUantum Interference Device |
| SVGA | Super Video Graphics Array |
| SWG | Standard Wire Gauge |
| SXGA | Super eXtended Graphics Array |
| TAPPI | Technical Association of the Pulp and Paper Industry |
| TCD | Thermal Conductivity Detector |
| TCR | Temperature Coefficient of Resistance |
| TDLAS | Tunable Diode LAser Spectroscopy |
| TDM | Time Division Multiplexing |
| TDR | Time Domain Reflectometry |
| TDS | Total Dissolved Solids |
| TEDS | Transducer Electronic Data Sheet |
| TEOM | Tapered Element Oscillating Microbalance |
| TFT | Thin Film Transistor |
| TIB | Transformer Isolated Barrier |
| TIM | Transmitter Interface Module |
| TOF | Time Of Flight |
| TTL | Transistor Transistor Logic |
| UHV | UltraHigh Vacuum |
| USEPA | United States Environment Protection Agency |
| UXGA | Ultra eXtended Graphics Array |
| VCO | Voltage Controlled Oscillator |
| VGA | Video Graphics Array |
| VSMOW | Vienna Standard Mean Ocean Water |
| WCOT | Wall Coated Open Tubular |
| WDS | Wavelength Dispersive Spectrometry |
| WDX | Wavelength Dispersive X-ray (analysis) |
| XDCR | Transducer |
| XFS | X-ray Fluorescence Spectroscopy |
| XGA | eXtended Graphics Array |
| XPS | X-ray Photoelectric Spectroscopy |
| XRD | X-Ray Diffractometry |
| XRF | X-Ray Fluorescence |
| YAG | Yttrium Arsenic Garnet |

Introduction

1.1 Measurements

Measurements are made or measurement systems are set up for one or more of the following functions:

1. To *monitor* processes and operations
2. To *control* processes and operations
3. To carry-out some *analysis*.

The functions are elaborated below.

Monitoring

Thermometers, barometers, anemometers, water, gas, electricity meters only indicate certain quantities. Their readings do not perform any control functions in the ordinary sense. These measurements are made for monitoring purposes only.

Control

The thermostat in a refrigerator or geyser determines the temperature of the relevant environment and accordingly switches OFF or ON the cooling or heating mechanism to keep the temperature constant, i.e. to control the temperature. A single system sometimes may require many controls. For example, an aircraft needs controls from altimeters, gyroscopes, angle-of-attack sensors, thermocouples, accelerometer, etc.

Controlling a variable is rather an involved process and as such it is a subject of study by itself.

Analysis

Measurements are also made to

1. test the validity of predictions from theories,
2. build empirical models, i.e. relationships between parameters and quantities associated with a problem, and
3. characterise materials, devices and components.

In general, these requirements may be called *analysis*.

1.2 Instruments

Measurements are made with the help of instruments. Instruments, in general, consist of a few elements. But before we go into the contents of a generalised instrument, let us define what we mean by an instrument.

An instrument can be defined as a device or a system which is designed in such a way that it maintains a functional relationship between a prescribed property of a substance and a physical variable, and communicates this relationship to a human observer by some ways and means. For example, a mercury-in-glass thermometer is an instrument, because it maintains a linear relationship between thermal expansion of mercury (prescribed property) and temperature (physical variable) and communicates this relationship to us through a graduated scale.

A generalised instrument can be schematically represented as shown in Fig. 1.1. It consists of

- 1 A transducer
- 2 A signal conditioner and transmitter, and
- 3 A display/recording device.

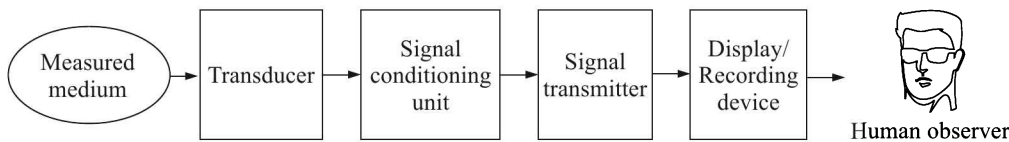


Fig. 1.1 A generalised instrument.

Transducer

A transducer senses the physical variable to be measured (i.e. measurand) and converts it to a suitable signal, preferably an electrical one.

One point has to be noted in this context. All transducers extract some energy from the measured medium which implies that the measurand is always disturbed by the measurement system. Therefore, *a perfect measurement is theoretically impossible*.

We will consider transducers in general in Chapter 5 and a few representative transducers for measurement of a few non-electrical quantities in Chapters 6 to 11.

Signal Conditioner and Transmitter

The signal generated by the transducer may need to be amplified, attenuated, integrated, differentiated, modulated, converted to a digital signal, and so on. The signal conditioner performs one or more such tasks. Since electrical signals have distinct advantages in this respect, more so with the development of electronics, a signal conditioner is now basically an electronic gadget. We will, however, discuss basics of signal conditioning in Chapter 16.

Signal transmitters are necessary for remote measurements. Remote measurements and control, called telemetry, is a highly-developed subject. We will exclude this topic from our consideration.

Display/Recording Device

The purpose of this element of an instrument is obvious—to communicate the information about the measurand to the human observer or to present it in an intelligible form. This aspect of instrumentation is discussed in Chapter 17.

We will study the subject element-wise. But before doing that we need to study the static (Chapter 2) and dynamic (Chapter 4) characteristics of instruments, and understand how to estimate errors (Chapter 3) because all these matters determine the performance of an instrumentation system.

Static Characteristics of Instruments

By static characteristics we mean attributes associated with static measurements or measurement of quantities which are constant or vary very slowly with time. For example, the measurement of emf (electromotive force) of a cell or the resistance of a resistor at constant temperature are both static measurements.

Static characteristics of instruments can broadly be divided into two categories—desirable and undesirable—each consisting of a few characteristics as shown in Fig. 2.1.

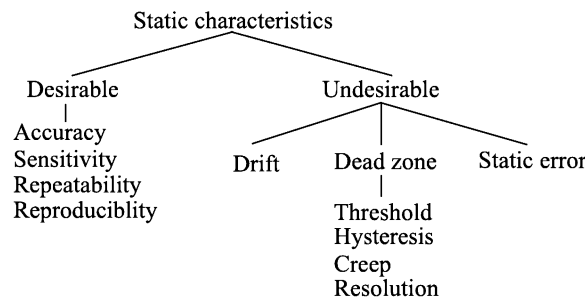


Fig. 2.1 Static characteristics tree.

2.1 Desirable Characteristics

The nature of these characteristics is discussed below.

Accuracy

Accuracy determines the closeness of an *instrument reading* to the true value of the measurand. Suppose, a known voltage of 200 V is being measured by a voltmeter and the successive readings are 204, 205, 203, 203 and 205 volts. So, the accuracy is about 2.5%. Here, though the repeatability of readings is not too bad, the accuracy is low because the instrument may be having a large calibration error. Hence, the accuracy can be improved upon by better calibration of the instrument.

Precision

Precision is another term which is often used in the same connotation as the accuracy. But in reality precision is different from accuracy. In the above example, the reading can be expressed as 204 ± 1 V, which means that the precision is a little less than 0.5% in this case.

Precision is, therefore, related to the repeatability¹ of the instrument reading and is a characteristic of the instrument itself. To improve the precision of an instrument, its design and construction have to be improved upon.

Symbolically, therefore, if a denotes accuracy, p the precision and c the calibration error then $a = p + c$.

Precision of a measurement also depends on what is called the number of significant figures. An example will perhaps make the point clear. Suppose the resistance of a conductor is being measured by an analogue ohmmeter. The ohmmeter indicates the true value, but the observer is unable to read the exact value because of lack of graduation beyond a certain number of decimals. Thus, though the instrument is showing the correct value, the precision of the measurement depends upon the number of significant figures to which the observer can read the value. And in an involved measurement where many measurands have to be combined, the number of significant figures plays a crucial role to determine the precision of the ultimate measurement. We discuss below this aspect in somewhat greater detail.

Significant figures

Significant figures convey information regarding the magnitude of precision of a quantity. For example, if a measurement reports that the line voltage is 220 V, it means that the line voltage is closer to 220 V than it is to 219 V or 221 V. Alternatively, if the reported value is 220.0 V, it means that the value is closer to 220.0 V than it is to 219.9 V or 220.1 V. Talking in terms of significant figures, it is 3 in the former case and 4 in the latter case. Significant figures play an important role in figuring out the final value in an involved measurement. Suppose, four resistors of values 28.4, 4.25, 56.605 and 0.76 ohms are connected in series.

$$\begin{array}{r} 28.4 \\ 4.25 \\ 56.605 \\ 0.76 \\ \hline 90.015 \end{array}$$

What should be the value for the total resistance? The general tendency is to report the result obtained by a straightforward addition, i.e. 90.015 ohms. But a close look reveals that this result conveys a wrong information regarding the precision of the measurement. If we signify doubtful figures by italics, it will be evident from the calculation shown above that the value when reported as 90.0 ohms would convey the right information regarding its precision. The simple rules arriving at such figures in mathematical manipulation of data are now enumerated.

Addition and subtraction. After performing the operation write the result rounded to the same number of *decimal places* as the least accurate figure.

¹See Section 2.1 at page 9.

Multiplication and division. After the operation round the result to the same number of *significant figures* as the least accurate number.

Example 2.1

Four capacitors of values 45.1, 3.22, 89.309 and 0.48 μF , are connected in parallel. Find the value of the equivalent capacitor to the appropriate number of significant figures.

Solution

The straightforward addition yields 138.109 μF . Rounding it off to the same number of decimal places as the least accurate figure, namely 45.1 μF , the acceptable value is 138.1 μF .

Example 2.2

A current of 3.12 A is flowing through a resistor of 53.635 Ω . Find the value of the voltage drop across the resistor to the appropriate number of significant figures.

Solution

The straightforward multiplication yields 167.3412 V. Rounding it off to three significant figures—the same as that of 3.12—the value to be reported is 167 V.

Sensitivity

Sensitivity is defined as the absolute ratio of the increment of the output signal (or response) to that of the input signal (or measurand). Stated mathematically, $S = \Delta q_o / \Delta q_i$ where q_i and q_o are input and output quantities respectively.

Suppose in a mercury-in-glass thermometer the meniscus moves by 1 cm when the temperature changes by 10°C. Its sensitivity is, therefore, 1 mm/°C.

The sensitivity of a voltmeter, however, is expressed in ohm/volt. A voltmeter is considered to be more sensitive if it draws less current from the circuit which, in turn, is ensured by the high resistance of the voltmeter that has to be connected in parallel with the circuit. For this reason, the sensitivity of a voltmeter varies inversely with the current required for full-scale deflection (FSD). Thus,

$$\text{Sensitivity} = \frac{1 \text{ (V)}}{I_{\text{FSD}} \text{ (A)}} \text{ ohm/volt}$$

where, I_{FSD} is the current required for FSD of the meter movement.

Example 2.3

What is the sensitivity of a voltmeter having 50 μA FSD?

Solution

The required sensitivity is given by

$$\text{Sensitivity} = \frac{1}{50 \times 10^{-6}} = 20,000 \text{ ohm/volt}$$

Lab quality voltmeters should have a minimum sensitivity of 20 k Ω /volt.

Linearity and nonlinearity

If the functional relationship between the input quantity and the output reading of an instrument is linear, we call it a linear instrument. For example, a mercury-in-glass thermo-

meter is a linear instrument while a simple thermocouple arrangement for measuring temperature is nonlinear.

The sensitivity of a linear instrument is constant while that of a nonlinear one varies from range to range as will be evident from Fig. 2.2.

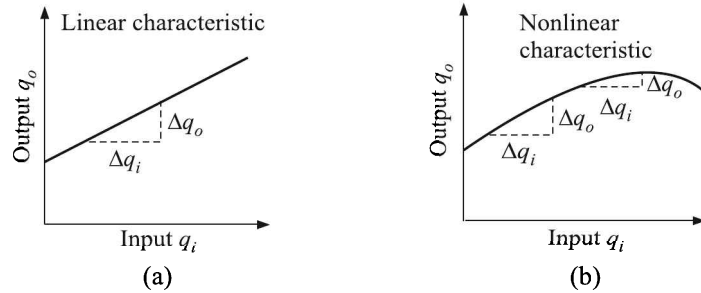


Fig. 2.2 Sensitivity: (a) linear instrument where the sensitivity is constant over the entire range, (b) nonlinear instrument where sensitivity varies from one range to another.

A perfectly linear instrument is rather difficult to realise, because almost all the so-called linear instruments show some deviation from linearity. This deviation may assume one of the following three forms as illustrated in Fig. 2.3.

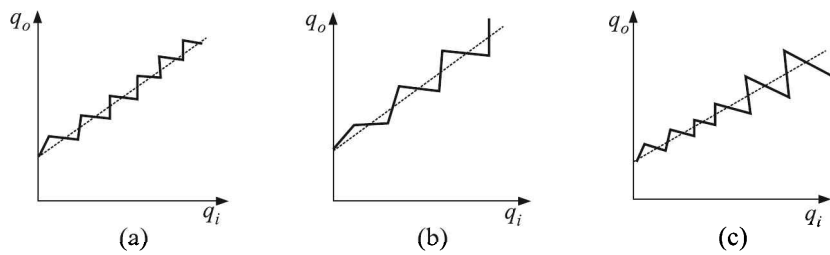


Fig. 2.3 Deviation from linearity: (a) oscillation with fixed amplitude, (b) oscillation with varying amplitude, and (c) combined type oscillation around the best-fit straight line.

1. The actual output of the instrument may oscillate with the same amplitude around the best-fit straight line. In this case, the nonlinearity is expressed in terms of the amplitude (or maximum deviation). The amplitude is calculated as the \pm of the full scale deflection (FSD).
2. The actual output of the instrument may oscillate around the best-fit straight line, but the amplitude of oscillation varies with the input value. Here, the nonlinearity is expressed as a function of the input value. Actually, the slopes of lines connecting positive and negative deviations are determined and the one with a higher deviation from the best-fit line is used to express the per cent nonlinearity with respect to the input value.
3. The actual output may oscillate with a fixed amplitude around the best-fit straight line over a certain range and then the amplitude may become a function of the input over the rest. In that case, two computations are made—one for the fixed amplitude part and expressed as $\pm\%$ of the FSD, and another for the varying amplitude part, expressed as $\pm\%$ of the input value. Nonlinearity is expressed in terms of the higher value.

Example 2.4

The output of a temperature transducer is recorded over its full-scale range of 25°C as shown below:

| | | | | | | |
|------------------------------|-----|-----|------|------|------|------|
| Calibration temperature (°C) | 0.0 | 5.0 | 10.0 | 15.0 | 20.0 | 25.0 |
| Output reading (°C) | 0.0 | 5.0 | 9.8 | 14.8 | 19.9 | 25.0 |

Determine (a) the static sensitivity of the device, and (b) the maximum nonlinearity of the device.

Solution

- Let q_i be the calibration temperature in °C
 q_o be the output reading in °C
 S be the sensitivity = $\Delta q_o / \Delta q_i$
 D be the deviation from the calibration temperature
 Δl be the nonlinearity = $100D/\text{FSD} = 4D$ since $\text{FSD} = 25^\circ\text{C}$.

Then, we have

| q_i (°C) | Δq_i (°C) | q_o (°C) | Δq_o (°C) | S | D (°C) | Δl (%) |
|------------|-------------------|------------|-------------------|------|----------|----------------|
| 0.0 | | 0.0 | | | 0.0 | 0.0 |
| 5.0 | 5.0 | 5.0 | 5.0 | 1.0 | 0.0 | 0.0 |
| 10.0 | 5.0 | 9.8 | 4.8 | 0.96 | -0.2 | 0.8 |
| 15.0 | 5.0 | 14.8 | 5.0 | 1.0 | -0.2 | 0.8 |
| 20.0 | 5.0 | 19.9 | 5.1 | 1.02 | -0.1 | 0.4 |
| 25.0 | 5.0 | 25.0 | 5.1 | 1.02 | 0.0 | 0.0 |

Thus $S_{\min} = 0.96$ and maximum nonlinearity = 0.8%

So far we discussed about the nonlinearity of an output when it is expected to be linear even from theoretical considerations. But there are cases where the output is not expected to be linear, and we have to resort to linear approximations for convenience. For example, the emf-temperature relation of a thermocouple can be written to a first approximation as

$$E = \alpha T + \beta T^2 \quad (2.1)$$

where, α and β are constants for a given thermocouple and T is the temperature of the hot junction, the cold junction being kept at 0°C. For such a thermocouple, if we assume emfs are E_1 and E_2 at temperatures T_1 and T_2 and that a linear relationship exists between emf and temperature, then

$$E_{\text{linear}} = \left(\frac{E_1 - E_2}{T_1 - T_2} \right) T \equiv \alpha' T \quad (2.2)$$

where, $\alpha' = \frac{E_1 - E_2}{T_1 - T_2}$. In this case, the nonlinearity N may be expressed as

$$N = E_{\text{actual}} - E_{\text{linear}} = (\alpha - \alpha')T + \beta T^2 \quad (2.3)$$

Alternatively, the nonlinearity may be defined in terms of the nonlinear term βT^2 in expression (2.1). We will consider a case in Example 10.1.

Repeatability

Repeatability is defined broadly as the measure of agreement between the results of successive measurements of the output of a measurement system for repeated applications of a given input in the same way and within the range of calibration of the measurement system. The tests should be made by the same observer, with the same measuring equipment, on the same occasion (i.e. successive measurements should be made in a relatively short span of time), without mechanical or electrical disturbance, and calibration conditions such as temperature, alignment of loading, and the timing of readings held constant as far as possible.

Reproducibility

Reproducibility is defined as the closeness of the agreement between the results of measurements of the same physical quantity carried out under *changed conditions of measurement*. A valid statement of reproducibility requires specification of the particular conditions changed and typically refers to measurements made weeks, months, or years apart. It would also measure, for example, changes caused by dismantling and re-assembling the equipment.

Reproducibility, therefore, determines precision of an instrument. The related undesirable characteristic is *drift*.

2.2 Undesirable Characteristics

As discussed before, the undesirable characteristics of instruments can be divided into three categories—drift, dead zone and static errors.

Drift

Drift denotes the change in the indicated reading of an instrument over time when the value of the measurand remains constant. If there is no drift, the reproducibility is 100%.

Several causes contribute to the drift. Stray electromagnetic fields, mechanical vibrations, changes in superincumbent temperature or pressure, Joule heating of the components of the instrument, etc. are some of the causes. In the case of suspended coil permanent magnet moving coil (PMMC) instruments the release of internal strain of the suspension wire causes drift of the zero-setting.

Dead Zone

Four phenomena—hysteresis, threshold, creep and resolution—contribute to the dead zone.

Hysteresis

Not all the energy put into a system while loading is recoverable upon unloading. For example, a spring balance may show one set of readings when the weight is increased in steps and another set of readings when the weight is decreased in steps. As a result the pointer reading vs weight plot may have the appearance of Fig. 2.4.

The loading and unloading curves do not coincide because of consumption of some energy by the internal friction of the solid, and also because of the external sliding friction between

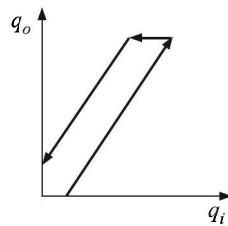


Fig. 2.4 Hysteresis effects shown in an exaggerated way.

components of the instrument. This phenomenon, which is akin to the one experienced during magnetising and demagnetising a magnetic material, is called ‘hysteresis’.

Threshold

Suppose an instrument is in its zero position, i.e. there is no input to it. If now an input is gradually applied to it, the instrument will require some minimum value of input before it shows any output (Fig. 2.5).

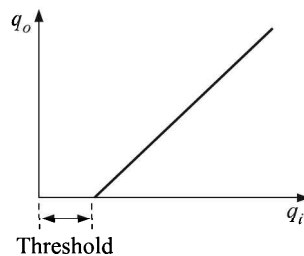


Fig. 2.5 Threshold effect.

This minimum input which is necessary to activate an instrument to produce an output is termed its *threshold*.

Creep

A measurement system may take some time to adjust fully to a change in the applied input, and the creep of a transducer is usually defined as the change of output with time following a step increase in the input from one value to another. Many instrument manufacturers specify the creep as the maximum change of output over a specified time after increasing the input from zero to the rated maximum input. Figure 2.6 shows an example of a creep curve where the transducer exhibits a change in output from R_1 to R_2 over a period of time from t_1 to t_2 after a step change between 0 and t_1 . In figures this might be, say, 0.03% of the rated output over 30 minutes.

Creep recovery is the change of output following a step decrease in the applied input to the transducer, usually from the rated maximum input to zero. For both creep and creep recovery, the results will depend on how long the applied input has been at zero or the rated value respectively before the input is changed.

Resolution

Even above the threshold input, an instrument needs a minimum *increment* in input to produce a perceptible output. This minimum necessary increment is called the *resolution* of

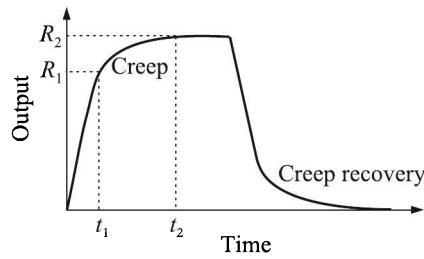


Fig. 2.6 Creep and creep recovery.

the instrument. Thus, resolution which denotes the smallest measurable *change in input* is similar to sliding friction while threshold signifies the smallest *initial input* resembling the static friction.

Example 2.5

An analogue ammeter has a linear scale of 50 divisions. Its full-scale reading is 10 A and half a scale division can be read. What is the resolution of the instrument?

Solution

1 scale division = $10/50$ A = 0.2 A. Thus, resolution = $1/2$ scale division = $(0.2/2)$ A = 0.1 A.

Example 2.6

The dead-zone in a pyrometer is 0.125% of the span. The instrument is calibrated from 800 to 1800°C. What temperature change must occur before it is detected?

Solution

The span is $(1800 - 800) = 1000^\circ\text{C}$. The dead zone is 0.125% of 1000°C , i.e. 1.25°C . Hence, no change in temperature below 1.25°C can be detected.

Static Errors

The tree in Fig. 2.7 depicts the classification of errors.

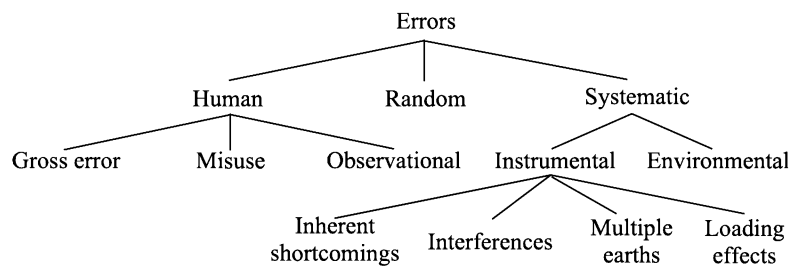


Fig. 2.7 The error tree.

Human errors

One may see from the tree that these errors can be subdivided into the following three classes—gross errors, misuse and observational errors.

Gross errors. Gross errors are basically human mistakes in reading or recording values. Suppose an instrument shows a value of 47.0 while the observer reads it as 42.0. Or, even if he reads the correct value, records it as 41.0. Such errors can be eliminated by automation or minimised by taking multiple readings of the same value at different times and by different observers.

Misuse. A casual approach on the part of the operator is the cause of this error. For example, in electrical measurements, if the leads are not connected firmly, or an ohmic contact² is not established or if the initial adjustment such as zero-checking is not done properly, or for a microvolt order measurement proper care is not taken to avoid thermo-emfs arising out of junctions of dissimilar metals, etc. errors will creep in. Alertness and perception on the part of the operator are the only remedy for such errors.

Observational errors. As distinct from gross errors or errors arising out of misuse, observational errors are caused by the observer's lack of knowledge in measurement methods. Parallax is one such error. There may be many more sources of observational errors from set-ups which depend on the so called eye estimation or human reflexes. One such example is the measurement of time period of a pendulum by a stopwatch. Here the precision of the measurement depends on the reflexes of the observer who clicks the stopwatch ON or OFF by noting the position of the bob of the pendulum visually.

Systematic errors

As shown in Fig. 2.7, this error may have two possible origins—instrumental and environmental.

Instrumental error. The instrumental error, in turn, may originate from four different causes—inherent shortcomings, interference, multiple earths and loading effects.

Inherent shortcomings. As the name implies, this error creeps in owing to malfunctioning of the components of instruments due to ageing, etc. For example, the spring of a galvanometer may become weak, thus changing its calibration. Therefore, to avoid this error the calibration of the instrument should be checked from time to time.

Interference. Thevenin's theorem states that a network consisting of linear impedances and voltage sources can be replaced with an equivalent circuit having a voltage source and a series impedance, while Norton's theorem states that such a network can be replaced with an equivalent circuit consisting of a current source in parallel with an impedance (Fig. 2.8).

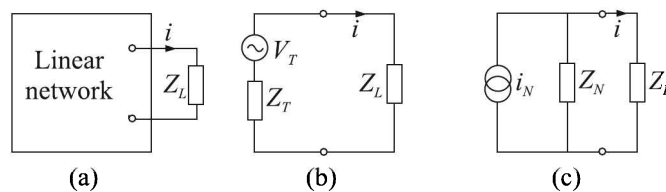


Fig. 2.8 (a) A linear circuit and its (b) Thevenin, and (c) Norton equivalents.

²If materials to be contacted are metals, it is easy to establish an ohmic contact between them through a proper cleaning of their surfaces. But if the contact is between a metal and a semiconductor, it is necessary to consider their Fermi levels or else, a rectifying contact may result. For a discussion on this, see *Solid State Electronic Devices*, 4th ed., by B G Streetman, Prentice-Hall of India (1993), pp 187-189.

Let us consider a voltage transmission system having a Thevenin voltage source V_T , Thevenin impedance Z_T , load impedance Z_L , cable resistance R_C and interference voltage V_I in series as shown in Fig. 2.9. Such interference is called the 'series mode interference'. Here, the current through the load is

$$i = \frac{V_T + V_I}{Z_T + R_C + Z_L}$$

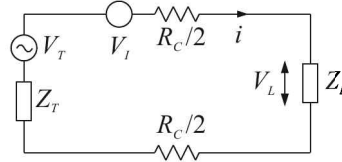


Fig. 2.9 Voltage transmission with series mode interference.

Therefore, the voltage across the load is

$$V_L = \frac{V_T + V_I}{Z_T + R_C + Z_L} \cdot Z_L \quad (2.4)$$

It is a normal practice to choose $Z_L \gg (R_C + Z_T)$ to ensure maximum voltage transfer to the load. With this condition Eq. (2.4) becomes

$$V_L = V_T + V_I \quad (2.5)$$

Equation (2.5) shows that the output contains unabated interference or noise voltage. The signal-to-noise ratio is defined as

$$\frac{S}{N} = 10 \log \frac{W_S}{W_N} \text{ dB} = 20 \log \frac{V_T}{V_I} \text{ dB}$$

where, W_S and W_N indicate signal and noise powers respectively. Thus, if $V_T = 1 \text{ mV}$ and $V_I = 0.1 \text{ mV}$, $S/N = 20 \text{ dB}$.

Next we consider a current transmission system having a Norton current source i_N along with the same interference voltage V_I in series, as shown in Fig. 2.10.

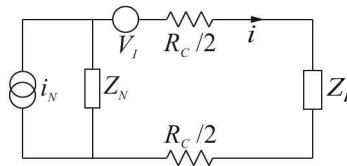


Fig. 2.10 Current transmission with series mode interference.

Then according to the rule of current division, current i through the load is

$$i = i_N \frac{Z_N}{Z_N + R_C + Z_L}$$

And the interference current through the load due to interference voltage is

$$i_I = \frac{V_I}{Z_N + R_C + Z_L}$$

Therefore, the total voltage across the load is

$$V_L = (i + i_I)Z_L = i_N Z_L \frac{Z_N}{Z_N + R_C + Z_L} + V_I \frac{Z_L}{Z_N + R_C + Z_L}$$

The normal practice is to make $Z_N \gg (R_C + Z_L)$ so that maximum current is transferred to the load. Then

$$V_L = i_N Z_L + V_I \frac{Z_L}{Z_N}$$

Since, $Z_L \ll Z_N$, the contribution of noise voltage to the output voltage is negligible in current transmission.

Next we consider a common mode interference where a common voltage V_c is added to both sides of the signal circuit (Fig. 2.11).

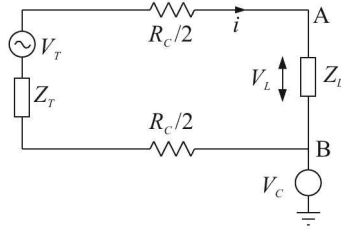


Fig. 2.11 Signal transmission with common mode interference.

Here, transmitted voltages at A and B are

$$V_A = V_c + V_T$$

$$V_B = V_c$$

Therefore,

$$V_L = V_A - V_B = V_T \quad (2.6)$$

Equation (2.6) shows that the common mode interference voltage V_c does not affect the voltage across the load.

Multiple earths. Although the earth potential is assumed to be zero, it may not be so everywhere owing to existence of leakage current from heavy electrical equipment. As a result, if a circuit is grounded at multiple points and if a potential difference exists between those points, then series and common mode interference may occur in measurements.

Consider the situation as shown in Fig. 2.12. As before, we assume, $Z_L \gg (Z_T + R_C)$. Hence negligible current flows through the ABCD loop.

But leakage impedances at the source Z_S and receiver Z_R exist and therefore a voltage difference V_E exists between the earth points at the source and the receiver. As a consequence, a current i_E flows in the circuit CFEB. It is given by

$$i_E = \frac{V_E}{Z_S + Z_E + Z_R + R_C/2} \quad (2.7)$$

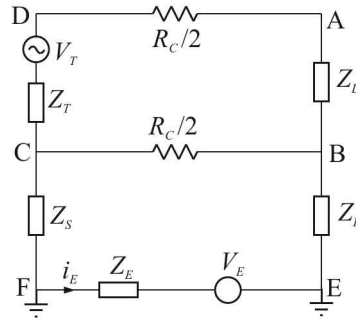


Fig. 2.12 Multiple earth situation.

Previously, potentials between (A and D) and (C and B) were equal because almost no current flowed through the ABCD loop. Now, because of i_E , potentials at these points are

$$\begin{aligned} V_C &= V_E - i_E(Z_E + Z_S) \\ V_B &= i_E Z_R \\ V_D &= V_A = V_E - i_E(Z_E + Z_S) + V_T \end{aligned} \quad (2.8)$$

V_B is common mode interference and its value is found from Eqs. (2.8) and (2.7) as

$$V_B = \frac{Z_R V_E}{Z_S + Z_E + Z_R + R_C/2}$$

The series mode interference can be obtained by figuring out the voltage across the load as

$$\begin{aligned} V_L &= V_A - V_B = V_E - i_E(Z_E + Z_S) + V_T - i_E Z_R \\ &= [V_E - i_E(Z_E + Z_S + Z_R)] + V_T \\ &= \frac{i_E R_C}{2} + V_T \end{aligned} \quad (2.9)$$

Thus, we find from Eq. (2.9) that the series mode interference is

$$V_I = \frac{i_E R_C}{2} = \frac{V_E R_C}{2(Z_S + Z_E + Z_R) + R_C}$$

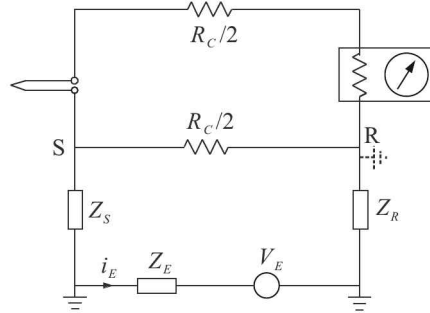
To minimise V_I as well as i_E , Z_S and Z_R should be as large as possible. But this may not always be possible in an industry. That may give rise to measurement error as will be clear from the following example.

Example 2.7

In a thermocouple installation the tip of the thermocouple touches the thermowell and the thermowell is bolted to a metal pipe which in turn is connected to one point in the earth plane. The emf generated is taken to a receiver 500 metres away by means of a cable of resistance 0.1Ω per metre. The receiver is isolated from earth but a 1 V potential difference exists between the source earth and the receiver earth. If this 1 V source has 1Ω impedance and the impedance between receiver and earth is $10^6 \Omega$ while that between the source and the earth is 10Ω , calculate the series mode interference voltage. What changes will be observed if the receiver is earthed?

Solution

The diagram for the problem is shown below.



Given,

$$\begin{aligned} R_C/2 &= (500 \times 0.1) \Omega = 50 \Omega \\ V_E &= 1 \text{ V} & Z_E &= 1 \Omega \\ Z_S &= 10 \Omega & Z_R &= 10^6 \Omega \end{aligned}$$

Therefore,

$$\begin{aligned} i_E &= \frac{V_E}{Z_S + Z_R + Z_E + R_C/2} \\ &= \frac{1}{10^6 + 10 + 1 + 50} = 0.999 \times 10^{-6} \text{ A} \end{aligned}$$

Series mode interference voltage, $V_I = i_E \cdot R_C/2 = 0.999 \times 10^{-6} \times 50 \cong 50 \mu\text{V}$. Now, if the receiver is earthed, $Z_R = 0$. Then,

$$i_E = \frac{V_E}{Z_S + Z_E + R_C/2} = \frac{1}{10 + 1 + 50} = 0.0164 \text{ A}$$

Therefore,

$$V_I = i_E \cdot R_C/2 = 0.0164 \times 50 = 0.82 \text{ V}$$

Thus, the series mode interference voltage increases from $50 \mu\text{V}$ to 820 mV when the receiver is earthed.

Loading effects. We have already mentioned that any measurement involves extraction of some energy, however small, from the measured medium changing thereby the value of the measurand from its pristine undisturbed state. This makes perfect measurement theoretically impossible.

This phenomenon of extraction of energy by a measurement system from the measured medium is known as the *loading effect*. We will consider this effect from the standpoint of electrical measurements.

Parallel or shunt connection (voltage measurement): Suppose the voltage across the terminals A and B of the circuit in Fig. 2.13 is being measured in the usual way. No sooner is the voltmeter attached to the terminals A and B, than the circuit is changed and the value of E_o is altered. The following analysis will make the point clear.

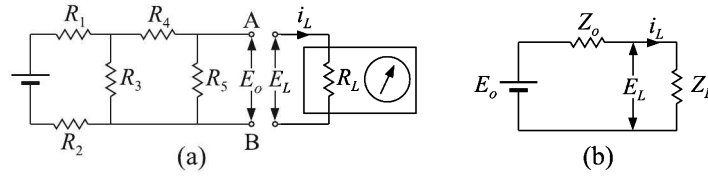


Fig. 2.13 Loading effect in voltage measurement: (a) schematic arrangement, (b) its Thevenin equivalent.

If Z_o is the true impedance between A and B, Z_L is the impedance of the loading circuit (here, the voltmeter), E_o is the true voltage between A and B, and E_L is the voltage seen by the loading circuit, their values may be derived as follows:

$$Z_o = \frac{\left[\frac{(R_1 + R_2)R_3}{R_1 + R_2 + R_3} + R_4 \right] R_5}{\frac{(R_1 + R_2)R_3}{R_1 + R_2 + R_3} + R_4 + R_5}$$

$$Z_L = R_m, \quad E_L = E_o - i_L Z_o = i_L Z_L \quad E_o = i_L (Z_o + Z_L)$$

Therefore,

$$\frac{E_L}{E_o} = \frac{i_L Z_L}{i_L (Z_o + Z_L)} = \frac{1}{1 + (Z_o/Z_L)}$$

or

$$E_L = \frac{E_o}{1 + (Z_o/Z_L)} \quad (2.10)$$

For the sake of simplicity, we made a dc analysis, but it holds true for ac as well. In fact, the voltage is modified both in magnitude and phase in the case of ac. The point will be clear from Example 2.11. In any case, it is clear from this analysis that the instrument will give true result if $Z_L \rightarrow \infty$, and a reasonably accurate one if $Z_L \gg Z_o$.

Series connection (current measurement): If AB in Fig. 2.14 is shorted, and i_o is the current flowing through the circuit, then $i_o = E_o/Z_o$. When the current in the circuit is measured by an ammeter of impedance Z_L , the current changes from i_o to i_L . Now,

$$i_L = \frac{E_o}{Z_o + Z_L} = \frac{i_o Z_o}{Z_o + Z_L} = \frac{i_o}{1 + (Z_L/Z_o)}$$

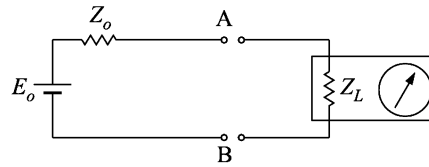


Fig. 2.14 Loading effect in current measurement.

Therefore, to minimise the measurement error, i.e. to make $i_L \rightarrow i_o$, it is necessary that $Z_o \gg Z_L$.

Example 2.8

What is the value of current in R in Fig. 2.15(a)? If an ammeter of resistance $2\text{ k}\Omega$ is used to measure the current what will it read? If an accuracy of 99% is desired, what should the ammeter resistance be?

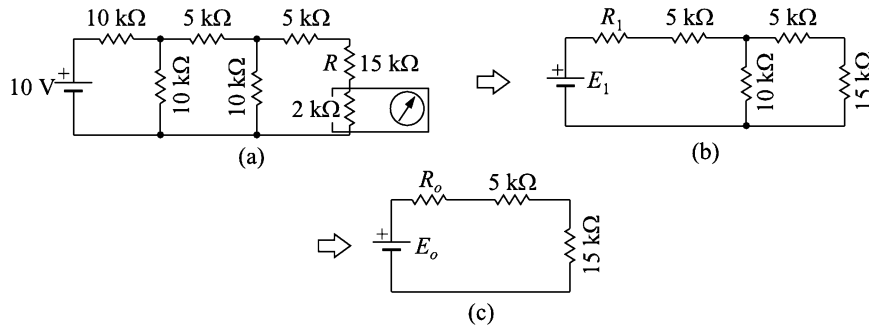


Fig. 2.15 Example 2.8.

Solution

Successive Thevenin-equivalents of the circuit are shown in Fig 2.15(b) and (c). From these we find

$$E_1 = \frac{10}{10 + 10} \times 10 \text{ V} = 5 \text{ V}$$

$$R_1 = 10 \text{ k}\Omega \parallel 10 \text{ k}\Omega = \frac{10 \times 10}{10 + 10} \text{ k}\Omega = 5 \text{ k}\Omega$$

Similarly,

$$E_o = 2.5 \text{ V} \quad R_o = 5 \text{ k}\Omega.$$

Actual value of the current through R is

$$I_o = \frac{E_o}{R_o + 5 + 15} \text{ mA} = \frac{2.5}{25} \text{ mA} = 100 \text{ }\mu\text{A}$$

The ammeter will read

$$I_L = \frac{E_o}{R_o + 5 + 15 + 2} \text{ mA} = \frac{2.5}{27} \text{ mA} = 92.6 \text{ }\mu\text{A}$$

For 99% accuracy,

$$\frac{I_L}{I_o} = \frac{1}{1 + \frac{R_L}{R_o + (5 + 15) \times 10^3}} = 0.99$$

\Rightarrow

$$\frac{1}{1 + \frac{R_L}{25 \times 10^3}} = 0.99$$

The left hand side can be written approximately as

$$1 - \frac{R_L}{25 \times 10^3} = 1 - 0.01$$

\Rightarrow

$$R_L = 25 \times 10^3 \times 0.01 = 250 \text{ }\Omega$$

Example 2.9

What is the true value of the voltage across the terminals A and B (Fig. 2.16)? What would a voltmeter of 20 kΩ/V sensitivity read on the 50 V and 10 V ranges?

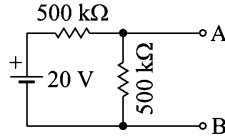


Fig. 2.16 Example 2.9.

Solution

Current in the circuit, $I = 20/(500 + 500)$ mA = 0.02 mA. Therefore, $E_o = (0.02 \times 10^{-3} \times 500 \times 10^3)$ V = 10 V. But the voltmeter offers different load resistances in its different ranges.

In the 50 V range: The load resistance, $R_L = (20 \times 10^3 \times 50)$ Ω = 10^6 Ω. Therefore,

$$E_L = \frac{10}{1 + \frac{250 \times 10^3}{10^6}} = 8.0 \text{ V}$$

In the 10 V range: $R_L = (20 \times 10^3 \times 10) = 2 \times 10^5$ Ω. So,

$$E_L = \frac{10}{1 + \frac{250 \times 10^3}{2 \times 10^5}} \approx 4.4 \text{ V}$$

Note: How a wrong setting of the instrument can introduce a huge error!

Example 2.10

What percentage error may be expected in measuring the voltage by the arrangement shown in Fig. 2.17(a)?

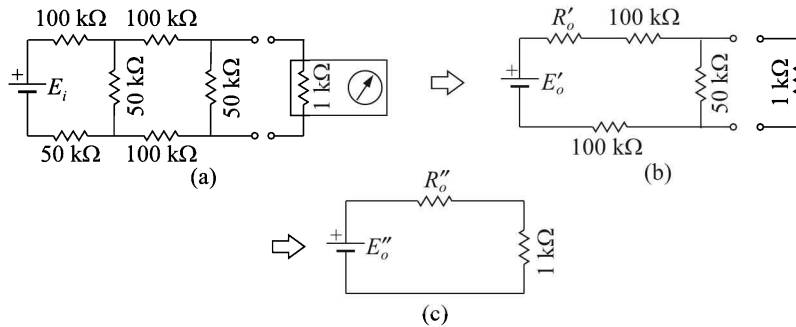


Fig. 2.17 Example 2.10.

Solution

Computing the Thevenin equivalent of the first stage on the left, we get

$$R'_o = \frac{100 \times (50 + 50)}{100 + 50 + 50} = 50 \text{ } \Omega$$

$$\begin{aligned}
 E'_o &= \frac{E_i}{100 + 50 + 50} \times 50 \\
 &= \frac{E_i}{4} \text{ V}
 \end{aligned}$$

Thus, the effective circuit becomes that as shown in Fig. 2.17(b). The Thevenin equivalent of this circuit is

$$\begin{aligned}
 R''_o &= \frac{(R'_o + 100)(100 + 50)}{R'_o + 100 + 50 + 100} \\
 &= 75 \Omega \\
 E''_o &= \frac{E'_o \times 50}{R'_o + 100 + 50 + 100} \\
 &= \frac{E_i}{4} \times \frac{50}{300} \\
 &= \frac{E_i}{24} \text{ V}
 \end{aligned}$$

The effective circuit now looks like Fig. 2.17(c). The voltage developed across the 1 k Ω resistance of the measuring instrument is

$$E_M = \frac{E''_o \times 1000}{R''_o + 1000} = E''_o \times \frac{1000}{1075} = 0.9302E''_o$$

Therefore, the error in measurement is

$$\frac{E''_o - E_M}{E''_o} \times 100 = (1 - 0.9302) \times 100 = 6.98\%$$

Example 2.11

An oscilloscope having an input resistance of 1 M Ω shunted by a 50 pF capacitor is connected across a circuit having an effective resistance of 10 k Ω (Fig. 2.18). If the open circuit voltage has 1.0 V peak for a sine wave, what voltage will the oscilloscope indicate when the frequency is (a) 100 kHz and (b) 1 MHz?

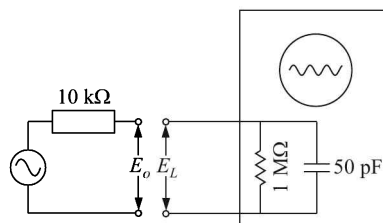


Fig. 2.18 Example 2.11.

Solution

(a) *100 kHz setting*: The relevant quantities expressed by usual symbols are:

$$X_c = \frac{1}{2\pi fC} = \frac{1}{2\pi \times 10^5 \times 50 \times 10^{-12}} = \frac{10^5}{\pi} \Omega$$

$$Z_L = \frac{R_L(-jX_c)}{R_L - jX_c} = \frac{10^6(-j10^5)/\pi}{10^6 - j10^5/\pi} \approx -j\frac{10^5}{\pi} \Omega$$

$$E_L = \frac{E_o}{1 + \frac{Z_o}{Z_L}} = \frac{1}{1 + \frac{j\pi}{10}} = \frac{1 - j\frac{\pi}{10}}{1 + (\frac{\pi}{10})^2}$$

$$= \frac{1 - j0.314}{1.098} = 0.91 - j0.286 = 0.954 \text{ V} \angle -17.4^\circ$$

(b) *1 MHz setting*: Here,

$$X_C = \frac{10^4}{\pi} \Omega \quad Z_L \cong -j\frac{10^4}{\pi} \Omega$$

$$E_L = \frac{1}{1 + j\pi} = \frac{1 - j\pi}{1 + \pi^2} = 0.092 - j0.289 = 0.303 \text{ V} \angle -72.3^\circ$$

Note: In the second case not only is the measured voltage much lower than the actual value, but also the phase change is substantial.

In this context of loading a measuring medium by a measuring instrument, we now discuss an important relevant theorem, called the *maximum power transfer theorem*.

Maximum power transfer theorem. This theorem states that for maximum power transfer to the load, the Thevenin-equivalent resistance R_o of the circuit should be equal to the load resistance, R_L .

Proof: Power transferred to the load,

$$P = \frac{E_L^2}{R_L} = \frac{E_o^2 R_L}{(R_o + R_L)^2} \quad (2.11)$$

$$= \frac{E_o^2}{R_L [1 + (R_o/R_L)]^2} \quad (2.12)$$

For maximum power transfer,

$$\frac{dP}{dR_L} = 0$$

Therefore,

$$\frac{E_o^2}{(R_o + R_L)^2} \left(1 - \frac{2R_L}{R_o + R_L}\right) = 0$$

or

$$R_L = R_o$$

In the case of ac circuits,

if $Z_o = \text{impedance of the source} = R_o + jX_o$

$Z_o^* = \text{complex conjugate of } Z_o$

$Z_L = \text{impedance of the load} = R_L + jX_L$

it can be shown that for maximum power transfer,

$$Z_L = Z_o^* = R_o - jX_o$$

Four points have to be noted in this context:

1. It is obvious that as $R_L \rightarrow 0, P \rightarrow 0$ [see Eq. (2.11)], and that as $R_L \rightarrow \infty, P \rightarrow 0$ [see Eq. (2.12)].
2. $P_{\max} = E_o^2/4R_o$.
3. Impedance matching is not critical, because,

$$\frac{P}{P_{\max}} = \frac{4(R_L/R_o)}{[1 + (R_L/R_o)]^2} \quad (2.13)$$

Hence, for a 10% deviation, i.e. $(R_L/R_o) = 1.1$ or 0.9 , $P/P_{\max} \cong 1$, which means power transfer is almost 100% for this impedance mismatch. For a 20% deviation, the corresponding figure is nearly 99% and even for a 100% deviation the power transfer is as much as 89%. Figure 2.19 offers a visual estimation of the amount of power transfer vis-a-vis the impedance mismatch as judged from the present theorem.

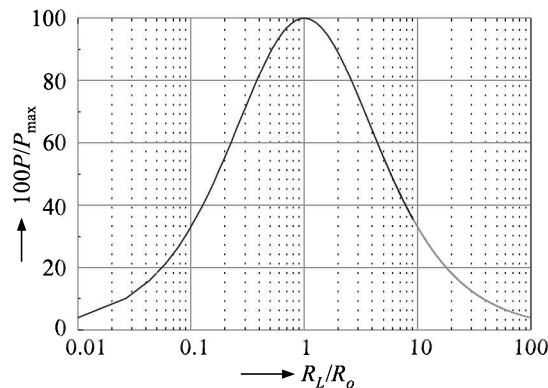


Fig. 2.19 Impedance matching characteristics.

4. At $R_L = R_o$, the current in the circuit $I = E_o/(2R_o)$. Therefore, the power absorbed by the load is

$$P_L = I^2 R_o = \frac{E_o^2}{4R_o}$$

But, the total available power is

$$P_T = I^2 (2R_o) = \frac{E_o^2}{2R_o}$$

Thus, the efficiency at maximum power transfer condition is

$$\frac{P_L}{P_T} = \frac{1}{2} = 50\%$$

Example 2.12

A source having an open circuit voltage of 20 V and an output impedance of $(0.5 + j1) \Omega$ is connected through a transmission network of impedance $(1.5 + j4) \Omega$. What should be the load impedance so that the maximum power will be delivered to it? Calculate the maximum deliverable power.

Solution

$Z_{\text{source}} = (0.5 + j1) \Omega$. $Z_{\text{trans}} = (1.5 + j4) \Omega$. Therefore, $Z_o = Z_{\text{source}} + Z_{\text{trans}} = (2 + j5) \Omega$. Hence for maximum power transfer, $Z_{\text{load}} = Z_o^* = (2 - j5) \Omega$. $E_o = 20$ V. So, the transferred power = $20^2 / (4 \times 2)$ W = 50 W.

Example 2.13

In a series circuit if E_o , R_o and R_L are the source voltage, source resistance and load resistance respectively, and P and P_{max} are power transferred to the load and the maximum power that can be transferred to the load respectively. Find the value of the ratio P/P_{max} in per cent when the source resistance is 50% of the load resistance.

Solution

Given, $R_L/R_o = 2$. Hence from Eq. (2.13), we have

$$\frac{P}{P_{\text{max}}} = \frac{4 \times 2}{(1 + 2)^2} = \frac{8}{9} = 88.9\%.$$

Example 2.14

A human nerve cell has an open circuit voltage of 80 mV and it can deliver a current of 5 nA through a 6 M Ω load. What is the maximum power available from the cell?

Solution

Given $E_o = 80$ mV, $I_L = 5$ nA and $R_L = 6$ M Ω . Now,

$$E_o = I_L(R_o + R_L)$$

which gives

$$R_o = \frac{E_o}{I_L} - R_L = \left(\frac{80 \times 10^{-3}}{5 \times 10^{-9}} - 6 \times 10^6 \right) \Omega = 10 \text{ M}\Omega$$

Therefore, the maximum available power is

$$P_{\text{max}} = \frac{(80 \times 10^{-3})^2}{4 \times 10 \times 10^6} = 0.16 \text{ nW}$$

Environmental error. Environmental factors, such as atmospheric pressure, temperature and humidity affect many measurements and consequently change certain parameters. Consider the simple length measurement with the help of a scale. Plastic scales, now very common,

change their lengths in humid conditions, while metal scales, although enjoying immunity from humidity, are affected by changes in temperature. Vibrations caused by running machinery or vehicles play havoc in measurements involving highly-sensitive instruments such as electron microscopes. Most electromagnetic instruments are affected by stray electromagnetic fields.

The remedy from such factors is, of course, controlling temperature and humidity or using vibration-free mountings for instruments and shielding electromagnetic fields. Theoretical corrections may be made for factors such as atmospheric pressure, gravity, etc. which cannot be controlled.

Random error

Even if all the sources of error, as narrated above, are taken care of, some fluctuation in the measured value remains. This fluctuation is caused by many random happenings such as cosmic ray showers, changes in geomagnetism, thunder-cloud activities, minor earth tremors, etc. The effect of such disturbances which we are not aware of, are grouped together and are termed *random* (or *residual*) errors. These errors can be estimated from a statistical treatment of data.

In the next chapter, we will consider the estimation of static errors in measurement.

Review Questions

- 2.1 What are the different errors encountered in measurements? Explain with suitable examples.
- 2.2 Define and explain briefly the static performance parameters of instruments.
- 2.3 Explain, giving an example, as to why the input effective resistance of the measuring instrument should be very high like that of a potentiometer or a voltage follower if we want to measure temperature by a thermocouple accurately.
- 2.4 What are various sources of gross, systematic and random errors in a process of measurement? How are these errors minimised?
- 2.5 What are various sources systematic errors? How do these errors influence the accuracy of measurements?
- 2.6 The true value of a voltage is 100 V. Values indicated by a measuring instrument are 104, 103, 105, 103 and 105 volts. Find the accuracy of the measurement and the precision of the instrument.
- 2.7 A voltmeter has a uniform scale with 100 divisions. The full-scale reading is 5 V and $\frac{1}{5}$ th of a division can be read. What is the resolution of the instrument?
- 2.8 The dead zone in a pyrometer is 0.125% of the span. The instrument is calibrated from 800 to 1800°C. What temperature change must occur before it is detected?
- 2.9 Define the terms accuracy, precision, resolution and sensitivity.
- 2.10 List the desirable and undesirable static characteristics of instruments.
- 2.11 Explain with an example the difference between accuracy and precision of measurement. On what factors does precision depend? How can the accuracy be improved upon?

2.12 Systematic errors can be classified as

- (a) Instrumental errors
- (b) Environmental errors
- (c) Observational errors.

Discuss the above types of errors giving suitable examples. Explain the measures taken to minimise these errors.

2.13 Explain with example the terms 'static sensitivity', 'linearity', 'hysteresis' and 'dead-zone' in an instrumentation system.

2.14 Choose the correct answers:

- (a) A $50\ \Omega$ resistor dissipates 2 W of power. The voltage across the resistor is
 - (i) 100 V
 - (ii) 25 V
 - (iii) 12.5 V
 - (iv) 10 V
- (b) The errors committed by a person in the measurement are
 - (i) gross errors
 - (ii) random errors
 - (iii) instrumental errors
 - (iv) environmental errors
- (c) A reading is recorded as $68.0\ \Omega$. The reading has
 - (i) three significant figures
 - (ii) five significant figures
 - (iii) four significant figures
 - (iv) none of the above
- (d) The degree of reproducibility among several independent measurements of the same value under reference conditions is known as
 - (i) accuracy
 - (ii) precision
 - (iii) linearity
 - (iv) calibration
- (e) In an instrument the smallest measurable input is known as
 - (i) threshold
 - (ii) resolution
 - (iii) dead zone
- (f) The threshold of an instrument is normally defined
 - (i) as the smallest measurable input change (non-zero value) which can be detected
 - (ii) as the smallest measurable input which can be detected
 - (iii) in terms of linearity of scale
 - (iv) as a function of drift

- (g) The term 'precision' used in instrumentation means
- gradual departure of the measured value from the calibrated value.
 - smallest increment in the measurand that can be detected by the instrument
 - maximum distance or angle through which any part of a mechanical system may be moved in one direction without causing motion of the next part
 - the ability of the instrument to give output readings close to each other, when the input is constant.
- (h) A voltmeter connected across the $10\text{ k}\Omega$ resistor in Fig. 2.20 reads 5 V . The voltmeter is rated at 1000 ohms/volt and has a full scale reading of 10 V .

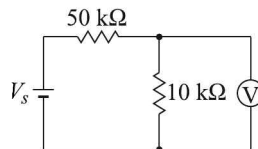


Fig. 2.20

The supply voltage V_s in volt is

- 30
 - 50
 - 55
 - 80
- (i) Threshold of a measurement system is
- the smallest change in input which can be detected
 - a measure of linearity of the system
 - the smallest input which can be detected
 - a measure of precision of the system
- (j) A common practice of reducing hysteresis error in the output for a given value of input is
- to maintain a high rate of change of input
 - to maintain a low rate of change of input
 - to take observations either in the ascending or in the descending order
 - to take observations both in the ascending and descending orders and then take average value of the output
- (k) The power supplied by the voltage source in the circuit, shown in Fig. 2.21, is

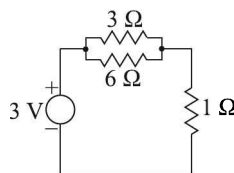


Fig. 2.21

- (i) 0 W
 - (ii) 1.0 W
 - (iii) 2.5 W
 - (iv) 3.0 W
- (l) The current I supplied by the dc voltage in the circuit, shown in Fig. 2.22, is

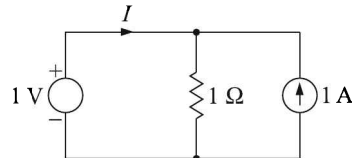


Fig. 2.22

- (i) 0 A
 - (ii) 0.5 A
 - (iii) 1 A
 - (iv) 2 A
- 2.15 Why is the linearity of an instrument an important specification? How is it expressed?
- 2.16 An ammeter has a range of 0 to 30 A. The instrument gave the following readings:

| | | | | | | |
|----------------------------|---|---|----|----|----|----|
| <i>Current flow (A)</i> | 0 | 5 | 10 | 15 | 20 | 25 |
| <i>Ammeter reading (A)</i> | 1 | 4 | 12 | 14 | 22 | 28 |

The nonlinearity of the instrument in terms of full scale reading (FSR) = ... % FSR

- 2.17 Define (any four) of the following:
- (a) Instrumental error
 - (b) Limiting error
 - (c) Calibration error
 - (d) Environmental error
 - (e) Random error
 - (f) Probable error
- 2.18 A circuit arrangement consists of a dc voltage source of 150 V in series with two resistors of value 100 kΩ and 50 kΩ respectively. It is desired to measure the voltage across the 50 kΩ resistor. Two voltmeters are available for this measurement: Voltmeter 1 with a sensitivity of 1 kΩ/V and Voltmeter 2 with a sensitivity of 20 kΩ/V. Both meters are used on their 50 V range. Calculate
- (a) The reading of each meter, and
 - (b) The error in each reading expressed as a percentage of the true value.

2.19 In the circuit shown in Fig. 2.23, the voltmeter is connected across AB. If the voltmeter has a resistance of $1200\ \Omega$, the measured value differs from the true value by ... V.

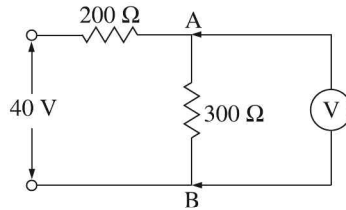


Fig. 2.23

2.20 Match the terms and their definitions:

- | | |
|-------------------|--|
| (a) Accuracy | (e) The ability to display the same reading when measuring a quantity |
| (b) Sensitivity | (f) Smallest perceptible change in the output |
| (c) Repeatability | (g) Error between the measured value and the absolute value |
| (d) Resolution | (h) Closeness of the measurement to the mean value |
| | (i) Ratio of change in the instrument reading to the change in the measured quantity |

Estimation of Static Errors and Reliability

While reporting a measured value of a quantity, it is necessary to indicate the possible error in the measurement. A question may be asked, how do we estimate a measurement error? Before going into that, let us define a few necessary parameters.

3.1 Definition of Parameters

The parameters we need to define are

1. Error
2. Scale range
3. Scale span
4. Limiting error
5. Probable error

Error

If X_m is the measured value of the quantity, and X_t is the true value of the quantity then the absolute static error is defined as

$$\varepsilon_0 = |X_m - X_t|$$

Often a relative static error is reported. Its definition is

$$\varepsilon_r = \frac{|X_m - X_t|}{X_t} = \frac{\varepsilon_0}{X_t} \quad (3.1)$$

Mostly, an error is much less than the true value making $\varepsilon_r < 1$. Now from Eq. (3.1), we get

$$\varepsilon_r = \frac{X_m}{X_t} - 1$$

$$\Rightarrow X_t = \frac{X_m}{1 + \varepsilon_r} \cong X_m(1 - \varepsilon_r)$$

While defining the absolute and relative static errors, we have used the term *true value*. The questions are, what is the true value of a physical quantity and how do we know it? It is said

that if we make an infinite number of measurements with the help of a *calibrated* measuring instrument and observe that the individual measurements agree between themselves within a specified degree of accuracy, we may assume that the measured value is the true value of the quantity.

Scale Range

The terms *scale range* is defined as follows. If X_{\min} and X_{\max} are the minimum and maximum values that an instrument can measure, then

$$\text{Scale range} = \text{Between } X_{\min} \text{ and } X_{\max}$$

Sometimes, the *dynamic range* of an instrument is specified.

Dynamic range

The dynamic range is defined as

$$\text{Dynamic range } N = \frac{\text{Range of operation}}{\text{Resolution}}$$

It is a common practice to specify dynamic range in dB as

$$\text{Dynamic range} = 20 \log_{10} N$$

Thus, an instrument having a 40 dB dynamic range means that it can handle input sizes of 100 to 1.

Example 3.1

A voltmeter has a range of [4 V, 20 V] and a resolution of 1 mV. The dynamic range of the instrument is

- (a) 21 dB (b) 60 dB (c) 72 dB (d) 84 dB

Solution

The range of operation of the instrument is $(20 - 4) = 16$ V and the resolution is 1×10^{-3} V. So,

$$\text{Dynamic range} = 20 \log \frac{16}{1 \times 10^{-3}} = 84 \text{ dB}$$

Therefore, the answer is (d).

Scale Span

If X_{\min} and X_{\max} are the minimum and maximum values that an instrument can measure, then the scale span is defined as

$$\text{Scale span} = X_{\max} - X_{\min}$$

Example 3.2

A voltmeter is calibrated between 10 V and 250 V. The scale span and scale range are respectively

- (a) 250, 250 (b) 240, 250 (c) 250, 240 (d) 240, 240

Solution

The nearest answer is (d). But it is better said that the scale range is 10 to 250 V.

3.2 Limiting Error

Suppose the length of a rod is being measured with the help of a vernier scale which has a vernier constant of 0.1 mm. One may measure the length only once by the vernier scale and report value as $L \pm 0.1$ mm, if the measured value is L mm. This reported error is called the *limiting error* (or *guarantee error*), because this is the maximum error which might have occurred during the measurement, assuming that the vernier scale has no calibration error. Many components (e.g. resistor, capacitor) or instruments are sold by manufacturers with some limits in their values or readings and indicated by gold or silver bands. These are limiting errors of the components.

Probable Error

Alternatively, in the foregoing example, one may measure the same length a number of times, take the arithmetic mean of the values obtained, calculate the error by one of the statistical methods (discussed later) and report the value as $L \pm \Delta L$ mm. This error may be termed *probable error*.

Estimation of limiting error for a single measurand is pretty simple, though it may be a bit involved for a measurement involving many measurands each having its own limiting error, while for probable errors even a single measurand demands attention.

We will take up both these methods of estimation one after another. It will be seen that the statistical treatment offers a more optimistic picture in the final analysis.

Combination of Limiting Errors

Suppose u and v are measurands and X is the final result. For simplicity we consider two measurands only though it is easy to generalise the results for many measurands. A few fundamental mathematical operations such as addition, subtraction, multiplication, division, raising to powers which connect the measurands to the final result are individually considered below.

Addition and subtraction. Here, $X = u \pm v$. Then

$$\frac{dX}{X} = \frac{du}{X} \pm \frac{dv}{X} = \frac{u}{X} \frac{du}{u} \pm \frac{v}{X} \frac{dv}{v}$$

But because errors are $\pm \delta u$ and $\pm \delta v$, we have for both addition and subtraction,

$$\frac{\delta X}{X} = \pm \left(\frac{u}{X} \frac{\delta u}{u} + \frac{v}{X} \frac{\delta v}{v} \right)$$

Multiplication and division. Here, either $X = uv$ or $X = u/v$. Taking logarithm of both sides, the two cases can be written as

$$\ln X = \ln u \pm \ln v$$

Taking differentials, we get

$$\frac{dX}{X} = \frac{du}{u} \pm \frac{dv}{v}$$

But, because of the nature of the errors as stated before,

$$\frac{\delta X}{X} = \pm \left(\frac{\delta u}{u} + \frac{\delta v}{v} \right)$$

Indices. Say, $X = u^m v^n$. On logarithmic differentiation,

$$\frac{dX}{X} = m \frac{du}{u} + n \frac{dv}{v}$$

Hence,

$$\frac{\delta X}{X} = \pm \left(m \frac{\delta u}{u} + n \frac{\delta v}{v} \right)$$

Example 3.3

Three resistors have the following values: $R_1 = 200 \Omega \pm 10\%$, $R_2 = 100 \Omega \pm 5\%$ and $R_3 = 50 \Omega \pm 5\%$. Determine the magnitude of the resultant resistances and the limiting errors if they are connected in (a) series, and (b) parallel.

Solution

(a) $R = R_1 + R_2 + R_3 = 350 \Omega$. $\frac{\delta R_1}{R_1} = 0.1$ and $\frac{\delta R_2}{R_2} = \frac{\delta R_3}{R_3} = 0.05$.

Therefore,

$$\begin{aligned} \frac{\delta R}{R} &= \frac{R_1}{R} \frac{\delta R_1}{R_1} + \frac{R_2}{R} \frac{\delta R_2}{R_2} + \frac{R_3}{R} \frac{\delta R_3}{R_3} \\ &= \frac{200}{350}(0.1) + \frac{100}{350}(0.05) + \frac{50}{350}(0.05) \cong 0.079 \end{aligned}$$

Thus, $R = 350 \Omega \pm 7.9\%$.

(b) Here $\frac{1}{R} = \frac{1}{200} + \frac{1}{100} + \frac{1}{50} = \frac{7}{200}$. Hence, $R \cong 28.6 \Omega$.

Again,
$$d\left(\frac{1}{R}\right) = d\left(\frac{1}{R_1}\right) + d\left(\frac{1}{R_2}\right) + d\left(\frac{1}{R_3}\right)$$

which gives
$$\frac{dR}{R^2} = \frac{dR_1}{R_1^2} + \frac{dR_2}{R_2^2} + \frac{dR_3}{R_3^2}$$

Therefore,

$$\begin{aligned} \frac{\delta R}{R} &= \frac{R}{R_1} \frac{\delta R_1}{R_1} + \frac{R}{R_2} \frac{\delta R_2}{R_2} + \frac{R}{R_3} \frac{\delta R_3}{R_3} \\ &= \frac{28.6}{200}(0.1) + \frac{28.6}{100}(0.05) + \frac{28.6}{50}(0.05) \cong 0.057 \end{aligned}$$

Thus,

$$R = 28.6 \Omega \pm 5.7\%$$

Example 3.4

The following are the data for a Hay's ac bridge: $R_1 = 1000 \Omega \pm 1$ part in 10,000, $R_2 = 16,800 \Omega \pm 1$ part in 10,000, $R_3 = 833 \pm 0.25 \Omega$, $C = 1.43 \pm 0.001 \mu\text{F}$. If frequency of the supply voltage is 50 ± 0.1 Hz and the formulae for L and R in the balanced condition of the bridge are

$$L = \frac{CR_1R_2}{1 + \omega^2C^2R_3^2} \quad \text{and} \quad R = \frac{R_1R_2R_3C^2\omega^2}{1 + \omega^2C^2R_3^2}$$

determine the values of L and R of the coil and their limits of error.

Solution

Given:

$$\frac{\delta R_1}{R_1} = \frac{\delta R_2}{R_2} = 1 \text{ part in } 10,000 = 0.01\%$$

$$\frac{\delta R_3}{R_3} = \frac{0.25}{8.33} \times 100\% = 0.03\%$$

$$\frac{\delta C}{C} = \frac{0.001}{1.43} \times 100\% \cong 0.07\%$$

$$\frac{\delta \omega}{\omega} = \frac{0.1}{50} \times 100\% = 0.2\%$$

Thus,

$$L = \frac{(1.43 \times 10^{-6})(1000)(16800)}{1 + (100\pi)^2(1.43 \times 10^{-6})^2(833)^2} \cong 21.1 \text{ H}$$

$$\frac{\delta L}{L} = 3\frac{\delta C}{C} + \frac{\delta R_1}{R_1} + \frac{\delta R_2}{R_2} + 2\frac{\delta R_3}{R_3} + 2\frac{\delta \omega}{\omega}$$

$$= [(3 \times 0.07) + 0.01 + 0.01 + (2 \times 0.03) + (2 \times 0.2)]\% = 0.69\%$$

Therefore,

$$L = 21 \text{ H} \pm 0.69\%$$

Similarly,

$$R = \frac{(1000)(16800)(833)(1.43 \times 10^{-6})^2(100\pi)^2}{1 + (100\pi)^2(1.43 \times 10^{-6})^2(833)^2} = 2477 \Omega$$

$$\frac{\delta R}{R} = \frac{\delta R_1}{R_1} + \frac{\delta R_2}{R_2} + 3\frac{\delta R_3}{R_3} + 4\frac{\delta C}{C} + 4\frac{\delta \omega}{\omega}$$

$$= [0.01 + 0.01 + (3 \times 0.03) + (4 \times 0.07) + (4 \times 0.2)]\% = 1.19\%$$

Therefore,

$$R = 2477 \Omega \pm 1.19\%$$

Example 3.5

A 0–10 ampere ammeter has a guaranteed accuracy of 1% of the full-scale deflection. The limiting error while reading 2.5 A is

- | | |
|--------|-----------------------|
| (a) 1% | (b) 2% |
| (c) 4% | (d) none of the above |

Solution

The full-scale deflection is 10 A. Therefore, the guaranteed accuracy is $10 \times 1\% = 0.1$ A. A 0.1 amp error in 2.5 amp amounts to

$$\frac{0.1}{2.5} \times 100 = 4\%$$

Therefore, the correct answer is (c).

Example 3.6

What is the percentage error in the measurement of kinetic energy of a body if percentage errors in the measurement of mass and speed are 2% and 3% respectively?

Solution

We know that the expression for kinetic energy is

$$E = \frac{1}{2}mv^2$$

Taking logarithm of both sides and then differentiating, we get

$$\begin{aligned} \frac{\delta E}{E} &= \frac{\delta m}{m} + 2\frac{\delta v}{v} \\ &= 0.02 + 2(0.03) = 0.08 \end{aligned}$$

Thus, the required error is 8%.

Method of Equal Effects

The most general method of finding what is called the *maximum possible error* (or *absolute error*) is as follows. If

$$X = f(p, q, r, \dots)$$

where p , q , r are individual measurands, then the maximum possible error ΔX of the quantity X is given by

$$\Delta X = \left| \frac{\partial f}{\partial p} \delta p \right| + \left| \frac{\partial f}{\partial q} \delta q \right| + \left| \frac{\partial f}{\partial r} \delta r \right| + \dots \quad (3.2)$$

where δp , δq , δr are measurement errors of variables p , q , r .

3.3 Statistical Treatment

As pointed out earlier, one has to resort to statistical methods to estimate random errors even for only one measurand. Statistical treatments are carried out on two kinds of samples—multi-sample and single-sample. By multi-sample test it is meant that the data have been acquired by different instruments, different observers and different methods, while the single-sample test signifies that instrument, observer and method remaining the same, the data have been acquired at different times.

Without going into details of the statistical methods and their mathematical backgrounds we consider only those aspects which are of importance to us in processing of data.

Suppose, we have measured the heights of 200 students of a college with the height of each student recorded in inches. A student's height may be any value such as 62.35 inches. It does not make sense to figure out how many students have heights of 62.35 inches. It may so happen that we may not find another student exactly 62.35 inches tall. It is better if we divide heights within a few 'class'es or 'cell's such as 56–58 inches, 58–61 inches and so on. The cell demarcation is arbitrary, but we care to see that each cell possesses a midpoint which is a convenient whole number. Let the grouped data so obtained be graphed as given in Fig. 3.1.

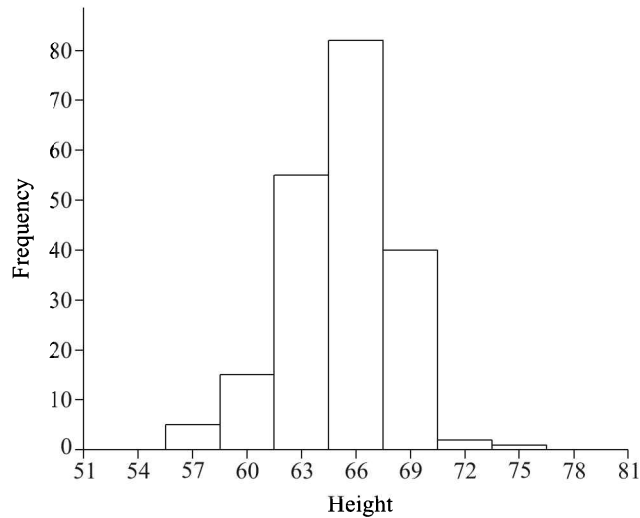


Fig. 3.1 Frequency distribution of the height of 200 students.

This graph is called the *frequency distribution* or *histogram*. Next, the question arises as to how we can characterise the frequency distribution of a sample with a single descriptive measure, or simple *statistic*. In fact, there are two highly useful descriptions: one is the *central point* of the distribution and the other is the *spread*.

Measures of Central Tendency

Statistically speaking, a central point or average is a value which is typical or representative of a set of data. Since the average tends to lie centrally within a set of data arranged according to magnitude, it is also called the central point or a *measure of central tendency*.

Generally six types of averages are defined. They are:

1. Mode
2. Median
3. Arithmetic mean or simply, mean
4. Geometric mean
5. Harmonic mean, and
6. Root mean square

Mode

Mode is defined as the most frequent value. In the frequency distribution of heights of students (Fig. 3.1), the mode value is 66 inches.

The mode may not exist, or even if it does it may not be unique.

Example 3.7

What are the mode values for the following three sets?

- (a) 2, 3, 5, 6, 9, 10, 10, 10, 11, 12, 12, 14, 18
- (b) 3, 5, 6, 9, 10, 11, 12, 14, 18
- (c) 2, 3, 5, 5, 5, 6, 9, 10, 11, 11, 11, 12, 12, 14, 18

Solution

The mode of set (a) is 10. Set (b) has no mode. Set (c) has two modes, 5 and 11.

Median

The median is the value below which half the values in the sample fall. So, if the number of data N , arranged according to magnitude, is odd, it is the value corresponding to the $(N \div 2 + 0.5)$ th data. If N is even, the median is represented by the average of the $(N \div 2)$ th and $(N \div 2 + 1)$ th points.

Example 3.8

Find the medians of the given sets of data:

- (a) 2, 3, 4, 4, 6, 7, 7, 7, 9
- (b) 3, 3, 7, 8, 12, 13, 16, 19

Solution

- (a) The number of data points is 9 which is odd. So, the $(9 \div 2) + 0.5 = 5$ th point, i.e. 6, is the median.
- (b) Here the number of data points, 8, is even. So, the median is the average of the $8 \div 2 = 4$ th and 5th points, i.e. $(8 + 12) \div 2 = 10$.

Arithmetic mean

Of all these types of averages, the arithmetic mean μ is considered the most probable value of the measurand and is defined as

$$\mu = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

where x_i 's are individual data and n is the number of measurements.

If the data x_1, x_2, \dots, x_k occur f_1, f_2, \dots, f_k times respectively, (i.e. occur with frequencies f_1, f_2, \dots, f_k), the arithmetic mean is

$$\mu = \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{f_1 + f_2 + \dots + f_k} = \frac{1}{n} \sum_{i=1}^n f_i x_i$$

Example 3.9

If 10, 16, 12 and 4 occur with frequencies 5, 3, 4 and 2 respectively then what is the arithmetic mean?

Solution

The arithmetic mean is

$$\mu = \frac{(5)(10) + (3)(16) + (4)(12) + (2)(4)}{5 + 3 + 4 + 2} = 11$$

Geometric mean

The geometric mean g_m of a set of n numbers $x_1, x_2, x_3, \dots, x_n$ is the n th root of the product of number. Written mathematically

Example 3.10

$$g_m = \sqrt[n]{x_1 x_2 x_3 \dots x_n}$$

The mass of a substance is being measured in a faulty common balance having unequal arm lengths. Show that the true mass of the substance is the geometric mean of the masses determined by placing the substance once on the left pan and next time on the right pan of the balance.

Solution

Let the true mass of the substance be m and the lengths of the left and right arms of the balance be x_1 and x_2 respectively. Initially, the substance is placed on the left pan and a mass m_1 on the right pan balances it. Then

$$m x_1 = m_1 x_2$$

or

$$m = m_1 \frac{x_2}{x_1}$$

Next, the mass is placed on the right pan and a mass m_2 on the left pan balances it. Then

$$m x_2 = m_2 x_1$$

or

$$x_2 = \frac{m_2}{m} x_1$$

Substituting this value of x_2 in the first equation, we get

$$m = m_1 \frac{x_2}{x_1} = \frac{m_1}{x_1} \cdot \frac{m_2}{m} x_1$$

or

$$m^2 = m_1 m_2$$

or

$$m = \sqrt{m_1 m_2}$$

Harmonic mean

The harmonic mean h_m of a set of n numbers $x_1, x_2, x_3, \dots, x_n$ is the reciprocal of the arithmetic mean of the reciprocals of number. Written mathematically,

$$h_m = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Example 3.11

A person travels from X to Y at an average speed of 60 km/h and returns by the same route at an average speed of 50 km/h. Find the average speed for the round trip.

Solution

Let the distance between the two places be x km. Then, if t_1 and t_2 be the time (in h) taken for the onward and return trips, we have

$$t_1 = \frac{x}{60}$$

$$t_2 = \frac{x}{50}$$

Therefore, the average speed v_{av} for the round trip is

$$v_{av} = \frac{\text{total distance}}{\text{total time}} = \frac{2x}{t_1 + t_2} = \frac{2x}{\frac{x}{60} + \frac{x}{50}} = \frac{2}{\frac{1}{60} + \frac{1}{50}}$$

We observe that v_{av} is nothing but the harmonic mean of the two speeds.

Root mean square

The root mean square (rms) or quadratic mean of a set of n numbers $x_1, x_2, x_3, \dots, x_n$ is defined as

$$\text{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

This averaging is quite common in engineering and physical applications.

Measures of Spread

The average height, in our earlier example, may be the most important characteristic (statistic) of the students. However, it is also equally important to know how the observations spread out or disperse around the average value. Like the measures of central tendency, there are several measures of spread or dispersion:

1. Deviation
2. Mean absolute deviation
3. Variance, and
4. Standard deviation

Deviation

Usually written as d_i , deviation is the scatter of an individual datum from the mean. Symbolically,

$$d_i = x_i - \mu$$

Obviously,

$$\sum d_i = 0$$

Mean absolute deviation

The mean absolute deviation¹ D of a set of data is defined as the average of absolute values of deviations, i.e.

$$D = \frac{1}{n} \sum_{i=1}^n |d_i| = \frac{1}{n} \sum_{i=1}^n |x_i - \mu|$$

If x_1, x_2, \dots, x_n occur with frequencies f_1, f_2, \dots, f_n respectively, then

$$D = \frac{1}{n} \sum_{i=1}^n f_i |x_i - \mu|$$

Example 3.12

Find the mean absolute deviation of heights of 100 male students of a class as given in table below.

| | | | | | |
|-----------------|---------|---------|---------|---------|---------|
| Height (in) | 60 – 62 | 63 – 65 | 66 – 68 | 69 – 71 | 72 – 74 |
| No. of students | 5 | 18 | 42 | 27 | 8 |

Solution

Here, the arithmetic mean

$$\begin{aligned} \mu &= \frac{(61 \times 5) + (64 \times 18) + (67 \times 42) + (70 \times 27) + (73 \times 8)}{5 + 18 + 42 + 27 + 8} \\ &= 67.45 \text{ in} \end{aligned}$$

The rest of the calculation is presented in the following table:

| Height (in) | $ x_i - \mu = x_i - 67.45 $ | f_i | $f_i x_i - \mu $ |
|-------------|-------------------------------|------------|-------------------|
| 60–62 | 6.45 | 5 | 32.25 |
| 63–65 | 3.45 | 18 | 62.10 |
| 66–68 | 0.45 | 42 | 18.90 |
| 69–71 | 2.55 | 27 | 68.85 |
| 72–74 | 5.55 | 8 | 44.40 |
| $\Sigma =$ | | 100 | 226.50 |

Therefore, the mean absolute deviation $D = \frac{226.50}{100} = 2.26$ in

Variance

Intuitively, the mean absolute deviation is a good measure of spread; but it is mathematically intractable. One difficulty is the problem of differentiating an absolute value function. To

¹Often referred to as *average deviation*.

obviate this difficulty, the *variance*, which is nothing but the mean squared deviation, was defined as

$$\text{Variance} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

The statistical theories distinguish between a *sample* and a *population* of data. The sample denotes that the number of data is less than or equal to 20 while the population indicates the size more than 20. According to the statistical theories, though the variance or MSD is a good measure of dispersion for a population, the divisor should be $(n - 1)$ rather than n to make it an *unbiased estimator* for a sample.

Standard deviation

Denoted by σ or s , standard deviation is defined in two ways—one for a sample ($n \leq 20$) and the other for a population ($n > 20$) as follows

$$\sigma = \sqrt{\frac{\sum d_i^2}{n}} \quad n > 20$$

$$s = \sqrt{\frac{\sum d_i^2}{n - 1}} \quad n \leq 20$$

Variance which is just the squared standard deviation is thus written as σ^2 or s^2 .

Example 3.13

A set of 10 independent measurements were made to determine the diameter of the bob of a simple pendulum. The measured values in cm were: 1.570, 1.597, 1.591, 1.562, 1.577, 1.580, 1.564, 1.586, 1.550 and 1.575. Determine (a) the arithmetic mean, (b) the average deviation, (c) the standard deviation, and (d) the variance.

Solution

The calculation is presented in a tabular form below with the last row in bold face letters indicating sums of corresponding columns:

| x_i | $ d $ | d^2 |
|---------------|--------------|------------------|
| 1.570 | 0.005 | 0.000025 |
| 1.597 | 0.022 | 0.000484 |
| 1.591 | 0.016 | 0.000256 |
| 1.562 | 0.013 | 0.000169 |
| 1.577 | 0.002 | 0.000004 |
| 1.580 | 0.005 | 0.000025 |
| 1.564 | 0.011 | 0.000121 |
| 1.586 | 0.011 | 0.000121 |
| 1.550 | 0.025 | 0.000625 |
| 1.575 | 0.000 | 0.000000 |
| 15.752 | 0.110 | 0.001866 |
| $\mu = 1.575$ | $D = 0.011$ | $s = 0.0143$ |
| | | $s^2 = 0.000204$ |

The measurement can thus be reported as 1.575 ± 0.014 cm where the indicated error is the standard deviation.

3.4 Error Estimates from the Normal (or Gaussian) Distribution

If a rather large number of careful measurements are carried out of a measurand, and the frequency of occurrence of a particular value is plotted against the corresponding values, the resulting histogram usually assumes the form of a bell-shaped curve called the *normal* or *Gaussian*² curve as shown in Fig. 3.2. Here we have plotted the Gaussian distribution curve for a typical height measurement of students of a class as given in Example 3.12 at page 39.

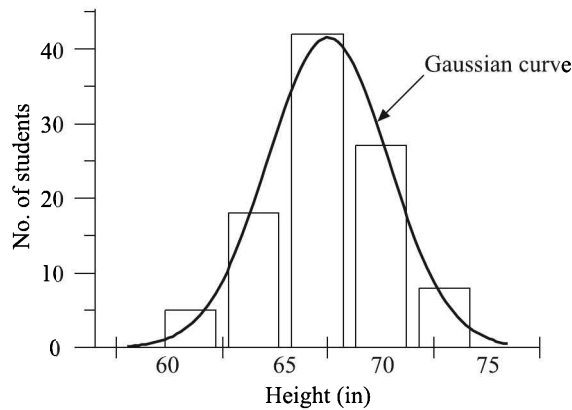


Fig. 3.2 Height vs. number of students of a class showing a Gaussian distribution.

Errors that are made in physical measurements do often have a normal distribution. This distribution is defined by the equation

$$y = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \right\} \quad (3.3)$$

where μ is the mean, σ is the standard deviation, and $\pi \cong 3.14159$.

It is easily seen from the following calculation that the total area bounded by the normal distribution curve and the abscissa is 1.

$$\begin{aligned} \int_{-\infty}^{+\infty} y dx &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp \left\{ -\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \right\} dx \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \exp(-u^2) du \quad \text{where } u = \frac{x - \mu}{\sigma\sqrt{2}} \\ &= \frac{\Gamma(1/2)}{\sqrt{\pi}} = 1 \end{aligned}$$

²Named after (Johann) Karl Friedrich Gauss (1777–1855), German mathematician. His contribution to electrostatics gave birth to the electromagnetic field theory apart from his seminal contributions to probability theory and other branches of mathematics.

Hence the area under the curve between $x = a$ and $x = b$, where, $a < b$, represents the probability that x lies between a and b and, therefore, can be written as $\Pr\{a < x < b\}$.

We note that in the very special case when $\mu = 0$ and $\sigma = 1$, the normal distribution of Eq. (3.3) reduces to

$$y = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \quad (3.4)$$

But more important, regardless of what μ and σ may be, the Eq. (3.3) can be translated to the form of Eq. (3.4) by defining $z = (x - \mu)/\sigma$, when it becomes

$$y = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

This is called the *standard form* of the normal distribution. A plot of the standard normal distribution is shown in Fig. 3.3.

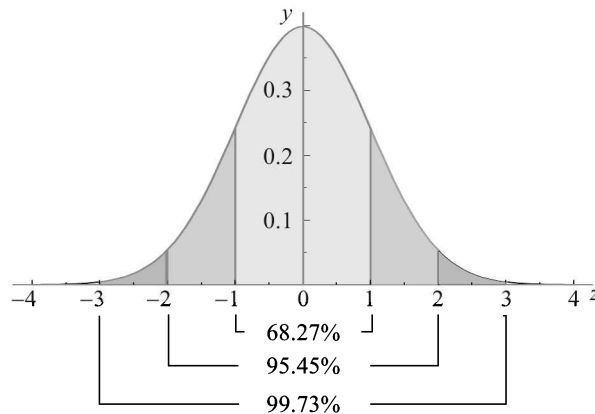


Fig. 3.3 Standard normal curve (Figures, e.g. 68.27%, show the probability within the indicated limits).

The probability of occurrence of a particular value between limits z_1 and z_2 given by the standard normal distribution is

$$\Pr\{z_1 \leq z \leq z_2\} = \int_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \quad (3.5)$$

Areas under this curve bounded by the ordinates at $z = 0$ and at any positive value of z are given in tabular form in Table 1, Appendix D at page 865. From such a table the area between any two z -values can be found by using the symmetry property of the curve, namely, $f(z) = f(-z)$. Thus, the probability of occurrence of a particular value within a certain limit can be found out from this table.

From the graph of this standardised (mean 0 and variance 1) normal curve as shown in Fig. 3.3 we observe that

| <i>Interval in z-parameter</i> | <i>Area under the interval</i> | <i>Standard deviation limits</i> | <i>Probability of lying a particular value within the limits</i> |
|------------------------------------|------------------------------------|--------------------------------------|--|
| $-1 \leq z \leq 1$ | 68.27% | $\pm \sigma$ | 68.27% |
| $-2 \leq z \leq 2$ | 95.45% | $\pm 2\sigma$ | 95.45% |
| $-3 \leq z \leq 3$ | 99.73% | $\pm 3\sigma$ | 99.73% |

Properties of the normal distribution. To sum up, the properties of the normal distribution are as follows:

1. The normal curve is symmetrical about the mean μ .
2. The mean is at the middle and it divides the area into halves.
3. The total area under the curve is equal to 1.
4. The curve is completely determined by its mean and standard deviation σ (or variance σ^2).
5. The probability of a continuous normal variable z found in a particular interval $[a, b]$ is the area under the curve bounded by $z = z_1$ and $z = z_2$ and is given by

$$\Pr\{z_1 \leq z \leq z_2\} = \int_{z_1}^{z_2} y dz$$

where y is the standard normal distribution and the area depends upon the values of μ and σ .

Example 3.14

Bolts produced in a factory have the specification 10 ± 0.2 cm where the specified error is the standard deviation. If one bolt is picked up at random from a heap, what is the probability that its length will lie between 9.9 and 10.1 cm?

Solution

Here the probability may be expressed as $\Pr\{9.9 \leq x \leq 10.1\}$. This may be written in the standardised form as

$$\Pr\left\{\frac{9.9 - 10}{0.2} \leq \frac{x - 10}{0.2} \leq \frac{10.1 - 10}{0.2}\right\} = \Pr\{-0.5 \leq z \leq 0.5\}$$

$$= 0.383 \quad (\text{the table lists } 0.1915 \text{ for } z = 0.5).$$

Thus, the probability is 38.3%.

Example 3.15

A mass of 10 kg is measured with an instrument and the readings are normally distributed with respect to the mean of 10 kg. Given that

$$\frac{2}{\sqrt{2\pi}} \int_0^{0.84} \exp\left(\frac{-\eta^2}{2}\right) d\eta = 0.6$$

and that 60 per cent of the recordings are found to be within 0.05 kg from the mean, the standard deviation of the data is

- (a) 0.02 (b) 0.04
 (c) 0.06 (d) 0.08

Solution

Given: $\Pr\{0 \leq \eta \leq 0.84\} = 0.6 = 60\%$, and for $(x - \mu) = 0.05$ Probability = 60%

By definition,

$$\frac{x - \mu}{\sigma} = \eta$$

So,

$$\frac{0.05}{\sigma} = 0.84$$

which gives

$$\sigma = 0.06$$

Therefore, the answer is (c).

Example 3.16

The average life of a certain type of instrument is 10 years, with a standard deviation of 2 years. If the manufacturer is willing to replace only 3% of the instruments that fail, how long a warranty should he offer, assuming that the lives of the instruments follow a normal distribution?

Solution

Here, $\mu = 10$ and $\sigma = 2$. The corresponding normal distribution curve is given in Fig. 3.4.

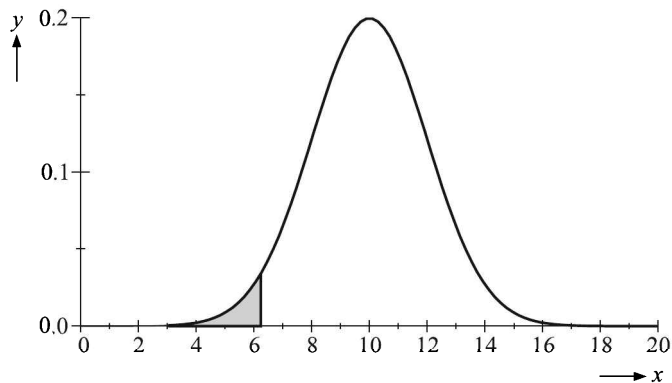


Fig. 3.4 Standard normal curve for $\sigma =$ and $\mu = 10$ (Example 3.16). The warranty is provided for the years that lie within the shaded area.

Let X be the life of an instrument and x be the warranty period. We need to find the value (in years) that will give us the bottom 3% of the distribution, i.e. the shaded area in Fig. 3.4. These are the instruments that the manufacturer is willing to replace under the warranty. Stated mathematically,

$$\Pr\{X < x\} = 0.03$$

The area of the lower half of the curve is 0.5, the total area being 1. So, to find the z value from the table, we need to know the area of the remaining portion which is $(0.5 - 0.03) = 0.47$. The corresponding nearest value, 0.4699, in the table is $z = -1.88$.

Since $z = \frac{x - \mu}{\sigma}$, we can write:

$$\frac{x - 10}{2} = -1.88$$

$$\Rightarrow x = 10 - (2)(1.88) = 6.24$$

So the warranty period should be 6.24 years.

Probable Error

Gaussian distribution helps us estimate the probable error of any measurement. The probable error is defined as the 50% confidence limit of the population parameters corresponding to a measured value. Let us elaborate it.

Suppose, we have made n measurements of a quantity and the resultant standardised deviation is $\pm\zeta$. For what value of ζ will there be a 50% chance that any value chosen at random will lie between $-\zeta$ and $+\zeta$? In other words, for what value of the limits $-\zeta$ and ζ will the area under the standard normal curve be 0.5? Stated mathematically, for what ζ

$$\frac{1}{\sqrt{2\pi}} \int_{-\zeta}^{+\zeta} \exp\left(-\frac{1}{2}z^2\right) dz = \frac{1}{2} \quad (3.6)$$

Solving Eq. (3.6) either numerically by the method of iteration, or by interpolating values in the standard table, one can find that

$$\zeta = \pm 0.6745$$

It means that the corresponding deviation r which is called the *probable error* is

$$r = x - \mu = \pm 0.6745\sigma \quad (3.7)$$

Two more quantities—probable error of the mean and standard deviation of the mean—are of interest to us. They are defined now.

Probable error of the mean

Written as r_m , the probable error the mean is given by

$$r_m = \frac{r}{\sqrt{n}} = \frac{0.6745\sigma}{\sqrt{n}}$$

Standard deviation of the mean

Denoted by σ_m , the standard deviation of the mean is given by

$$\sigma_m = \frac{\sigma}{\sqrt{n}}$$

Probable error of combinations

If the final result X is generated from individual measurands p, q, r, \dots such that

$$X = f(p, q, r, \dots)$$

then the probable error of the combination, r_X , is given by³

$$r_X = \sqrt{\left(\frac{\partial X}{\partial p}\right)^2 r_p^2 + \left(\frac{\partial X}{\partial q}\right)^2 r_q^2 + \left(\frac{\partial X}{\partial r}\right)^2 r_r^2 + \dots} \quad (3.8)$$

Equation (3.8) is called the *root-sum square* (rss) *formula*.

We have given the relation for the probable error here. Since the probable error is nothing else than the scaled standard deviation, scaling factor being 0.6745, a similar relation can be written for the combination of standard deviations.

Example 3.17

If a physical quantity F is expressed in terms of three measurands x, y and z as $F = xy/z$ and if the limiting and random probable errors of measurands are $(\delta x, \delta y, \delta z)$ and $(\Delta x, \Delta y, \Delta z)$ respectively, find the expressions for

- (a) the resultant limiting error in F , and
- (b) the resultant random probable error in F

Solution

(a) Given,

$$F = \frac{xy}{z}$$

Taking logarithms of both sides, we get

$$\ln F = \ln x + \ln y - \ln z$$

Taking differentials, we get

$$\frac{dF}{F} = \frac{dx}{x} + \frac{dy}{y} - \frac{dz}{z}$$

Since errors are expressed as $\pm \delta x \dots$ etc, the resultant limiting error is given by

$$\frac{\delta F}{F} = \pm \left(\frac{\delta x}{x} + \frac{\delta y}{y} + \frac{\delta z}{z} \right)$$

(b) In this case,

$$\frac{\partial F}{\partial x} = \frac{y}{z}, \quad \frac{\partial F}{\partial y} = \frac{x}{z}, \quad \frac{\partial F}{\partial z} = -\frac{xy}{z^2}$$

Hence the resultant random probable error is

$$r_F = \sqrt{\frac{y^2}{z^2}(\Delta x)^2 + \frac{x^2}{z^2}(\Delta y)^2 + \frac{x^2 y^2}{z^4}(\Delta z)^2}$$

³see Appendix A at page 857 for a derivation.

Example 3.18

In a parallel circuit the current in one branch, I_1 , is (100 ± 2) A and in the other, I_2 , is (200 ± 5) A. Determine the total current considering errors as (a) limiting error, and (b) probable error.

Solution

(a) Total current, $I = I_1 + I_2$. Therefore,

$$\begin{aligned}\delta I &= \pm(\delta I_1 + \delta I_2) \\ &= \pm(2 + 5) = \pm 7 \text{ A}\end{aligned}$$

So,

$$I = 300 \pm 7 \text{ A}$$

(b) If the errors are probable errors, we have

$$r_I = \sqrt{\left(\frac{\partial I}{\partial I_1}\right)^2 r_{I_1}^2 + \left(\frac{\partial I}{\partial I_2}\right)^2 r_{I_2}^2}$$

Now, $\frac{\partial I}{\partial I_1} = \frac{\partial I}{\partial I_2} = 1$ and $r_{I_1} = \pm 2$, $r_{I_2} = \pm 5$

$$r_I = \sqrt{4 + 25} \cong \pm 5.4 \text{ A}$$

Therefore,

$$I = 300 \pm 5.4 \text{ A}$$

Note: The probable error gives a more optimistic estimate of the error than the limiting error.

Example 3.19

A capacitor of value (1.0 ± 0.1) μF is charged to a voltage of (20 ± 1) V, where the errors are standard deviations. Find the charge on the capacitor and its standard deviation.

Solution

If Q is the charge,

$$Q = CV = 1 \times 20 = 20 \mu\text{C}$$

$\therefore \frac{\partial Q}{\partial C} = V = 20 \text{ V}$ $\frac{\partial Q}{\partial V} = C = (1.0 \times 10^{-6}) \text{ F}$

$$\sigma_C = (\pm 0.1 \times 10^{-6}) \text{ F} \quad \sigma_V = \pm 1 \text{ V}$$

Therefore,

$$\begin{aligned}\sigma_Q &= \sqrt{\left(\frac{\partial Q}{\partial C}\right)^2 \sigma_C^2 + \left(\frac{\partial Q}{\partial V}\right)^2 \sigma_V^2} \\ &= \sqrt{(20^2)(0.1 \times 10^{-6})^2 + (1.0 \times 10^{-6})^2(1^2)} \\ &= \sqrt{5} \times 10^{-6} \cong 2.2 \mu\text{C}\end{aligned}$$

Hence,

$$Q = 20 \pm 2.2 \mu\text{C}$$

Example 3.20

In a circuit the value of the current I is $10 \text{ A} \pm 1\%$ and the voltage V across a resistor R of value $10 \Omega \pm 0.1\%$ is $100 \text{ V} \pm 1\%$, where the indicated errors are probable errors. The power consumed by the circuit can be calculated from the relations (a) $P = V^2/R$, and (b) $P = VI$. Calculate the probable errors in both the methods.

Solution

(a) Here,

$$\frac{\partial P}{\partial V} = \frac{2V}{R} = \frac{2 \times 100}{10} = 20, \quad \frac{\partial P}{\partial R} = -\frac{V^2}{R^2} = -\frac{(100)^2}{(10)^2} = -100$$

$$r_V = \pm 1 \quad r_R = \pm 0.01$$

From these we get,

$$r_P = \sqrt{\left(\frac{\partial P}{\partial V}\right)^2 r_V^2 + \left(\frac{\partial P}{\partial R}\right)^2 r_R^2}$$

$$= \sqrt{20^2 \times 1^2 + 100^2 \times 0.01^2}$$

$$= \sqrt{401} \cong \pm 20.02 \text{ W}$$

Hence

$$P = 1000 \text{ W} \pm 2.002\%$$

(b) In this case,

$$\frac{\partial P}{\partial V} = I = 10, \quad \frac{\partial P}{\partial I} = V = 100, \quad r_V = \pm 1, \quad \text{and} \quad r_I = \pm 0.1$$

Therefore,

$$r_P = \sqrt{\left(\frac{\partial P}{\partial V}\right)^2 r_V^2 + \left(\frac{\partial P}{\partial I}\right)^2 r_I^2}$$

$$= \sqrt{10^2 \times 1^2 + 100^2 \times 0.1^2} = \sqrt{200} \cong 14.14 \text{ W}$$

Thus

$$P = 1000 \text{ W} \pm 1.414\%$$

Note: This example shows how to choose parameters of a quantity so that the error in its measurement is a minimum.

Example 3.21

The diameter of a large bore is measured by using pin gauge of length L and the amount of rock at the end of the pin, W . The diameter D of the bore is calculated using the expression

$$D = L + \frac{W^2}{8L}$$

Repeated measurements of L gave a mean value of 500 mm with standard deviation σ_L of 0.02 mm and for rock W a mean value of 60 mm with standard deviation σ_W of 0.5 mm. Determine

- (a) The maximum possible error (absolute error) in the determination of D , based on 3.29σ limits.
- (b) Root-sum square error in the determination of D based on 3σ limits.

Solution

(a) At 3.29σ limit,

$$\delta L = 0.02 \times 3.29 = 0.0658 \text{ mm} \quad \delta W = 0.5 \times 3.29 = 1.645 \text{ mm}$$

From Eq. (3.2), we get

$$\begin{aligned} \Delta D &= \left| \frac{\partial D}{\partial L} \delta L \right| + \left| \frac{\partial D}{\partial W} \delta W \right| \\ &= \left(1 - \frac{W^2}{8L^2} \right) \delta L + \left(\frac{W}{4L} \right) \delta W \\ &= \left[1 - \frac{60^2}{8(500)^2} \right] (0.0658) + \left[\frac{60}{4(500)} \right] (1.645) \\ &= 0.115 \text{ mm} \end{aligned}$$

(b) At 3σ limit,

$$r_L = 0.02 \times 3 = 0.06 \text{ mm} \quad r_W = 0.5 \times 3 = 1.5 \text{ mm}$$

From Eq. (3.8),

$$\begin{aligned} r_D &= \sqrt{\left(\frac{\partial D}{\partial L} r_L \right)^2 + \left(\frac{\partial D}{\partial W} r_W \right)^2} \\ &= \sqrt{\left(1 - \frac{W^2}{8L^2} \right)^2 r_L^2 + \left(\frac{W}{4L} \right)^2 r_W^2} \\ &= \sqrt{\left(1 - \frac{60^2}{8 \times 500^2} \right)^2 (0.06)^2 + \left(\frac{60}{4 \times 500} \right)^2 (1.5)^2} \\ &= 0.075 \text{ mm} \end{aligned}$$

Note: The instrument makers often mention a quantity called *tolerance* of the instrument. It is stated either as an absolute amount such as $\pm 0.5^\circ\text{C}$ or as a percentage such as $\pm 1\%$. This tolerance may either be the limiting error or the standard deviation.

3.5 Chi-Square Test

While calculating the probable error by Eq. (3.7), we tacitly assumed that our data closely followed a normal or Gaussian distribution. Although in most physical measurements data do follow such a distribution, it is somewhat better to check quantitatively that the distribution is close to Gaussian. Chi-square (χ^2) goodness-of-fit test helps us to do that.

χ^2 is a measure of discrepancy existing between the observed and the expected frequencies in a given range (denoted by o and e respectively) and is defined as

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \dots + \frac{(o_n - e_n)^2}{e_n} = \sum_{i=1}^n \frac{(o_i - e_i)^2}{e_i} \quad (3.9)$$

If $\chi^2 = 0$, the observed and the expected frequencies agree exactly, and therefore, the data exactly obey a Gaussian distribution, while if $\chi^2 > 0$, there is some discrepancy. The larger the value of χ^2 , the greater is the discrepancy.

In practice, the computed value of χ^2 , given by Eq. (3.9) is compared with some critical value, such as $\chi_{.95}^2$ or $\chi_{.99}^2$, which correspond to 95% and 99% confidence limits. These values are available in tabular form in which values are listed for different confidence limits against number of degrees of freedom ν , which is defined as

$$\nu = i - 1 - m, \quad (3.10)$$

where i is the total number of groups and m is the number of population parameters.

If in the above comparison, the calculated value were less than that listed in the table, we would conclude that the observed frequencies do not differ significantly from the expected ones and would accept that the data are Gaussian at the corresponding level of significance.

We should also look at the situation where χ^2 is near zero, such as 0.05 or 0.01. If the computed value of χ^2 is less than the listed value for $\chi_{.05}^2$ or $\chi_{.01}^2$, we may conclude that the agreement is *too good*. The following example will make the procedure clear.

Example 3.22

Test the goodness-of-fit of the data given below

| Height (m) | 1.52–1.56 | 1.57–1.61 | 1.62–1.66 | 1.67–1.71 | 1.72–1.76 |
|-----------------|-----------|-----------|-----------|-----------|-----------|
| No. of students | 5 | 18 | 42 | 27 | 8 |

Solution

The work may be organised as in Table 3.1.

$$\mu = \frac{(1.54)(5) + (1.59)(18) + (1.64)(42) + (1.69)(27) + (1.74)(8)}{100} = 1.6475 \text{ m}$$

$$\sigma = \sqrt{\frac{5(1.54 - 1.6475)^2 + 18(1.59 - 1.6475)^2 + 42(1.64 - 1.6475)^2 + 27(1.69 - 1.6475)^2 + 8(1.74 - 1.6475)^2}{100}}$$

$$= \sqrt{\frac{27(1.69 - 1.6475)^2 + 8(1.74 - 1.6475)^2}{100}} = 0.04867$$

Note: In Table 3.1:

1. z for class-boundaries have been calculated from $z = (x - \mu)/s$, where μ and s have been calculated by the standard procedure.
2. Areas under normal curve have been obtained using standard table.

Table 3.1 Example 3.22

| Heights (m) | Class boundaries x | z for class boundaries | Area under the normal curve from 0 to z | Area for each class | Expected frequency e | Observed frequency o |
|-------------|----------------------|--------------------------|---|---------------------|------------------------|------------------------|
| 1.52–1.56 | 1.515 | −2.72 | 0.4967 | | | |
| | | | | 0.0413 | 4.13 or 4 | 5 |
| 1.57–1.61 | 1.565 | −1.70 | 0.4554 | 0.2068 | 20.68 or 21 | 18 |
| 1.62–1.66 | 1.615 | −0.67 | 0.2486 | *0.3892 | 38.92 or 39 | 42 |
| | 1.665 | 0.36 | 0.1406 | | | |
| 1.67–1.71 | 1.665 | 0.36 | 0.1406 | 0.2771 | 27.71 or 28 | 27 |
| 1.72–1.76 | 1.715 | 1.39 | 0.4177 | | | |
| | | | | 0.0743 | 7.43 or 7 | 8 |
| | 1.765 | 2.41 | 0.4920 | | | |

* Obtained by adding with the area corresponding to previous z value, because there is a change of sign.

3. The area for each class has been calculated by subtraction from the area under the preceding z , except in one case where it has been added because z has changed sign there.
4. Expected frequency has been obtained by multiplying the area for each class by the total frequency 100.

Now,

$$\chi^2 = \frac{(5 - 4)^2}{4} + \frac{(18 - 21)^2}{21} + \frac{(42 - 39)^2}{39} + \frac{(27 - 28)^2}{28} + \frac{(8 - 7)^2}{7} = 1.08$$

Here, $i = 5$ and $m = 2$ (namely, μ and σ); therefore, $\nu = 5 - 1 - 2 = 2$.

From the standard table⁴ for *Percentile Values for the Chi-Square Distribution* we get, for $\nu = 2$, $\chi^2_{.95} = 5.99$. Since the value we obtained is much lower than this, we conclude that the fit of the data is *very good*. Now, for $\nu = 2$, $\chi^2_{.05} = 0.103$. Since our value is greater than this value, we may say that the fit is not *too good*.

3.6 Curve Fitting Methods

More often than not the data show a scatter and we draw an average curve by what we call the *eye estimation*. But there are a few methods of drawing the most probable curve provided, of course, we know the relation between the parameters. We discuss three methods, namely

1. Method of sequential differences
2. Method of extended differences, and
3. Method of least squares

for a linear relation such as, $y = a_0 + a_1x$.

⁴See Table 2, Appendix D at page 866.

Method of Sequential Differences

If there are n pairs of data, such as $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, then according to the method of sequential differences, $(n - 1)$ slopes are found from adjacent points as

$$\begin{aligned} a_1 &= \frac{y_2 - y_1}{x_2 - x_1} \\ a_2 &= \frac{y_3 - y_2}{x_3 - x_2} \\ &\vdots \\ a_{n-1} &= \frac{y_n - y_{n-1}}{x_n - x_{n-1}} \end{aligned}$$

The mean of the slopes is given by

$$\bar{a} = \frac{1}{n-1} \sum_{i=1}^{n-1} a_i$$

Next, the coordinates of the centroid of data points is obtained from the relations

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Now, the intercept a_0 is figured out from the relation

$$a_0 = \bar{y} - \bar{a}\bar{x}$$

Example 3.23

Find the most probable straight line for the following data by the method of sequential differences:

| | | | | | | |
|-----|---|---|---|---|----|----|
| x | 1 | 3 | 4 | 6 | 11 | 14 |
| y | 1 | 2 | 4 | 5 | 8 | 10 |

Solution

The calculation is shown in a tabular form below.

| x | y | $y_n - y_{n-1}$ | $x_n - x_{n-1}$ | a |
|-----------------|---------------|-----------------|-----------------|-----------------------|
| 1 | 1 | | | |
| 3 | 2 | 1 | 2 | 0.5 |
| 4 | 4 | 2 | 1 | 2 |
| 6 | 5 | 1 | 2 | 0.5 |
| 11 | 8 | 3 | 5 | 0.6 |
| 14 | 10 | 2 | 3 | 0.667 |
| 39 | 30 | | | $\bar{a} = (4.267/5)$ |
| $\bar{x} = 6.5$ | $\bar{y} = 5$ | | | $= 0.8534$ |

The intercept is, therefore,

$$a_0 = 5 - 0.8534 \times 6.5 = -0.5471$$

The equation of the straight line fit is

$$y = -0.5471 + 0.8534x$$

Method of Extended Differences

In the method of extended differences, the entire data is divided into two equal groups. For this purpose, the number of data points should be even ($2n$). If the number is not even, one point which is suspect, is ignored. Suppose, after division of data into two groups, the following is the configuration:

$$\begin{aligned} \text{Group A (low values)} &: && (x_1, y_1), (x_2, y_2), \dots (x_n, y_n) \\ \text{Group B (high values)} &: && (x_{n+1}, y_{n+1}), (x_{n+2}, y_{n+2}), \dots (x_{2n}, y_{2n}) \end{aligned}$$

From these the slopes are calculated by using corresponding values of two groups as follows:

$$\begin{aligned} a_{n+1} &= \frac{y_{n+1} - y_1}{x_{n+1} - x_1} \\ a_{n+2} &= \frac{y_{n+2} - y_2}{x_{n+2} - x_2} \\ &\vdots \\ a_{2n} &= \frac{y_{2n} - y_n}{x_{2n} - x_n} \end{aligned}$$

As in the method of sequential differences, the mean slope is calculated from the n number of available slopes as

$$\bar{a} = \frac{1}{n} \sum_{i=n+1}^{2n} a_i$$

The rest of the procedure is identical to that of the method of sequential differences.

This method is likely to yield a better result because two averaging processes are done simultaneously, one implicitly being done by choosing to calculate slopes over long distances between data points.

Example 3.24

Find the most probable straight line for the data given in Example 3.23 by the method of extended differences.

Solution

The calculation is shown in a tabular form below.

| x | y | $y_n - y_{n-3}$ | $x_n - x_{n-3}$ | a |
|-----------------|---------------|-----------------|-----------------|----------------------------|
| 1 | 1 | | | |
| 3 | 2 | | | |
| 4 | 4 | | | |
| 6 | 5 | 4 | 5 | 0.8 |
| 11 | 8 | 6 | 8 | 0.75 |
| 14 | 10 | 6 | 10 | 0.6 |
| 39 | 30 | | | $\bar{a} = \frac{2.15}{3}$ |
| $\bar{x} = 6.5$ | $\bar{y} = 5$ | | | $= 0.7167$ |

The intercept is, therefore,

$$a_0 = 5 - 0.7167 \times 6.5 = 0.3415$$

The equation of the straight line fit is

$$y = 0.3415 + 0.7167x$$

Method of Least Squares

The method of least squares is a statistical procedure. Without going into details of their derivation, we give below the relevant working formulae for the least square fit of a straight line.

$$a_0 = \frac{\Sigma x^2 \Sigma y - \Sigma x \Sigma xy}{n \Sigma x^2 - (\Sigma x)^2} \quad (3.11)$$

$$a_1 = \frac{n \Sigma xy - \Sigma x \Sigma y}{n \Sigma x^2 - (\Sigma x)^2} \quad (3.12)$$

These relations can be obtained by solving simultaneously the following *normal equations* of the straight line equation:

$$\Sigma y = na_0 + a_1 \Sigma x \quad (3.13)$$

$$\Sigma xy = a_0 \Sigma x + a_1 \Sigma x^2 \quad (3.14)$$

Thus if a relation is given as

$$y = a_0 + a_1 x + a_2 x^2 \quad (3.15)$$

the corresponding normal equations can be set up as

$$\Sigma y = na_0 + a_1 \Sigma x + a_2 \Sigma x^2$$

$$\Sigma xy = a_0 \Sigma x + a_1 \Sigma x^2 + a_2 \Sigma x^3$$

$$\Sigma x^2 y = a_0 \Sigma x^2 + a_1 \Sigma x^3 + a_2 \Sigma x^4$$

One may note that these equations are obtained by multiplying Eq. (3.15) by 1, x and x^2 respectively and summing both sides of the resulting equations. This technique can be extended to obtain normal equations for polynomial of any order.

The least square curve fitting is also known as *regression analysis*.

Problems in polynomial fitting. However, not all types of data can be fit to polynomials. This was demonstrated by Runge⁵ when he fit data based on a simple function given by

$$y = \frac{1}{1 + 25x^2}$$

in an interval of $[-1, 1]$. Let us consider six equally spaced points in $[-1, 1]$ given by

| | | | | | | |
|---------------------|----------|------|------|-----|-----|----------|
| x | -1.0 | -0.6 | -0.2 | 0.2 | 0.6 | 1.0 |
| $y = 1/(1 + 25x^2)$ | 0.038461 | 0.1 | 0.5 | 0.5 | 0.1 | 0.038461 |

⁵Carl David Tolmé Runge (1856–1927) was a German mathematician and physicist.

The following fifth order polynomial can be fit through these points:

$$y = 0.56731 + 1.0004 \times 10^{-11}x - 1.7308x^2 - 3.3651 \times 10^{-11}x^3 + 1.2019x^4 + 3.1378 \times 10^{-11}x^5$$

Figure 3.5 shows the original curve and the polyfit curve.

It is obvious from the figure that the polynomial fitting fails here. It can be shown that a higher order polynomial will produce even a worse fitting.

Nonlinear relations can sometimes be converted to linear ones by suitable transformation of variables. A brief discussion is given later at page 56.

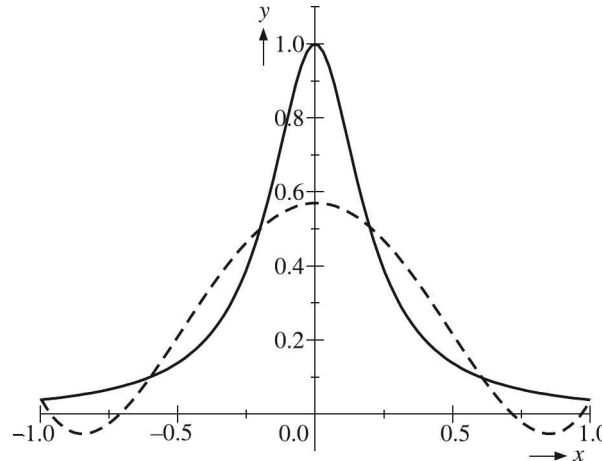


Fig. 3.5 Fifth order polynomial fitting (dashed curve) with six equally spaced points of the function $y = 1/(1 + 25x^2)$ (continuous curve).

Example 3.25

Find, by the method of least squares, the most probable straight line for the data given in Example 3.23.

Solution

The calculation is presented in the following tabular form, the last row showing the summed columns:

| x | y | x^2 | xy |
|-----------|-----------|------------|------------|
| 1 | 1 | 1 | 1 |
| 3 | 2 | 9 | 6 |
| 4 | 4 | 16 | 16 |
| 6 | 5 | 36 | 30 |
| 11 | 8 | 121 | 88 |
| 14 | 10 | 196 | 140 |
| 39 | 30 | 379 | 281 |

Substituting these values in the respective formulae, we get

$$a_0 = \frac{379 \times 30 - 39 \times 281}{6 \times 379 - 39^2} = 0.5458$$

$$a_1 = \frac{6 \times 281 - 39 \times 30}{753} = 0.6853$$

Therefore

$$y = 0.5458 + 0.6853x$$

A plot of the data and lines obtained from different methods are shown in Fig. 3.6. It may be seen that the method of extended differences and the method of least squares yield comparable results, but the result obtained from the method of sequential differences is not happy for this set of data.

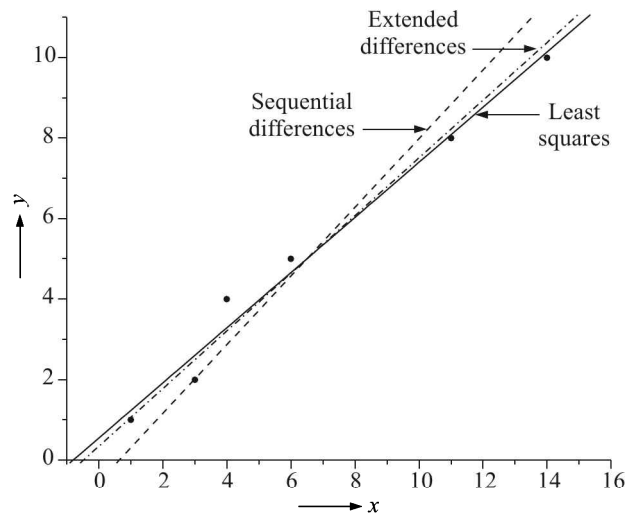


Fig. 3.6 Plot of data (●) and lines obtained from different methods: - - -, sequential differences; · - · -, extended differences; —, least squares.

Linearisation

The term linearisation refers to the following two situations:

1. Finding a linear approximation to a function at a given point or a domain in a statical system.
2. Assessing the local stability of an equilibrium point of a system of nonlinear differential equations or discrete function in dynamical systems.

The methods find use in fields such as engineering, physics, economics and ecology.

Static systems

Suppose, we need to linearise the characteristics of a transducer such as a thermistor. This can be achieved through two ways:

1. Software
2. Hardware

Software linearisation. It is possible to get data for the resistance vs. temperature of the thermocouple at discrete points such as (R_0, T_0) , (R_1, T_1) , \dots , (R_n, T_n) . The data given in Example 10.6 at page 397 are plotted in Fig. 3.7(a). Clearly, it is a nonlinear plot.

We know that in a short interval, the data can be fit to a function of the form

$$\frac{1}{T} = \frac{1}{T_0} + \frac{1}{B} \ln \left(\frac{R_T}{R_0} \right) \quad [\text{see Eq. (10.11) at page 394}]$$

If we plot the data as $1/T$ vs. $\ln R_T$, we get a linear curve as shown in Fig. 3.7(b). From this linear curve it is easier to know the temperature corresponding to a measured resistance of the thermistor.

Here is another example.

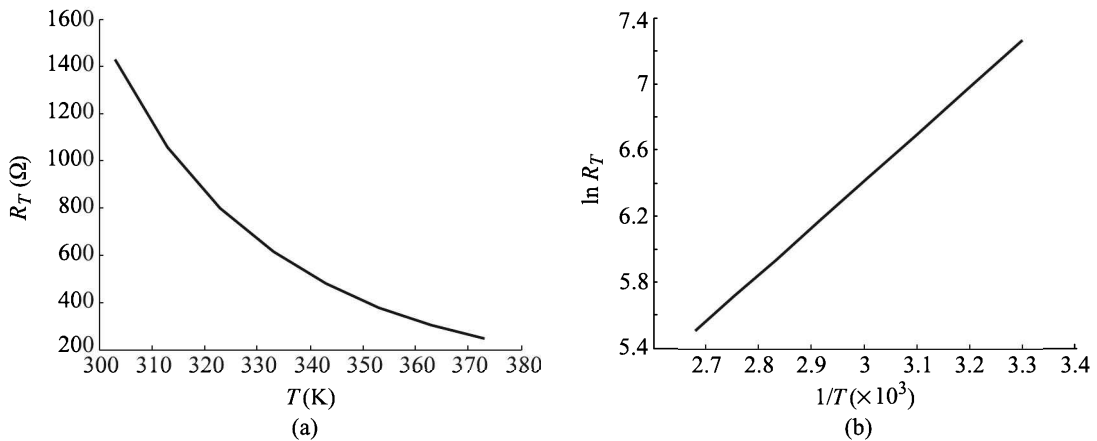


Fig. 3.7 (a) Temperature vs. resistance plot, and (b) $1/T$ vs. $\ln R_T$ plot. Data taken from Example 10.6 of page 397.

Example 3.26

The following table gives experimental values of the pressure p of a given mass of gas corresponding to various values of the volume V . Thermodynamics says that a relationship of the form $pV^\gamma = C$ should exist between the variables where γ and C are constants. Obtain a linear relation for the data.

| | | | | | | |
|---|-------|--------|--------|--------|--------|--------|
| Volume, $V(\text{cm}^3)$ | 889.8 | 1012.7 | 1186.4 | 1453.5 | 1943.5 | 3179.1 |
| Pressure, p (kg/cm^2) | 4.30 | 3.48 | 2.64 | 2.00 | 1.35 | 0.71 |

Solution

Taking logarithm of both sides, the given thermodynamic relation can be re-written as

$$\ln p + \gamma \ln V = \ln C$$

or

$$\ln p = \ln C - \gamma \ln V$$

or

$$y = a_0 + a_1 x$$

where, $\ln V = x$, $\ln p = y$, $a_0 = \ln C$ and $a_1 = -\gamma$. The data for solving the normal equations [Eqs. (3.13) and (3.14)] can be calculated as follows:

| $x = \ln V$ | $y = \ln p$ | x^2 | xy |
|----------------|---------------|-----------------|----------------|
| 6.7910 | 1.4586 | 46.1176 | 9.9053 |
| 6.9204 | 1.2470 | 47.8916 | 8.6299 |
| 7.0787 | 0.9708 | 50.1077 | 6.8718 |
| 7.2817 | 0.6931 | 53.0236 | 5.0473 |
| 7.5722 | 0.3001 | 57.3389 | 2.2725 |
| 8.0644 | -0.3425 | 65.0338 | -2.7620 |
| 43.7084 | 4.3271 | 319.5132 | 29.9648 |

Therefore,

$$a_0 = \frac{4.3271 \times 319.5132 - 43.7084 \times 29.9648}{6 \times 319.5132 - (43.7084)^2} = 10.947$$

$$a_1 = \frac{6 \times 29.9648 - 43.7084 \times 4.3271}{6 \times 319.5132 - (43.7084)^2} = -1.404$$

The original data and the linearised data are shown in Fig. 3.8.

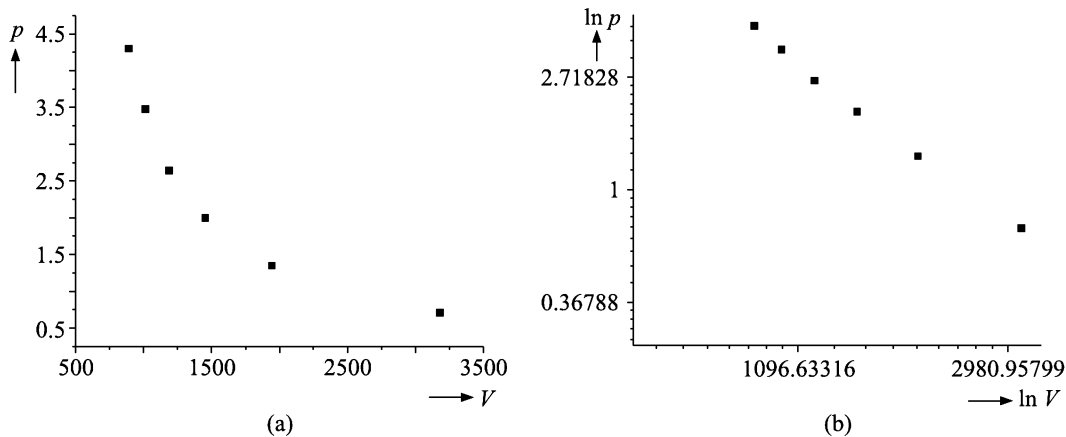


Fig. 3.8 (a) Original data, and (b) linearised data (Example 3.26).

In this way appropriate functional forms can be found to represent the data in a chosen interval and then the characteristic curve can be made linear. For very short intervals, a first approximation Taylor expansion can be resorted to even *extrapolate data* as shown in the following example.

Example 3.27

The output y of a transducer is given by $y = \sqrt{x}$ where x is the input. The value of the output is 2.3452 for an input of 5.5. What will be the linearised output when the input is 5.6?

Solution

The given functional form $f(x) = \sqrt{x}$ is nonlinear. Using the Taylor expansion we can write it in a linearised form at $x = a$ as

$$y = f(a) + f'(a)(x - a) = \sqrt{a} + \frac{1}{2\sqrt{a}}(x - a) \quad \left[\because f'(x) = \frac{1}{2\sqrt{x}} \right] \quad (i)$$

We are given that for $a = 5.5$, $f(a) = 2.3452$. Putting this value in the linear from given by Eq. (i), we get for $x = 5.6$

$$y = 2.3452 + \frac{1}{2 \times 2.3452}(5.6 - 5.5) = 2.3665$$

The actual value is 2.3664.

Hardware linearisation. Perhaps the most obvious and simple way to improve linearity of an output from a system is to use some kind of feedback as shown schematically in Fig. 3.9.

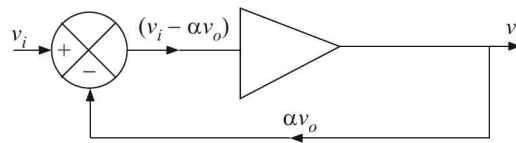


Fig. 3.9 Negative feedback.

If the output v_o is expressed in terms of a power series of the input v_i as

$$v_o = a \cdot v_i + b \cdot v_i^2 + c \cdot v_i^3 + \dots$$

then though the feedback αv_o we have

$$\begin{aligned} v_o &= a \cdot (v_i - \alpha v_o) + b \cdot (v_i - \alpha v_o)^2 + c \cdot (v_i - \alpha v_o)^3 + \dots \\ &\cong a \cdot (v_i - \alpha v_o) \quad [\text{when } (v_i - \alpha v_o) \text{ is small}] \end{aligned}$$

Apart from this, some other methods of hardware linearisation have been discussed in Sections 6.2 at page 187, 8.6 at page 308, 10.3 at page 395 and 16.1 at page 751.

However, the issue of hardware linearisation is becoming less important, because contemporary data-acquisition systems have built-in software to convert data to the desired physical quantity, making hardware linearisation unnecessary.

Dynamic systems

Because Laplace transform can be applied to only a linear differential equation, it is sometimes necessary to linearise a nonlinear differential equation. The goal of such linearisation is to study the behaviour of a system when it is perturbed from its steady-state. We will study a simple system where the first order Taylor expansion can be used for our purpose.

Consider a system where liquid level in a tank has to be maintained as shown in Fig. 3.10. Liquid is being poured in at a rate of Q . If the velocity of the outflow is v then

$$v^2 = 2gh$$

or

$$v = \sqrt{2gh}$$

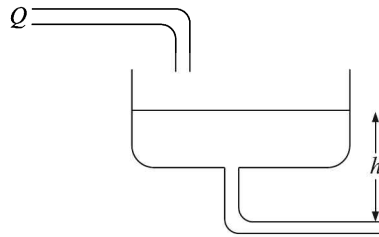


Fig. 3.10 Maintenance of liquid level in a tank.

where h is the height of the liquid column at any instant. The discharge coefficient as well as cross-sectional area of the pipe will also come into play. So, let the outflow rate be $\beta\sqrt{h}$, where β is a constant. Now, the dynamical situation can be described by

$$Q(t) = A \frac{dh}{dt} + \beta\sqrt{h} \quad (3.16)$$

which is a nonlinear equation. We can linearise it as follows.

If the inflow rate at the steady-state is Q_s and h_s is the corresponding height of the liquid column, we have

$$Q_s = \beta\sqrt{h_s} \quad (3.17)$$

Now, through the first order Taylor expansion around the steady-state value, we have from Eq. (3.16)

$$Q(t) = A \frac{dh}{dt} + \beta \left[\sqrt{h_s} + \frac{1}{2\sqrt{h_s}}(h - h_s) \right]$$

or

$$Q(t) - Q_s = A \frac{dh}{dt} + \frac{\beta}{2\sqrt{h_s}}(h - h_s) \quad [\text{using Eq. (3.17)}] \quad (3.18)$$

If we put $Q_{\text{in}}(t) = Q(t) - Q_s$ and $H = h - h_s$, we have from Eq. (3.18)

$$A \frac{dH}{dt} + \frac{\beta}{2\sqrt{h_s}}H = Q_{\text{in}}(t) \quad \left[\because \frac{dh}{dt} = \frac{dH}{dt} \right]$$

This is indeed a linear differential equation.

Spline Interpolation of Data

Often it is required to interpolate data between two data points. More often than not a linear interpolation is resorted to. If the data interval is small, a linear interpolation may yield good result. But, if we face a situation where we have to interpolate between scant data of a quickly varying function, as shown in Fig. 3.11 where obviously linear interpolations will produce unreliable data, it is better to use the spline interpolation.

The most common spline interpolations used are linear, quadratic and cubic splines. The linear spline interpolation being similar to the first order Taylor expansion, we will consider the quadratic spline interpolation.

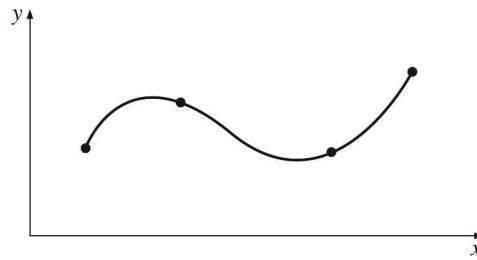


Fig. 3.11 Interpolation from scant data of a quickly varying function.

Quadratic spline

Suppose, our data points are $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n), \dots$. We want to fit quadratic splines through the data. The splines are given by

$$\begin{aligned} f(x) &= a_1x^2 + b_1x + c_1 & x_0 \leq x \leq x_1 \\ &= a_2x^2 + b_2x + c_2 & x_1 \leq x \leq x_2 \\ &\vdots \\ &= a_nx^2 + b_nx + c_n & x_{n-1} \leq x \leq x_n \end{aligned}$$

In these n equations, the number of coefficients is $3n$. In order to find $3n$ coefficients, we need to set up as many equations and then solve them simultaneously. The equations are set up as follows:

1. Since each quadratic spline passes through two consecutive data points, we have

$$\begin{aligned} a_1x_0^2 + b_1x_0 + c_1 &= f(x_0) \\ a_1x_1^2 + b_1x_1 + c_1 &= f(x_1) \\ a_2x_1^2 + b_2x_1 + c_2 &= f(x_1) \\ a_2x_2^2 + b_2x_2 + c_2 &= f(x_2) \\ &\vdots \\ a_nx_{n-1}^2 + b_nx_{n-1} + c_{n-1} &= f(x_{n-1}) \\ a_nx_n^2 + b_nx_n + c_n &= f(x_n) \end{aligned}$$

This condition gives us $2n$ equations since each of n quadratic splines passes through two consecutive data points.

2. We need to match the slopes at the meeting point of two splines so that the first derivatives of two consecutive splines are continuous at the interior points. For example, the derivative of the first spline

$$\begin{aligned} &a_1x^2 + b_1x + c_1 \\ \text{is} &2a_1x + b_1 \end{aligned}$$

The derivative of the second spline

$$\begin{aligned} &a_2x^2 + b_2x + c_2 \\ \text{is} &2a_2x + b_2 \end{aligned}$$

Since the two are equal at $x = x_1$, we have

$$2a_1x_1 + b_1 = 2a_2x_1 + b_2$$

or

$$2a_1x_1 + b_1 - 2a_2x_1 - b_2 = 0$$

In a similar way, we get for other interior points:

$$2a_2x_2 + b_2 - 2a_3x_2 - b_3 = 0$$

$$2a_3x_3 + b_3 - 2a_4x_3 - b_4 = 0$$

\vdots

$$2a_{n-1}x_{n-1} + b_{n-1} - 2a_nx_{n-1} - b_n = 0$$

We will have $(n - 1)$ such equations because there are $(n - 1)$ interior points between n data points.

3. The total number of equations so far obtained is $2n + (n - 1) = 3n - 1$. We need one more equation. That is obtained by assuming that the first spline is linear, which gives

$$a_1 = 0$$

4. All $3n$ equations are solved simultaneously to obtain $3n$ coefficients.

Example 3.28

The resistance-temperature data of a thermistor is given as follows:

| | | | | |
|----------------------------|--------|-------|-------|-------|
| T ($^{\circ}\text{C}$) | 30 | 50 | 70 | 100 |
| R_T (Ω) | 1428.9 | 800.0 | 479.3 | 246.3 |

Find the calibration curve using quadratic splines, and find the temperature corresponding to 918.57Ω . What is the per cent error from the calculated value obtained from the equation

$$\frac{1}{T} = A + B \ln R_T$$

where, $A = 7.4084 \times 10^{-4}$ and $B = 3.5232 \times 10^{-4}$?

Solution

Since there are four data points, three quadratic splines will pass through them. Thus,

$$\begin{aligned} f(R_T) &= a_1R_T^2 + b_1R_T + c_1 & 1428.9 \leq R_T \leq 800.0 \\ &= a_2R_T^2 + b_2R_T + c_2 & 800.0 \leq R_T \leq 479.3 \\ &= a_3R_T^2 + b_3R_T + c_3 & 479.3 \leq R_T \leq 246.3 \end{aligned}$$

We need to find the coefficients $a_1, b_1, c_1 \dots a_3, b_3, c_3$. The equations are obtained as follows:

1. Each quadratic spline passes through two consecutive data points.

$a_1R_T^2 + b_1R_T + c_1$ passes through $R_T = 1428.9$ and $R_T = 800.0$. Therefore,

$$\begin{aligned} a_1(1428.9)^2 + b_1(1428.9) + c_1 &= (30 + 273) = 303 \\ \Rightarrow 2.0418 \times 10^6 a_1 + 1428.9b_1 + c_1 &= 303 \end{aligned} \tag{i}$$

$$\begin{aligned} a_1(800.0)^2 + b_1(800.0) + c_1 &= (50 + 273) = 323 \\ \Rightarrow 6.4000 \times 10^5 a_1 + 800.0b_1 + c_1 &= 323 \end{aligned} \tag{ii}$$

$a_2R_T^2 + b_2R_T + c_2$ passes through $R_T = 800.0$ and $R_T = 479.3$. Therefore,

$$\begin{aligned} & a_2(800.0)^2 + b_2(800.0) + c_2 = (50 + 273) = 323 \\ \Rightarrow & 6.4000 \times 10^5 a_2 + 800.0b_2 + c_2 = 323 \end{aligned} \quad \text{(iii)}$$

$$\begin{aligned} & a_2(479.3)^2 + b_2(479.3) + c_2 = (70 + 273) = 343 \\ \Rightarrow & 2.2973 \times 10^5 a_2 + 479.3b_2 + c_2 = 343 \end{aligned} \quad \text{(iv)}$$

$a_3R_T^2 + b_3R_T + c_3$ passes through $R_T = 479.3$ and $R_T = 246.3$. Therefore,

$$\begin{aligned} & a_3(479.3)^2 + b_3(479.3) + c_3 = (70 + 273) = 343 \\ \Rightarrow & 2.2973 \times 10^5 a_3 + 479.3b_3 + c_3 = 343 \end{aligned} \quad \text{(v)}$$

$$\begin{aligned} & a_3(246.3)^2 + b_3(246.3) + c_3 = (100 + 273) = 373 \\ \Rightarrow & 6.0664 \times 10^4 a_3 + 246.3b_3 + c_3 = 373 \end{aligned} \quad \text{(vi)}$$

2. Quadratic splines have continuous derivatives at the interior data points.

At $R_T = 800.0$

$$\begin{aligned} & 2a_1(800.0) + b_1 - 2a_2(800.0) - b_2 = 0 \\ \Rightarrow & 1600a_1 + b_1 - 1600a_2 - b_2 = 0 \end{aligned} \quad \text{(vii)}$$

At $R_T = 479.3$

$$\begin{aligned} & 2a_2(479.3) + b_2 - 2a_3(479.3) - b_3 = 0 \\ \Rightarrow & 958.6a_2 + b_2 - 958.6a_3 - b_3 = 0 \end{aligned} \quad \text{(viii)}$$

3. Assuming that the first spline $a_1R_T^2 + b_1R_T + c_1$ is linear, we have

$$a_1 = 0 \quad \text{(ix)}$$

Arranging Eqs. (i) to (ix) in matrix form, we get

$$\begin{bmatrix} 2.0418 \times 10^6 & 1428.9 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6.4000 \times 10^5 & 800.0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6.4000 \times 10^5 & 800.0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.2973 \times 10^5 & 479.3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2.2973 \times 10^5 & 479.3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6.0644 \times 10^4 & 246.3 & 1 \\ 1600 & 1 & 0 & -1600 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 958.6 & 1 & 0 & -958.6 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \\ a_3 \\ b_3 \\ c_3 \end{bmatrix} = \begin{bmatrix} 303 \\ 323 \\ 323 \\ 343 \\ 343 \\ 373 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

or

$$\begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \\ a_3 \\ b_3 \\ c_3 \end{bmatrix} = \begin{bmatrix} 2.0418 \times 10^6 & 1428.9 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6.4000 \times 10^5 & 800.0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6.4000 \times 10^5 & 800.0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.2973 \times 10^5 & 479.3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2.2973 \times 10^5 & 479.3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6.0644 \times 10^4 & 246.3 & 1 \\ 1600 & 1 & 0 & -1600 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 958.6 & 1 & 0 & -958.6 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 303 \\ 323 \\ 323 \\ 343 \\ 343 \\ 373 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\
 = \begin{bmatrix} 0 \\ -0.0318 \\ 348.4412 \\ 0.0001 \\ -0.1843 \\ 409.4310 \\ 0.0002 \\ -0.2406 \\ 422.9138 \end{bmatrix}$$

Therefore, the splines are given by

$$\begin{aligned}
 f(R_T) &= 0R_T^2 - 0.0318R_T + 348.4412 & 1428.9 \leq R_T \leq 800.0 \\
 &= 0.0001R_T^2 - 0.1843R_T + 409.4310 & 800.0 \leq R_T \leq 479.3 \\
 &= 0.0002R_T^2 - 0.2406R_T + 422.9138 & 479.3 \leq R_T \leq 246.3
 \end{aligned}$$

At $R_T = 918.57 \Omega$

$$\begin{aligned}
 f(918.57) &= -0.0318(918.57) + 348.4412 \\
 &= 319.22 \text{ K}
 \end{aligned}$$

The calculated value from the given equation is $317.9999 \cong 318 \Omega$. So the error is

$$\frac{319.22 - 318}{319.22} \times 100 = 0.38\%$$

Note: The spline interpolation is not an overall curve fitting method. It simply is a better method of interpolation. The linear spline interpolation, which is similar to first order Taylor expansion, can be called a linearisation method over a short interval. But it will not be correct to say that quadratic or cubic splines are linearisation methods.

3.7 Reliability Principles

So far we have talked about the statistical methods of data handling and checking the reliability of the obtained data. Now, we will deal with the reliability of the measurement system itself. Underpinning the design of safety and protection of processes lies the calculations about the reliability of measurement systems and elements which are integral parts of processes or total systems. Three terms are defined in this context:

1. Reliability
2. Unreliability
3. Mean failure rate

Reliability

The reliability $R(t)$ of a measurement system or an element is defined as the probability that it will operate:

1. To an agreed level of performance
2. For a specified period of time
3. Under specified conditions
4. When used for the manner and purpose for which it was intended

Suppose, the agreed level of performance of a voltmeter is an accuracy of $\pm 2\%$ and the warranty period is 1 year. It should not be used above 40° and its maximum input should be 220 V. As long as the instrument is used in the specified conditions and it gives readings within this specified error limits, we consider it reliable. If it does not, although the system may be otherwise all right, it will be considered to have failed.

Because reliability is defined as a probability, its value always lies between 0 and 1. Quantitatively, reliability is defined as

$$R(t) = \Pr \{0 \text{ failures in } [0, t] \mid \text{no failure at } t = 0\}$$

Let n_s be the number of elements that working correctly at time t , n_f be the number of elements that have failed at time t , and n_0 be the total number of elements. Then,

$$R(t) = \frac{n_s}{n_0} = \frac{n_s}{n_s + n_f} \quad (3.19)$$

We have denoted $R(t)$ as a function of time because it is a common experience that an instrument that has just been checked and calibrated has $R(t) = 1$, but after a lapse of, say, six months it might have a reliability of 0.9. It is also a common experience that the reliability of a system always decreases with time.

Unreliability

The unreliability, normally denoted by $F(t)$, is the complement of reliability. Mathematically, it can be written as

$$F(t) = \frac{n_f}{n_0} = \frac{n_f}{n_s + n_f}$$

Since an element or system can only have failed or not failed, the total probability, that is, the sum of reliability and unreliability, must be unity. Thus,

$$R(t) + F(t) = 1$$

The unreliability, too, is a function of time, because as the reliability of a system decreases with time, unreliability, being the complement of reliability, increases with time. It is a probability value, or a number between 0 and 1, that indicates the likelihood that a system cannot continuously operate up to a specified point in time.

Mean Failure Rate

The mean failure rate λ is the average of faults per device per unit time. It is commonly expressed in terms of failures per year or failures per million hours (FPMH). For example, if a television has a failure rate of five per million hours, it means that one can watch one million one hour-long television shows and may experience a failure during only five shows.

Before we go into more theoretical considerations regarding $R(t)$ and $F(t)$, we need to divide the measurement systems into two categories:

1. Unrepairable
2. Repairable

Unrepairable Systems

Unrepairable systems are those which are discarded once a fault occurs. Typical examples that are cited are artificial satellites, sub-sea valves, missiles, microprocessors, hard disc drives, etc. Much of the reliability theory has been developed in relation to unrepairable systems. A time-independent parameter which is used to specify an unrepairable system is called the *mean time to failure* (MTTF).

Mean time to failure

Failure rates are calculated from MTTF data. Suppose, n_0 number of new devices comprise an unrepairable system. The devices and their conditions are identical and they are allowed to operate until they fail. The time taken for each device to fail is noted and once it fails, it is taken out of service. The average of these times, when all n_0 devices have failed, is the MTTF. Thus, if we say the total survival time or 'up time' for the i -th failure is T_i , then

$$\text{MTTF} = \frac{\text{Total up time}}{\text{Total number of failures}} = \frac{\sum_{i=1}^{n_0} T_i}{n_0} \quad (3.20)$$

Obviously then,

$$\lambda = \frac{\text{Total number of failures}}{\text{Total up time}} = \frac{n_0}{\sum_{i=1}^{n_0} T_i} = \frac{1}{\text{MTTF}} \quad (3.21)$$

If the units of λ are number of failures per hour, then those of MTTF are hours.

We observe that at time $t = 0$, n_0 devices survive and at $t = T$, 0 devices survive. So, from Eq. (3.19) the reliability $R(i)$ is

$$R(i) = \frac{\text{Number of devices survived at } t = T_i}{\text{Total number of devices at } t = 0} = \frac{n_0 - i}{n_0} \quad (3.22)$$

From Eq. (3.22) we can construct a survival table as shown in Table 3.2. The $R(i)$ vs. t plot

Table 3.2 Survival table of devices

| Time (t) | Number of survivors | Reliability [$R(i)$] |
|--------------|---------------------|------------------------|
| 0 | n_0 | 1 |
| T_1 | $n_0 - 1$ | $\frac{n_0 - 1}{n_0}$ |
| \vdots | \vdots | \vdots |
| T_i | $n_0 - i$ | $\frac{n_0 - i}{n_0}$ |
| \vdots | \vdots | \vdots |
| T | 0 | 0 |

looks like Fig. 3.12 where rectangles have heights of $1/n_0$.

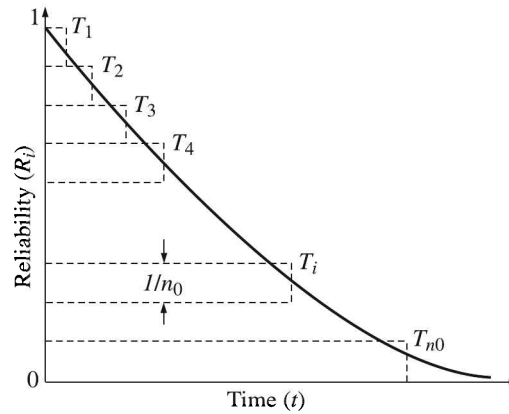


Fig. 3.12 Reliability vs. time plot for unreparable systems.

The area under the curve is given by

$$\begin{aligned} \text{Area} &= \left(T_1 \cdot \frac{1}{n_0}\right) + \left(T_2 \cdot \frac{1}{n_0}\right) + \left(T_3 \cdot \frac{1}{n_0}\right) + \cdots + \left(T_{n_0} \cdot \frac{1}{n_0}\right) \\ &= \frac{\sum_{i=1}^{n_0} T_i}{n_0} \end{aligned} \quad (3.23)$$

Comparing Eqs. (3.20) and (3.23), we find that the area is the MTTF itself. By making $n_0 \rightarrow \infty$, we can write Eq. (3.23) as

$$\text{MTTF} = \int_0^{\infty} R(t) dt \quad (3.24)$$

Reliability $R(t)$ and failure rate λ are related. Let us consider a simple analysis here.

Relation between $R(t)$ and λ

We consider n_0 identical devices put to operation at time $x = 0$. If n_f devices failed at $x = t$ and if no repairs were done, then the number of surviving devices n_s is

$$n_s = n_0 - n_f \quad (3.25)$$

Now, suppose an additional Δn_f devices fail during the next time interval Δx . Then, if we assume that the mean failure rate is approximately equal to the failure probability per unit time,

$$\lambda \approx \frac{\text{Failure probability}}{\text{Time interval}} = \frac{\Delta n_f / n_s}{\Delta x} = \frac{1}{n_s} \frac{\Delta n_f}{\Delta x} \quad (3.26)$$

Equation (3.26) gives us in the limit

$$\lambda = \frac{1}{n_s} \cdot \frac{dn_f}{dx} \quad (3.27)$$

Differentiating Eq. (3.25) we get

$$\frac{dn_s}{dx} = - \frac{dn_f}{dx} \quad (3.28)$$

So, from Eqs. (3.27) and (3.28) we get

$$\frac{dn_s}{dx} = -\lambda n_s \quad (3.29)$$

Assuming λ to be constant, we get by integrating Eq. (3.29)

$$\int_{n_0}^{n_s} \frac{dn_s}{n_s} = -\lambda \int_0^t dx$$

or

$$n_s = n_0 \exp(-\lambda t) \quad (3.30)$$

Now, reliability $R(t)$ is the ratio of n_s to n_0 . Therefore, we get from Eq. (3.30)

$$R(t) = \exp(-\lambda t) = \exp\left(-\frac{t}{\text{MTTF}}\right) \quad (3.31)$$

We reiterate that while arriving at the result given by Eq. (3.31), we have assumed that

1. There are no repairs, and
2. The mean failure rate is constant

This reliability model is eminently known as that based on the exponential distribution function. There are other models based on distributions like

1. Normal or Gaussian distribution
2. Log-normal distribution
3. Γ distribution
4. Weibull distribution

But we will carry on our analysis with the simple exponential distribution.

Repairable Systems

As the name suggests, the repairable systems are those in which when the system fails, faulty parts are repaired and/or replaced, and the system is put back into service.

A particularly useful metric for repairable systems is the *mean time between failures* (MTBF). Let n_0 identical repairable systems or devices be tested over a period of time T . Once a fault occurs, it is recorded, the device repaired and put back into service. If T'_i be the down time of the i -th failure (Fig. 3.13), the total down time for n_f failures is $\sum_{i=1}^{n_f} T'_i$ and the *mean down time* (MDT) is

$$\text{MDT} = \frac{\sum_{i=1}^{n_f} T'_i}{n_f}$$

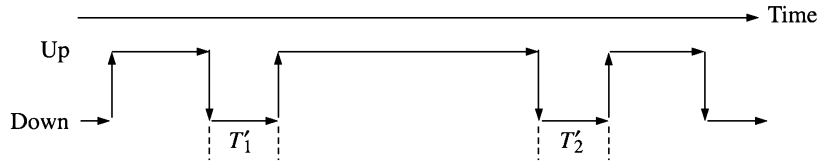


Fig. 3.13 Failure pattern of repairable systems.

Obviously, the total up time is then

$$\text{Total up time} = n_0 T - \sum_{i=1}^{n_f} T'_i = n_0 T - n_f (\text{MDT})$$

The MTBF is, therefore,

$$\text{MTBF} = \frac{\text{Total up time}}{\text{Total number of failures}} = \frac{n_0 T - n_f (\text{MDT})}{n_f} \quad (3.32)$$

and the corresponding mean failure rate is

$$\lambda = \frac{1}{\text{MTBF}} \quad (3.33)$$

$$= \frac{n_f}{n_0 T - n_f (\text{MDT})} \quad (3.34)$$

Example 3.29

What is the probability that a repairable element of a system will survive till its MTBF?

Solution

According to Eq. (3.33), the mean failure rate λ of the element is equal to $(1/\text{MTBF})$. Therefore, from Eq. (3.31) the reliability of the element is

$$R_{\text{MTBF}} = \exp(-1) = 0.3677$$

In other words, the probability that the element will survive till its MTBF is 36.8%.

Example 3.30

Calculate the MTBF and the mean failure rate if 100 faults were recorded for 300 transducers of a system during 1.5 years, the mean down time being 1 day.

Solution

$\text{MDT} = 1 \text{ day} = \frac{1}{365} \text{ yr}$. Therefore, from Eq. (3.32)

$$\text{MTBF} = \frac{(300)(1.5) - (100)(1/365)}{100} = 4.497 \text{ yrs}$$

$$\text{Mean failure rate } \lambda = \frac{1}{4.497} = 0.222 \text{ per yr}$$

Note: Sometimes, the vendors mention a term *mean time to repair* (MTTR) which is lower than MDT because their MTTR is based on the assumption that a fully trained technician, complete with appropriate spares and test equipment, is ready for 24 h a day and that failed equipment is immediately available for repairs. So, the MTTR is rather an optimistic estimate while MDT, being the sum of MTTR and other delays, is more realistic.

Availability

The availability A is the probability that a system will be functioning correctly when needed. In other words, it is the fraction of the total up time during a test interval. That is,

$$\begin{aligned} A &= \frac{\text{Total up time}}{\text{Test interval}} \\ &= \frac{\text{Total up time}}{\text{Total up time} + \text{Total down time}} \\ &= \frac{(n_f)(\text{MTBF})}{(n_f)(\text{MTBF}) + (n_f)(\text{MDT})} \\ &= \frac{\text{MTBF}}{\text{MTBF} + \text{MDT}} \end{aligned} \quad (3.35)$$

Example 3.31

Calculate the availability of the system of Example 3.30.

Solution

In Example 3.30, we had $MTBF = 4.497$ yrs and $MDT = (1/365)$ yr.

Therefore, from Eq. (3.35)

$$A = \frac{4.497}{4.497 + (1/365)} = 0.999$$

In other words, the availability is 99.9%.

Note: An availability of 99.9% is often called *three nines availability*. The one nine availability is not 9%, but 90%, two nines, 99% and five nines, 99.999%.

Unavailability

The unavailability U is the complement of availability. Since these are probabilities,

$$U = 1 - A$$

$$= \frac{MDT}{MTBF + MDT}$$

Substituting $1/\lambda$ for $MTBF$ [see Eq. (3.33)], we get

$$U = \frac{\lambda(MDT)}{1 + \lambda(MDT)} \approx \lambda(MDT)$$

assuming $\lambda(MDT) \ll 1$.

Hazard rate

Often a *hazard rate* or *instantaneous failure rate* is defined if λ is not constant. Written as $\lambda(t)$, it is defined as

$$\lambda(t) = \frac{\text{Failure probability}}{\Delta t} = \frac{\Delta n_f}{n_0 \Delta t}$$

where Δt is a small span of time.

Bath tub curve

The instantaneous failure rate of a typical system or element varies throughout its life as shown in Fig. 3.14. The shape of the curve has made it known as the *bath tub curve*.

The bath tub curve reveals three distinct phases, each with its own characteristics.

Infant mortality phase. The first phase is so called because it occurs early in the life of the system. This phase occurs owing to faulty design and manufacturing faults. Modifications are made and faulty elements are repaired or replaced with good ones. This decreases the failure rate. The majority of this phase occurs during testing by the manufacturer or during commissioning.

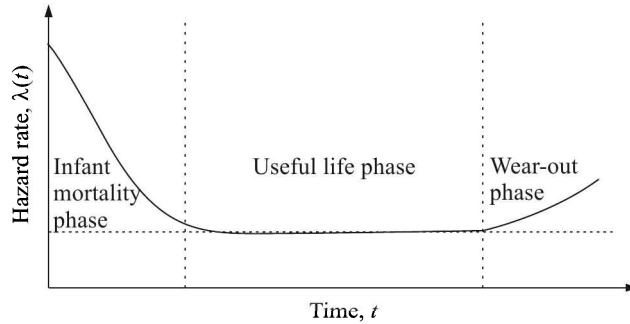


Fig. 3.14 The bath tub curve.

Useful life phase. The useful life phase is characterised by a constant failure rate. Here, though all bad components have been removed, random failure of components does not normally make $\lambda = 0$. All the metrics related to repairable systems pertain to this phase.

Wear-out phase. In this phase, $\lambda(t)$ slowly increases with time as individual elements start reaching their designed life-spans.

System Reliability

From the hardware point of view, measurement systems usually comprise of sensors, signal conditioners and display devices connected in series. But to provide alternative paths in case of failures, sometimes some of them are connected in parallel as well. We will consider system reliability in each type of connection.

Elements in series

Consider a system of n elements in series with failure rates of $\lambda_1, \lambda_2, \dots, \lambda_n$ as shown in Fig. 3.15.

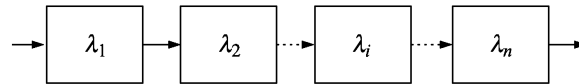


Fig. 3.15 Elements in series.

The system will remain up as long as every element remains up; if one element fails, the system fails. Consequently, the system reliability is the product of individual element reliabilities

$$R_{\text{series}} = R_1 \cdot R_2 \dots R_i \dots R_n = \prod_{i=1}^n R_i \quad (3.36)$$

Substituting the values of R 's from Eq. (3.31), we get

$$\begin{aligned} \exp(-\lambda_{\text{series}}t) &= \exp(-\lambda_1 t) \cdot \exp(-\lambda_2 t) \dots \exp(-\lambda_i t) \dots \exp(-\lambda_n t) \\ &= \exp[-(\lambda_1 + \lambda_2 + \dots + \lambda_n)t] \end{aligned} \quad (3.37)$$

From Eq. (3.37), we get the expression for the failure rate of a system of elements connected in series as

$$\lambda_{\text{series}} = \sum_{i=1}^n \lambda_i$$

For repairable systems, availability, rather than reliability, is more important. From the same argument as that of reliability, we get for the availability

$$A_{\text{series}} = \prod_{i=1}^n A_i$$

The unavailability can be obtained by substituting $A_i = 1 - U_i$, expanding the expression obtained and neglecting second and higher order terms (because $0 < U \ll 1$) to yield

$$U_{\text{series}} = U_1 + U_2 + \dots + U_n = \sum_{i=1}^n U_i \quad (3.38)$$

Thus, according to Eq. (3.38) the overall unavailability of a system, having elements connected in series, is the sum of unavailability of individual elements.

Elements in parallel

Parallel elements do naturally occur in some systems. But more often than not, parallelism is deliberately introduced to increase reliability. This is known as *redundancy*.

Consider a system of n elements connected in parallel as shown in Fig. 3.16.

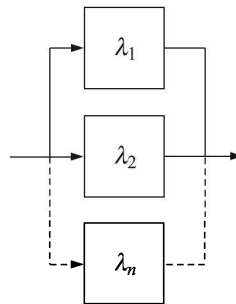


Fig. 3.16 Elements in parallel.

Here, the overall system will fail if every single element fails; if one element survives, the system survives. If U_1, U_2, \dots, U_n are the unavailabilities and if they are independent, then the probability that the overall system will be unavailable provided element 1 is unavailable and the element 2 is unavailable and ... element n is unavailable. Whence

$$U_{\text{parallel}} = U_1 \cdot U_2 \cdot \dots \cdot U_n = \prod_{i=1}^n U_i$$

By the same argument, the unreliability of a parallel system is given by

$$F_{\text{parallel}} = \prod_{i=1}^n F_i \quad (3.39)$$

Common mode failure

So far we have tacitly assumed that reliabilities of elements of a system are independent of each other. However, in practice, it is quite common that a single fault affects many elements. Such failures are called as *common mode failures*.

Common mode failures have the effect of reducing the availability. Thus, for two elements in parallel having unavailabilities U_1 and U_2 , the system unavailability is

$$U_{\text{sys}} > U_1.U_2$$

This kind of situation is handled by defining

$$\lambda_{\text{sys}} = \lambda_{\text{ind}} + \lambda_{\text{dep}} \quad (3.40)$$

where λ_{ind} is the failure rate of the system assuming all its elements are independent. The contribution from dependencies λ_{dep} is evaluated from

$$\lambda_{\text{dep}} = \alpha.\lambda_{\text{max}} \quad (3.41)$$

where λ_{max} is the failure rate of the most unreliable element and α is the dependency factor. Values of α are tabulated. Their values depend on the context, the amount of redundancy in the system and the diagnostic coverage. Combining Eqs. (3.40) and (3.41), we get

$$\lambda_{\text{sys}} = \lambda_{\text{ind}} + \alpha.\lambda_{\text{max}}$$

Voting systems

The word *vote* literally means ‘formal indication of a choice between two or more courses of action’⁶. To enable redundancy, many voting strategies are introduced in measurement systems. Some typical voting strategies for continuous operations of systems are like:

1. The average of two closest of three analogue input channels is accepted.
2. The middle value from three analogue input channels is accepted. In case one of them fails, the lower of the remaining two is chosen.
3. The average of two analogue inputs is accepted. In case, one of them is out of a specified range, the other is accepted.

We have shown an arrangement for the second option in Fig. 3.17.

The measurand is the temperature of a body and the three channels are three identical thermistors, each having a mean failure rate of 1.2 per year. The three thermistors are connected to a selector which chooses the middle value and sends the signal to the converter and display systems successively. The mean failure rates of the selector, converter and display are 0.1 per year, 0.1 per year and 0.2 per year respectively. Suppose, the measured signals are 4.8 mV, 5.0 mV and 4.9 mV. The middle value selector selects 4.9 mV and sends it to the converter for its conversion to the temperature value in the desired scale. If the second thermistor fails, then the selector chooses 4.8 mV.

⁶ *Pocket Oxford Dictionary*, Oxford University Press (2004).

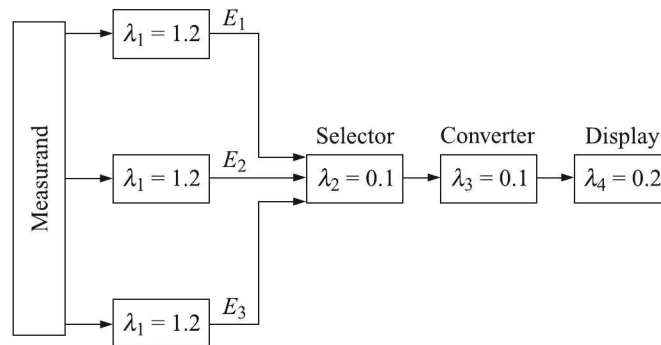


Fig. 3.17 Middle value selector voting system.

To analyse the reliability of the system during a test period of 0.5 year, we may, following Eq. (3.39), replace the three parallel thermistors with a single element of unreliability

$$F_1 = (1 - \exp(-\lambda_1 t))^3 = [1 - \exp\{-(1.2)(0.5)\}]^3 = 0.0918$$

Therefore, the reliability of the parallel thermistors is

$$R_1 = 1 - F_1 = 0.9082$$

We may note here that the reliability of a single thermistor is $\exp\{-(1.2)(0.5)\} = 0.5488$. So, the redundancy has almost doubled the reliability. The reliabilities of selector, converter and display are

$$R_2 = R_3 = \exp\{-(0.1)(0.5)\} = 0.9512 \quad R_4 = \exp\{-(0.2)(0.5)\} = 0.9048$$

Utilising Eq. (3.36), we get for the overall reliability of the voting system as

$$R_{\text{sys}} = \prod_{i=1}^4 R_i = (0.9082)(0.9512)^2(0.9048) = 0.7407$$

This yields the value of the unreliability of the overall system as $F_{\text{sys}} = 1 - 0.7407 = 0.2593$ which is nearly three times the unreliability of the parallel thermistors.

Review Questions

- 3.1 Discuss how limiting errors in y can be computed from the measurement of two quantities u and v , each having limiting errors when (a) $y = u + v$ and (b) $y = u/v$.
- 3.2 The relative errors in measurement of power P , voltage V and current I are $\pm 0.5\%$, $\pm 1\%$ and $\pm 1\%$ respectively. If power factor $\cos \phi = P/VI$, calculate (a) the relative error in power factor measurement, (b) the uncertainty in the power factor if the above errors were specified as uncertainties.
- 3.3 The measured value of a capacitor is $205 \mu\text{F}$, whereas its true value is $200.4 \mu\text{F}$. Determine the relative error.

- 3.4 Three resistors have the following ratings: $R_1 = 47 \Omega \pm 4\%$, $R_2 = 65 \Omega \pm 4\%$ and $R_3 = 55 \Omega \pm 4\%$. Determine the magnitude and limiting error in ohms and in percentage of the resistance if these resistances are connected in series.
- 3.5 A voltmeter and an ammeter are used to determine the power dissipated in a resistor. Both the measurements are guaranteed to be accurate within $\pm 1\%$ at full-scale deflection. If the voltmeter reads 80 V on its 150 V range and the ammeter reads 70 mA on its 100 mA range, determine the limiting error for the power calculation.
- 3.6 The formula for the unknown resistor measured with the help of Wheatstone bridge is $R_x = R_2 R_3 / R_1$, where $R_1 = 100 \Omega \pm 0.5\%$, $R_2 = 1000 \Omega \pm 0.5\%$ and $R_3 = 842 \Omega \pm 0.5\%$. Determine (a) the value of R_x , (b) its limiting error in per cent and in ohm.
- 3.7 The following 10 observations were recorded while measuring a voltage: 41.7, 42.0, 41.8, 42.0, 42.1, 41.9, 42.5 and 41.8 volts. Find (a) the mean, (b) the standard deviation, (c) the probable error of one reading, and (d) the probable error of the mean.
- 3.8 The following values were recorded during the measurement of a resistance: 147.2, 147.4, 147.9, 148.1, 147.1, 147.5, 147.6 and 147.5 ohms. Taking arithmetic mean as the central value, calculate (a) standard deviation, (b) probable error of one reading, and (c) probable error of the mean.
- 3.9 Ten observations of resistance made in an experiment are: 100.4 Ω , 99.2 Ω , 101.1 Ω , 100.5 Ω , 99.8 Ω , 102.0 Ω , 99.9 Ω , 101.7 Ω , 100.8 Ω , 101.2 Ω . Calculate (a) the arithmetic mean, (b) the standard deviation, and (c) the variance.
- 3.10 Indicate the correct choice:
- The current through a resistance is measured with uncertainties: $I = 4 \text{ A} \pm 0.5\%$, $R = 100 \Omega \pm 0.2\%$. The uncertainty in the measurement of power is
 - 1600 W $\pm 0.01\%$
 - 1600 W $\pm 0.2\%$
 - 1600 W $\pm 0.5\%$
 - 1600 W $\pm 1.02\%$
 - To measure 2 volts, if one selects a 0–100 volt range voltmeter which is accurate to within $\pm 1\%$, the error in one's measurement may be up to
 - $\pm 0.02\%$
 - $\pm 1\%$
 - $\pm 2\%$
 - $\pm 50\%$
 - A thermometer is calibrated from 150 to 200°C. The accuracy specified is $\pm 0.25\%$. The maximum static error in measurement is
 - $\pm 0.5^\circ\text{C}$
 - $\pm 0.375^\circ\text{C}$
 - $\pm 0.125^\circ\text{C}$
 - $\pm 0.0125^\circ\text{C}$

- (d) The radius of a sphere is given as 40.0 ± 0.5 mm. The estimated error in its mass is:
- (i) $\pm 3.75\%$
 - (ii) $\pm 1.25\%$
 - (iii) $\pm 12.5\%$
 - (iv) $\pm 0.125\%$
- (e) A large number of 230Ω resistors are obtained by combining 120Ω resistors with a standard deviation of 4.0Ω and 110Ω resistors with a standard deviation of 3.0Ω . The standard deviation of the 230Ω resistors thus formed will be
- (i) 3.5Ω
 - (ii) 5.0Ω
 - (iii) 7.0Ω
 - (iv) 12.0Ω
- (f) The calibration data for a pressure compensator of a pump is given below

| | | | | | | | | | | | |
|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Input x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Output y | 9.5 | 8.4 | 7.8 | 7.4 | 6.1 | 5.4 | 5.2 | 4.6 | 3.2 | 1.9 | 1.1 |

For the given data, the slope of the best-fit line applying the least squares method is:

- (i) 0.921
 - (ii) -0.803
 - (iii) 0.819
 - (iv) -0.945
- (g) The measurements of a source voltage are 5.9 V, 5.7 V and 6.1 V. The sample standard deviation of the readings is:
- (i) 0.013
 - (ii) 0.04
 - (iii) 0.115
 - (iv) 0.2
- (h) The reliability of an instrument refers to:
- (i) measurement changes due to temperature variation
 - (ii) degree to which repeatability continues to remain within specified limits
 - (iii) the life of the instrument
 - (iv) the extent to which the characteristics remain linear
- (i) Using the given data points tabulated below, a straight line passing through the origin is fitted using the least squares method. The slope of the line is

| | | | |
|-----|-----|-----|-----|
| x | 1.0 | 2.0 | 3.0 |
| y | 1.5 | 2.2 | 2.7 |

- (i) 0.9
(ii) 1.0
(iii) 1.1
(iv) 1.5
- 3.11 For the following given data,
 $x_1 = 49.7$, $x_2 = 50.1$, $x_3 = 50.2$, $x_4 = 49.6$, $x_5 = 49.7$
calculate the following
- (a) Arithmetic mean
(b) Deviation of each value
(c) Algebraic sum of the deviations
(d) Average deviation
(e) Standard deviation
- 3.12 A voltmeter reading of 70 V on its 100 V range and an ammeter reading of 80 mA on its 150 mA range are used to determine the power dissipated in a resistor. Both these instruments are guaranteed to be accurate within $\pm 1.5\%$ at full-scale deflection. Determine the limiting error of the power.
- 3.13 The voltage across a resistor is 200 V, with a probable error of $\pm 2\%$, and the resistance is 42Ω with a probable error of $\pm 1.5\%$. Calculate (a) the power dissipated in the resistor, and (b) the percentage error in the answer.
- 3.14 Calculate the absolute standard deviation and the coefficient of variation of the results of the following calculations. Round off each result so that it contains only significant digits. The numbers in parentheses are absolute standard deviations.
- (a) $y = 5(\pm 0.03) + 0.75(\pm 0.001) - 4.2(\pm 0.001)$
(b) $y = 67.1(\pm 0.25) \times 1.05(\pm 0.02) \times 10^{-17}$
(c) $y = 2.9(\pm 0.001) \div 179(\pm 2)$
- 3.15 If the variable X is dependent upon the experimental variables p, q, r, \dots each of which varies in a random and independent way, obtain the relationship of the variable X with the variances of the variables p, q, r, \dots .
- 3.16 Two resistors R_1 and R_2 are connected in series and then in parallel. The values of resistances are: $R_1 = 100.0 \Omega \pm 0.1\%$, $R_2 = 50 \Omega \pm 0.06\%$. Calculate the uncertainty in the combined resistance of both series and parallel arrangements.
- 3.17 For the following given data, calculate (a) the arithmetic mean, (b) the average deviation, (c) the algebraic sum of deviations, (d) the standard deviation, and (e) the variance. Data: $x_1 = 99.7$, $x_2 = 100.1$, $x_3 = 100.2$, $x_4 = 99.6$, $x_5 = 99.7$.
- 3.18 The voltage generated by a circuit is equally-dependent on the value of three resistors and is given by the following equation

$$V_{\text{out}} = \frac{R_1 R_2}{R_3}$$

If the tolerance of each resistor is 0.1%, what is the maximum error of the generated voltage?

- 3.19 The radius of curvature of a concave surface is measured by a ring spherometer using the formula

$$\frac{D^2}{8h} + \frac{h}{2}$$

where D is the diameter of the ring and h is the sagitta. Assuming that the sagitta is very small compared to the radius of curvature, calculate the nominal radius of curvature and its uncertainty. Given, $D = 40 \text{ mm} \pm 0.05 \text{ mm}$, and $h = 0.4 \text{ mm} \pm 0.001 \text{ mm}$. The errors indicated are standard deviations.

- 3.20 A method for checking large radius is shown in Fig. 3.18. The equation for the radius may be derived as

$$R = \frac{C^2}{8(d-h)} - \frac{h}{2}$$

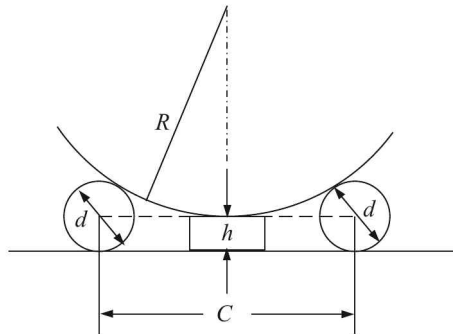


Fig. 3.18

Find the value of R and calculate its uncertainty of measurement. Given,

| <i>Dimension</i> | <i>Nominal size (mm)</i> | <i>Uncertainty (mm)</i> |
|------------------|--------------------------|-------------------------|
| C | 125 | ± 0.05 |
| d | 25 | ± 0.001 |
| h | 1.25 | ± 0.0005 |

Dynamic Characteristics of Instruments

As the name implies, these characteristics describe the behaviour of a system with time when some input is given to the system. Suppose a resistor R and a capacitor C are connected in series. If now a voltage source E_i is included in the circuit, charge builds up on the capacitor and the voltage across the resistor slowly builds up. This behaviour of the circuit can aptly be described by setting up an appropriate differential equation and solving it for certain boundary conditions. Of course, while setting up the equation, certain simplifying assumptions need be made to make the problem tractable. Here in the RC circuit, it is tacitly assumed that neither the capacitor is leaky nor the resistor dissipates any energy in the form of heat. An actual instrumentation system is rather complex and therefore, it is necessary to introduce certain ideal conditions to make mathematical studies tractable. This is called *mathematical modelling* of the problem. Once a model has been built, the response of the system, termed the *dynamic response*, is studied with respect to a few idealised inputs. Transfer functions of systems are very useful to study their responses.

4.1 Transfer Function

The generalised relation between a particular input $q_i [\equiv q_i(t)]$ and the corresponding output q_o with proper simplifying assumptions, can be written in the form

$$a_n \frac{d^n q_o}{dt^n} + \dots + a_1 \frac{dq_o}{dt} + a_0 q_o = b_m \frac{d^m q_i}{dt^m} + \dots + b_1 \frac{dq_i}{dt} + b_0 q_i \quad (4.1)$$

where a 's and b 's are combination of system parameters assumed to be constant.¹

Taking Laplace transform² and assuming all initial conditions equal to zero, we get from Eq. (4.1)

$$(a_n s^n + \dots + a_1 s + a_0) Q_o(s) = (b_m s^m + \dots + b_1 s + b_0) Q_i(s)$$

The transfer function $G(s)$ is defined as

$$G(s) \equiv \frac{Q_o(s)}{Q_i(s)} = \frac{b_m s^m + \dots + b_1 s + b_0}{a_n s^n + \dots + a_1 s + a_0} \quad (4.2)$$

¹Any system satisfying a relation of the type given by Eq. (4.1) is called a *linear time-invariant* system. See Appendix B at page 858 for the definition of such a system.

²See Appendix C at page 861 for a brief introduction to Laplace transform.

Properties of Transfer Function

1. The transfer function is a general relation between the Laplace transforms of the output and input quantities $Q_o(s)$ and $Q_i(s)$. It is not the instantaneous ratio of the time-varying quantities $q_o(t)$ and $q_i(t)$. For example, the relation between the current i and the emf e in an LCR circuit is given by

$$e(t) = L \frac{di(t)}{dt} + Ri(t) + \frac{1}{C} \int i(t) dt$$

Taking Laplace transform, we get

$$\mathcal{L}\{e(t)\} \equiv E(s) = sLI(s) + RI(s) + \frac{I(s)}{sC}$$

Therefore, the transfer function,

$$G(s) = \frac{I(s)}{E(s)} = \frac{1}{sL + R + (1/sC)}$$

2. The transfer function does not give any insight about the structure of the system.
3. It offers a symbolic picture about the dynamic characteristics of the system as shown in Fig. 4.1.
4. If the transfer functions of individual components of the system are known, the overall characteristics of the system can be determined just by taking their product (Fig. 4.1), provided the loading effect between connected devices can be neglected.

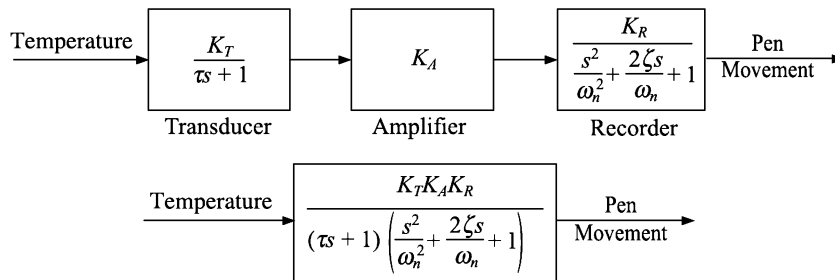


Fig. 4.1 Individual transfer functions are multiplied to get the final transfer function.

So far while discussing dynamic characteristics, we have tacitly assumed that the inputs to the system are time-varying and we want to study the dynamic response of the system at different intervals of time. Such kind of a study is called a *time domain analysis*.

But time domain analysis is rather cumbersome and, of course, not necessary if the input varies periodically with time, such as $q_i = A_i \sin \omega t$. The output quantity q_o in such cases will also be a sine wave, once the transients die out. The only changes that are expected are in the amplitude and the phase of the output. Since the input and output frequency are the same, the output is completely specified by giving the amplitude ratio A_o/A_i , and the phase shift angle ϕ . Thus, the response of a system to a periodic input is completely studied if the amplitude ratio and phase shift are studied as a function of frequency (Fig. 4.2). This analysis is termed *frequency domain analysis*.

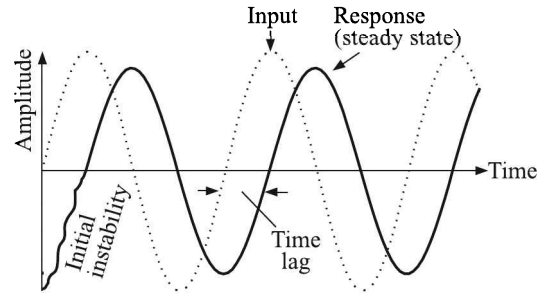


Fig. 4.2 A typical response of a system to a sinusoidal input

While a full-fledged time domain analysis with a sine wave input will also lead to the same results, much quicker and easier methods are offered by the concept of sinusoidal transfer function which is obtained simply by substituting $j\omega$ for s in the operational transfer function. Thus, the frequency domain transfer function for Eq. (4.2) is given by

$$G(j\omega) = \frac{Q_o(j\omega)}{Q_i(j\omega)} = \frac{b_m(j\omega)^m + \cdots + b_1j\omega + b_0}{a_n(j\omega)^n + \cdots + a_1j\omega + a_0}$$

$G(j\omega)$ for a given frequency ω is a complex quantity. Any complex quantity, $a + jb$, can be expressed in the polar form $M\angle\phi$ where $M (= \sqrt{a^2 + b^2})$ is the magnitude and $\phi (= \tan^{-1} \frac{b}{a})$ is the angle. It can be proved that M and ϕ corresponding to $G(j\omega)$ equal the amplitude ratio and the leading phase angle, respectively. By leading phase angle we mean that if the output lags behind the input, ϕ is negative (Fig. 4.3).

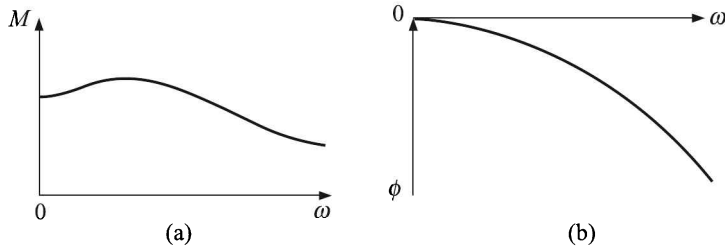


Fig. 4.3 Frequency response of system.

Thus, the sinusoidal transfer function for an LCR circuit is

$$G(j\omega) = \frac{X(j\omega)}{F(j\omega)} = \frac{1}{j\omega L + R + (1/j\omega C)}$$

4.2 Standard Inputs to Study Time Domain Response

Standard inputs generally used to study the dynamic behaviour of measurement systems and their Laplace transforms are as follows:

Step Input

The functional form [Fig. 4.4(a)] is given by

$$f(t) = \begin{cases} 0 & \text{if } t \leq 0 \\ A & \text{if } t > 0 \end{cases}$$

The Laplace transform of the step input is

$$\mathcal{L}\{f(t)\} = \frac{A}{s}$$

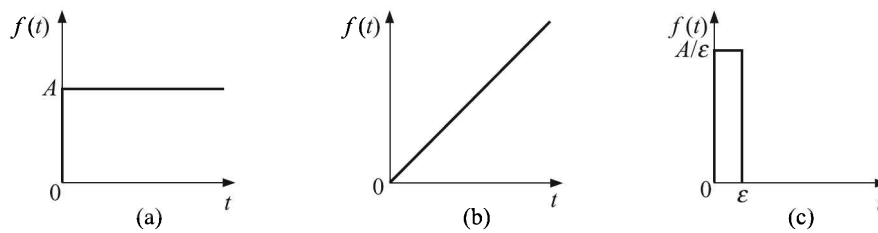


Fig. 4.4 Standard inputs: (a) step function, (b) ramp function, and (c) impulse function.

Ramp Input

The functional form [Fig. 4.4(b)] and the Laplace transform are

$$f(t) = At \quad \text{and} \quad F(s) = \frac{A}{s^2}$$

Impulse Input

The impulse function, [Fig. 4.4(c)] related to the Dirac δ -function is defined as

$$f(t) = A\delta(t)$$

where,

$$\delta(t) = \begin{cases} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} & \text{for } 0 \leq t \leq \epsilon \\ 0 & \text{for } t > \epsilon \end{cases}$$

By definition, $\int_0^{\infty} \delta(t) dt = 1$. Therefore, the corresponding Laplace transform³ is $\mathcal{L}\{\delta(t)\} = 1$. Thus,

$$F(s) = A$$

Armed with this background knowledge we will now study dynamic responses of different orders of instruments. But before that we will see what characteristics we want to watch. In other words, what are the dynamic characteristics of instruments.

³See Appendix C.1 at page 863 for a derivation.

4.3 Dynamic Characteristics

Like the static characteristics, dynamic characteristics can also be divided into two basic categories:

1. Desirable
2. Undesirable

The tree looks like Fig. 4.5.

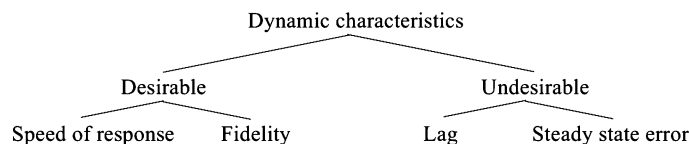


Fig. 4.5 Dynamic characteristics tree.

Although the names of characteristics are self-explanatory, we briefly discuss what they mean.

Speed or response. The speed of response indicates how quickly the system reacts to the input signal.

Fidelity. The fidelity of a dynamic system denotes how faithfully the system outputs the input signal and what is the distortion, if any.

Lag. As the name implies, the lag indicates what time the system takes to output the input signal.

Steady state error. The steady state error is defined as

$$e_{ss} = \lim_{t \rightarrow \infty} e_m$$

where,

$$e_m = q_i - \frac{q_o}{K}$$

Here, q_i is the input signal

q_o/K is the normalised output

K is the amplification factor

It will be seen that in the ultimate analysis, the desirable and undesirable characteristics all depend on the speed of response, which, in turn, is related to the time constant τ of the system.

We will study dynamic characteristics of instruments according to their order. The question is, what do we mean by the order of an instrumentation system?

Order of a system

The order of a dynamic system is indicated by the highest power of the Laplace variable s in the rationalised denominator of the corresponding transfer function.

Suppose, the transfer function of an instrumentation system is given by

$$G(s) = \frac{s + 9}{s^2 + 2s + 9}$$

Here, the denominator is given by

$$s^2 + 2s + 9$$

which consists of a power of 2 for the Laplace variable s . Therefore, it represents a second order system. Now, the same transfer function can be written as

$$G(s) = \frac{1 + (9/s)}{s + 2 + (9/s)}$$

Does it indicate that it is a first order system? The answer is obviously 'no' because the denominator contains a fraction.

Note: The power of s being related to the degree of the differential equation describing the instrumentation system, actually it is the degree of the differential equation which determines the order of a dynamic system. But then, more often than not the systems are described by their transfer functions.

Let us now embark upon the study of different dynamic systems.

4.4 Zero Order Instrument

Suppose all a 's and b 's except a_0 and b_0 in Eq. (4.1) are zero, degenerating it to an algebraic equation

$$a_0 q_o(t) = b_0 q_i(t) \quad (4.3)$$

A *zero order* instrument is defined as one which closely obeys this equation over its range of operation. On rearranging Eq. (4.3), we get

$$q_o(t) = \frac{b_0}{a_0} q_i(t) = K q_i(t) \quad (4.4)$$

where $K = b_0/a_0 =$ static sensitivity.

Equation (4.4) being an algebraic relation, it is apparent that the output q_o faithfully follows the input q_i with no distortion or time lag of any sort. The zero order instrument, therefore, may be considered as ideal having a perfect dynamic response.

A potentiometer used for measuring displacements⁴ may be shown to be a zero-order instrument. In such an arrangement, a wire-wound resistance, provided with a sliding contact, is excited with a voltage. Assuming that the resistance is distributed linearly along its length L , we have

$$E_o = \frac{x}{L} E_i \equiv Kx \quad (4.5)$$

where E_o and E_i are the output and input voltages, x is the displacement and K is a constant.

⁴See Section 6.2 at page 173.

Note: We have made the following tacit assumptions to write Eq. (4.5):

1. The increase in resistance is continuous. But in actuality, for a wire-wound type potentiometer the wound wire has a finite diameter and hence the resistance increases in steps as the sliding contact (called *wiper*) moves (see Fig. 6.6 at page 175).
2. The winding is purely resistive, which is not true because all such windings have inductive and capacitive effects.
3. Electric loading by the voltage measuring instrument is negligible. In case there is loading, the relation will not be linear.
4. There is no mechanical loading by the sliding contact (i.e. wiper) and, therefore, no heat generation during sliding motions.

4.5 First Order Instrument

If the dynamic relation between the input and output of an instrument assumes the form

$$a_1 \frac{dq_o(t)}{dt} + a_0 q_o(t) = b_0 q_i(t)$$

it is called a *first order* instrument. However, this relation can be written with two rather than three coefficients as follows:

$$\tau \frac{dq_o(t)}{dt} + q_o(t) = K q_i(t) \quad (4.6)$$

Here, $K = b_0/a_0$ and $\tau = a_1/a_0$. τ is called the time constant because it has the dimension of time in physical processes.

Taking Laplace transform of Eq. (4.6), we get

$$\tau s Q_o(s) + Q_o(s) = K Q_i(s)$$

or

$$(1 + \tau s) Q_o(s) = K Q_i(s)$$

Therefore, the transfer function is given by

$$G(s) \equiv \frac{Q_o(s)}{Q_i(s)} = \frac{K}{1 + \tau s}$$

The familiar *RC* circuit is an example of a first order arrangement. Here the relation between e_i (= voltage, input) and Q (= charge, output) is given by,

$$R \frac{dQ}{dt} + \frac{Q}{C} = e_i$$

or

$$\tau \frac{dQ}{dt} + Q = K e_i$$

where, $\tau = RC$ and $K = C$.

Examples of First Order Instruments

Mercury-in-glass thermometer

The common mercury-in-glass thermometer, (Fig. 4.6) behaves as a first order instrument as can be seen from the analysis given below.

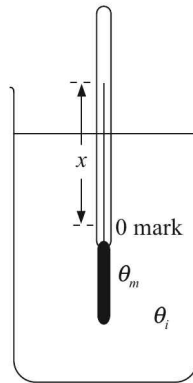


Fig. 4.6 Mercury-in-glass thermometer.

- Let V be the volume of the bulb
 A_b be the area of the bulb conducting heat
 γ_a be the coefficient of apparent expansion of mercury
 θ_m be the temperature attained by mercury at any instant
 x be the corresponding height of the mercury column (output)
 A be the area of cross-section of the capillary tube
 K_g be the thermal conductivity of glass
 ρ be the density of mercury
 c be the specific heat of mercury
 θ_i be the temperature of the liquid (input)
 θ be the wall thickness of the glass bulb,

then
$$x = \frac{\gamma_a V \theta_m}{A}$$

which gives
$$\theta_m = \frac{Ax}{\gamma_a V}$$

Now, the heat conducted from the liquid to mercury through the bulb (i.e. heat lost by the liquid) during the interval dt equals

$$\frac{K_g A_b}{\theta} (\theta_i - \theta_m) dt$$

The corresponding heat gained by mercury in the bulb is

$$(\rho c V) d\theta_m$$

Equating heat lost to heat gained and substituting the value of θ_m , we get after some rearrangement

$$\tau \frac{dx}{dt} + x = K\theta_i$$

where,

$$\tau = \frac{\rho c V \theta}{K_g A_b}$$

$$K = \frac{\gamma_a V}{A}$$

Thermocouple

The other common temperature measuring devices, such as thermocouple and thermistor, are also first order systems. To verify that, we consider a thermocouple which is dipped into a hot liquid. For simplicity, we assume that

- (a) the heat transfer takes place only by conduction
- (b) the emf vs temperature curve of the thermocouple is linear
- (c) the other junction of the thermocouple is kept at room temperature

If A is the heat transfer area of the thermocouple

K_t is the thermal conductivity of the thermocouple material

θ is the temperature attained by the thermocouple at any instant

θ_i is the temperature of the hot liquid

m is the mass of the thermocouple junction

c is the specific heat of the junction material

E is the developed emf in the thermocouple

we have

$$E = K\theta \quad (4.7)$$

where K is a constant. Heat conducted from the liquid to the thermocouple junction during a small time interval dt is

$$K_t A (\theta_i - \theta) dt$$

The corresponding heat gained by the thermocouple junction is

$$mc d\theta$$

Thus,

$$mc d\theta = K_t A (\theta_i - \theta) dt$$

which gives

$$\frac{mc}{K_t A} \frac{dE}{dt} + E = K\theta_i \quad [\text{applying Eq. (4.7)}] \quad (4.8)$$

or

$$\tau \frac{dE}{dt} + E = K\theta_i \quad (4.9)$$

where

$$\tau = \frac{mc}{K_t A}$$

Equation (4.9) shows that the thermocouple is a first order system.

Dynamic Response of First Order Instruments

Step response

As shown before, here $Q_i(s) = A/s$. Therefore,

$$Q_o(s) = G(s)Q_i(s) = \frac{KA}{s(\tau s + 1)} = KA \left(\frac{1}{s} - \frac{\tau}{\tau s + 1} \right) = KA \left(\frac{1}{s} - \frac{1}{s + \frac{1}{\tau}} \right) \quad (4.10)$$

Taking inverse Laplace transform of Eq. (4.10), we get

$$q_o(t) = KA[1 - \exp(-t/\tau)] \quad (4.11)$$

On non-dimensionalising, Eq. (4.11) becomes

$$\frac{q_o}{KA} = 1 - \exp(-t/\tau) \quad (4.12)$$

Measurement error. The measurement error e_m is

$$e_m = q_i - \frac{q_o}{K} = A - A[1 - \exp(-t/\tau)]$$

$$\Rightarrow \frac{e_m}{A} = \exp(-t/\tau) \quad (4.13)$$

Steady-state error. From Eq. (4.13) we find that the steady-state error for the step input of a first order instrument is

$$e_{ss} = \lim_{t \rightarrow \infty} e_m = 0$$

An idea about the growth of the output with time can be obtained from Table 4.1 which has been shown in graphical form in Fig. 4.7.

Table 4.1 Values of non-dimensionalised parameters for step response of the first order instrument

| t/τ | 0 | 1 | 2 | 3 | 4 | 5 | ∞ |
|----------|-------|-------|-------|-------|-------|-------|----------|
| q_o/KA | 0.000 | 0.632 | 0.865 | 0.950 | 0.982 | 0.993 | 1.000 |

From these analyses we can infer that

1. The speed of response depends only on the value of τ .
2. The response reaches within 5% of its final value at 3τ .
3. The steady-state value can be assumed to have reached around 5τ .

Going back to the case of mercury-in-glass thermometer, we find that

1. Reducing τ means reducing ρ , c and V , and increasing K_g and A_b ; but reducing V means reducing A_b as well!
2. ρ , c can be reduced by choosing an appropriate fluid for the thermometer;
3. By lowering V , the static sensitivity is lowered. This, in turn, means that a fast responding thermometer is less sensitive.

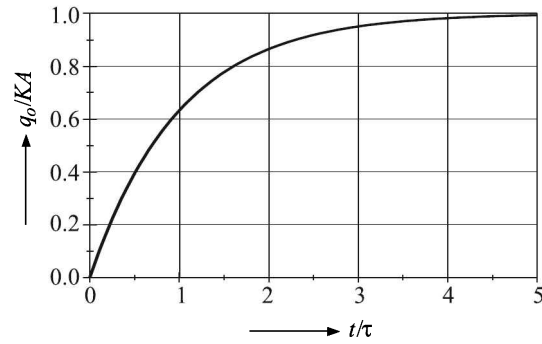


Fig. 4.7 Step response of the first order instrument.

Example 4.1

A thermometer, initially at 70°C , is suddenly dipped in a liquid at 300°C . After 3 s, the thermometer indicates 200°C . After what time is the thermometer expected to give a reliable reading, say well within 1% of the actual value?

Solution

This is obviously a case of a step input, the step being not from an initial zero value, but a finite value of $\theta_0 = 70^\circ\text{C}$. So if we denote θ_3 as the thermometer reading after 3 s, our equation is

$$\theta_3 = (300 - 70) \left[1 - \exp\left(-\frac{3}{\tau}\right) \right] + 70 = 200^\circ\text{C} \quad (\text{given})$$

or

$$\exp\left(-\frac{3}{\tau}\right) = \frac{230 - 130}{230} = \frac{10}{23}$$

or

$$-\frac{3}{\tau} = \ln 10 - \ln 23 = -0.8329$$

or

$$\tau \cong 3.6 \text{ s}$$

Since a reliable reading can be obtained at 5τ , the required time is 18 s.

Ramp response

Here $Q_i(s) = A/s^2$. Therefore,

$$Q_o(s) \equiv G(s)Q_i(s) = \frac{KA}{s^2(1 + \tau s)} = KA \left(\frac{1}{s^2} - \frac{\tau}{s} + \frac{\tau^2}{1 + \tau s} \right)$$

On taking inverse Laplace transform, we get

$$q_o(t) = KA \left[t - \tau \left\{ 1 - \exp\left(-\frac{t}{\tau}\right) \right\} \right] \quad (4.14)$$

Measurement error. The measurement error is

$$\begin{aligned}
 e_m &= At - A \left[t - \tau \left\{ 1 - \exp \left(-\frac{t}{\tau} \right) \right\} \right] \\
 &= A\tau \left\{ 1 - \exp \left(-\frac{t}{\tau} \right) \right\} \\
 &= \underbrace{-A\tau \exp \left(-\frac{t}{\tau} \right)}_{\text{transient error}} + \underbrace{A\tau}_{\text{steady-state error}}
 \end{aligned}
 \tag{4.15}$$

Steady-state error. The steady-state error is

$$e_{ss} = \lim_{t \rightarrow \infty} e_m = A\tau$$

The steady-state error obviously depends on τ which means that a small τ instrument is desirable.

Lag. An intriguing revelation is that the instrument reading always lags behind the actual value as if the instrument shows a value what the input was τ seconds ago (see Fig. 4.8). The situation will be clear from Example 4.2

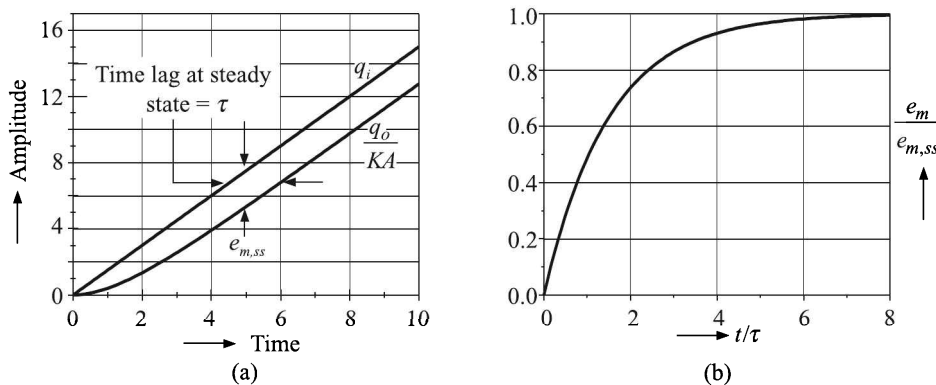


Fig. 4.8 Ramp response of the first order instrument: (a) actual response, and (b) error plot.

Example 4.2

A balloon carrying a first order thermometer ($\tau = 10$ s) rises through the atmosphere at the rate of 10 m/s and records the temperature and altitude readings back to the ground. At 3500 m the balloon says the temperature is 0°C . What is the true altitude at which 0°C occurs, provided the variation of temperature with altitude is 0.15°C per 30 m?

Solution

We assume that

- (i) there is no lag in the altimeter reading,
- (ii) the time lag for radio wave propagation is negligibly small, and
- (iii) the only lag is in the thermometer reading.

This is obviously a case of ramp response, the value of the ramp constant $A = 10$ m/s. So, at any instant the thermometer will read a value which occurred $A\tau$ distance ago. In other words, the lag in terms of distance is 10×10 m = 100 m. So, the true altitude at which 0°C occurs is $(3500 - 100)$ m = 3400 m.

Example 4.3

A first order thermometer was dipped in a temperature-controlled water-bath maintained at 100°C and the following time-temperature readings were obtained:

| | | | | | | |
|----------------------------------|-----|-----|-----|------|------|------|
| Time (s) | 0.0 | 3.0 | 8.0 | 11.0 | 15.0 | 18.0 |
| Temperature ($^\circ\text{C}$) | 20 | 60 | 90 | 95 | 98 | 99 |

Estimate the time constant of the thermometer. Determine the steady-state error when the thermometer is required to measure the temperature of a liquid which is being heated at a constant rate of 0.2°C/s .

Solution

If T_0 is the temperature of the bath ($= 100^\circ\text{C}$), and T is the temperature attained by the thermometer at any instant t , then since it is a first order instrument

$$T = T_0[1 - \exp(-t/\tau)]$$

where, τ is the time constant. From this relation, with re-arrangement of terms and taking logarithm of both sides, we get

$$\ln(T_0 - T) = \ln T_0 - \frac{1}{\tau}t$$

This is a linear relation of the form $y = a_0 + a_1x$ where $y = \ln(T_0 - T)$, $a_0 = \ln T_0$, $a_1 = -1/\tau$ and $x = t$. A least-square fit of the data can be made as follows:

| x | $(T_0 - T)$ | y | x^2 | xy |
|-----------|-------------|----------------|------------|----------------|
| 0 | 80 | 4.3820 | 0 | 0.0 |
| 3 | 40 | 3.6889 | 9 | 11.0667 |
| 8 | 10 | 2.3026 | 64 | 18.4208 |
| 11 | 5 | 1.6094 | 121 | 17.7034 |
| 15 | 2 | 0.6931 | 225 | 10.3965 |
| 18 | 1 | 0.0 | 324 | 0.0 |
| 55 | | 12.6760 | 743 | 57.5874 |

Substituting these values in Eq. (3.12), we get

$$-\frac{1}{\tau} = a_1 = \frac{6 \times 57.5874 - 55 \times 12.676}{6 \times 743 - 55^2} = -0.2454$$

or

$$\tau = 4.075 \text{ s}$$

Therefore, the steady-state error = $T_0\tau = 0.2 \times 4.075 = 0.8^\circ\text{C}$.

Impulse response

Here, $Q_i(s) = A$. Therefore,

$$Q_o(s) = \frac{KA}{1 + \tau s} = \frac{KA}{\tau} \frac{1}{s + (1/\tau)}$$

Taking inverse Laplace transform, we get

$$q_o(t) = \frac{KA}{\tau} \exp\left(-\frac{t}{\tau}\right)$$

It may be noted in this context that the realisation of an impulse input is rather impossible for a physical system. Because, for such an input, at $t = 0$, $q_i(t)$ has an infinite slope and it goes down to zero value at $t = \varepsilon$, where $\varepsilon \rightarrow 0$ (Fig. 4.9).

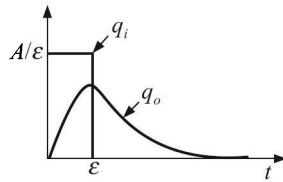


Fig. 4.9 Impulse response of first order instrument.

For any physical system to respond to such an input, it is necessary to transfer energy at an infinite rate which is not feasible. However, the derivative of a step function is an impulse function. Such a situation is observed at the output of a capacitor when a step input is applied to it.

Frequency response

Here,

$$G(j\omega) = \frac{K}{1 + j\omega\tau} = \frac{K}{\sqrt{1 + \omega^2\tau^2}} \angle \tan^{-1}(-\omega\tau)$$

Hence, the amplitude ratio and phase angle are given by

$$\frac{A_o}{A_i} = \frac{K}{\sqrt{1 + \omega^2\tau^2}} \quad (4.16)$$

$$\phi = \tan^{-1}(-\omega\tau) \quad (4.17)$$

The plots of amplitude ratio and phase angle vs. frequency [see Eqs. (4.16) and (4.17)] are shown in Fig. 4.10.

For an ideal frequency response, $\phi \rightarrow 0$, which means $\omega\tau \rightarrow 0$. Thus for a given τ , there will be some ω below which the measurement is accurate. Alternatively, if ω is high, the τ of the instrument needs to be low.

Example 4.4

An input of $2 \sin 3t + 0.5 \sin 30t$ is given to a first order instrument which has a time constant of 0.5 s. What is the response of the instrument?

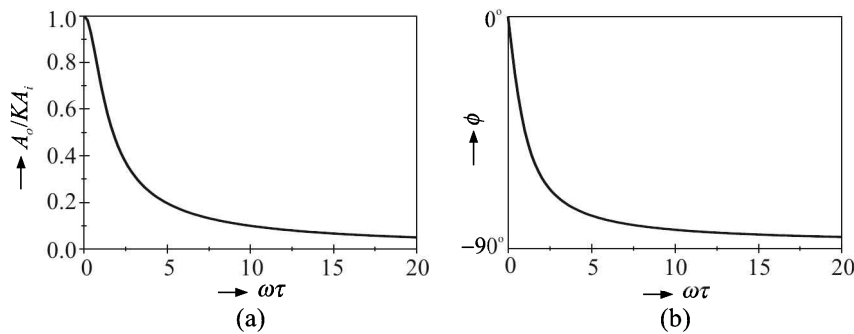


Fig. 4.10 Plots for a first order system: (a) amplitude ratio vs. frequency and (b) phase angle vs. frequency.

Solution

Here,

$$q_i(t) = 2 \sin 3t + 0.5 \sin 30t$$

For $\omega = 3$:

$$\frac{A_o}{A_i} = \frac{1}{\sqrt{1 + 9 \times 0.25}} = 0.5547$$

and

$$\phi = \tan^{-1}(-3 \times 0.5) = -56.3^\circ$$

For $\omega = 30$:

$$\frac{A_o}{A_i} = \frac{1}{\sqrt{1 + 900 \times 0.25}} = 0.0665$$

and

$$\phi = \tan^{-1}(-300 \times 0.25) = -86.2^\circ$$

Incorporating these values in the first equation, we get

$$\begin{aligned} q_o(t) &= 2 \times 0.5547 \sin(3t - 56.3^\circ) + 0.5 \times 0.0665 \sin(30t - 86.2^\circ) \\ &\cong 1.11 \sin(3t - 56^\circ) + 0.033 \sin(30t - 86^\circ). \end{aligned}$$

The input and output waveforms are shown in Fig. 4.11.

It can be seen that not only is there substantial reduction in the amplitude of the higher frequency, but also phase lag in both the frequencies is considerable. So, the instrument fails to measure the given input faithfully.

Fidelity. In the context of findings of Example 4.4, we discuss an important dynamic characteristic of instruments, namely, fidelity.

We have already stated that by fidelity of an instrument we mean its ability to reproduce an input signal faithfully. A dynamic signal, especially a periodic one, is characterised by its

- (i) amplitude
- (ii) frequency, and
- (iii) phase

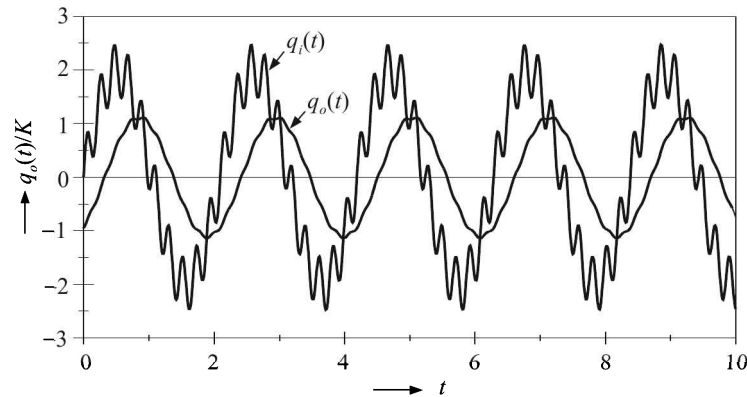


Fig. 4.11 Schematic presentation of frequency response (Example 4.4).

If an instrument reproduces these three characteristics faithfully, i.e. without any distortion, over a desired amplitude and frequency range, it is considered to be of high fidelity.⁵ When we say frequency of the signal, what we actually mean is different frequency components, may be higher harmonics of a base frequency, that are present in the signal. In Example 4.4, the instrument is obviously of low fidelity because its output not only has had a lower amplitude but also is almost shorn of harmonics of the input. The phase is also different.

Speed of response. The low fidelity of the instrument can be attributed to its rather high time constant. A lower value of the time constant is one of the contributing factors to a higher fidelity. We have already discussed the effect of τ on the step response of a first order instrument. There we found that the instrument reaches its steady-state value at nearly 5τ . We will see later that for a critically damped second order instrument, it takes about 8 times the value of the time constant to attain the steady state.

If an instrument reaches its steady-state value quickly, it is said that its *speed of response* is high. We found that a high speed of response is caused by a low time constant of the instrument.

Example 4.5

A first order thermometer is used to measure the temperature of air cycling at a rate of 1 cycle every 5 min. The time constant of the thermometer is 20 s. Calculate the attenuation of the indicated temperature in per cent. If the temperature undergoes a sinusoidal variation of $\pm 20^\circ\text{C}$, calculate the indicated variation in temperature.

Solution

Here, $\tau = 20$ s. $f = 1/5$ min = $1/300$ Hz = 0.0033 Hz. Therefore, amplitude of the indicated temperature variation is

$$\frac{20}{\sqrt{\omega^2\tau^2 + 1}} = \frac{20}{\sqrt{(2\pi \times 0.0033)^2(20)^2 + 1}} = \pm 20 \times 0.92 = \pm 18.45^\circ\text{C}$$

$$\text{Attenuation} = (1 - 0.92) \times 100\% = 7.8\%.$$

⁵ Commonly referred to as *hi-fi*.

4.6 Second Order Instrument

By definition the dynamic relation here is

$$a_2 \frac{d^2 q_o}{dt^2} + a_1 \frac{dq_o}{dt} + a_0 q_o = b_0 q_i$$

which gives

$$\frac{1}{\omega_n^2} \frac{d^2 q_o}{dt^2} + \frac{2\zeta}{\omega_n} \frac{dq_o}{dt} + q_o = K q_i \quad (4.18)$$

where

$$K = b_0/a_0 = \text{static sensitivity}$$

$$\omega_n = \sqrt{a_0/a_2} = \text{natural frequency}$$

$$\zeta = a_1/(2\sqrt{a_0 a_2}) = \text{damping ratio}$$

Taking Laplace transform of Eq. (4.18), we get

$$\frac{s^2}{\omega_n^2} Q_o + \frac{2\zeta s}{\omega_n} Q_o + Q_o = K Q_i$$

which gives the transfer function as

$$G(s) \equiv \frac{Q_o}{Q_i} = \frac{K}{(s^2/\omega_n^2) + (2\zeta s/\omega_n) + 1} = \frac{K\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

A good example of a second order system is provided by the mass-spring-damper system as shown in Fig. 4.12.

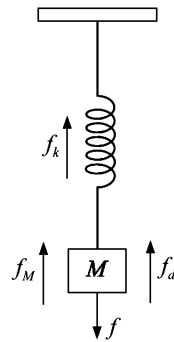


Fig. 4.12 Mass-spring-damper system.

When a force f is applied to the mass M downwards, it is acted upon by

1. An inertial force $f_M = M(d^2x/dt^2)$
2. A restoring force because of the spring stiffness $f_k = kx$
3. Another restoring force owing to viscous friction with the air $f_d = D(dx/dt)$

Therefore, the following dynamic equation may be set up for the system:

$$M \frac{d^2x}{dt^2} + D \frac{dx}{dt} + kx = f \quad (4.19)$$

where, k is the stiffness constant of the spring and D is the damping constant. Comparing Eq. (4.19) with Eq. (4.18), we may write

$$\text{Natural frequency} \quad \omega_n = \sqrt{\frac{k}{M}} \quad (4.20)$$

$$\text{Damping ratio} \quad \zeta = \frac{D}{2\sqrt{kM}} \quad (4.21)$$

Example 4.6

A 2 g mass when suspended from a simple spring, causes a deflection of 5 mm. What is the natural frequency of oscillation of the system?

Solution

Force acting on the mass M downwards = Mg , where g is the acceleration due to gravity. Force acting on it upwards due to the elasticity of the spring = kx , where k is the spring constant and x is the deflection. In equilibrium, these forces balance whence we get $k = Mg/x$.

Therefore, the natural frequency of oscillation is given by

$$f_n = \frac{1}{2\pi} \omega_n = \frac{1}{2\pi} \sqrt{\frac{k}{M}} = \frac{1}{2\pi} \sqrt{\frac{g}{x}} = \frac{1}{2\pi} \sqrt{\frac{980}{0.5}} = 7.05 \text{ Hz}$$

Example 4.7

A d'Arsonval galvanometer, which is a second order instrument, is designed in such a way that its damping ratio is 0.65 and the natural frequency of undamped oscillation is 4 Hz.

- If the sensitivity of the movement is doubled by using a spring of smaller stiffness, calculate the new damping ratio and the new natural frequency.
- If the damping ratio is now restored to its original value by altering the moment of inertia of the system, determine the newer natural frequency of the system.

Solution

This is a rotational system where, according to Newtonian mechanics, the applied torque to the system is balanced by torques owing to (i) inertia, (ii) viscous damping, and (iii) torsion of the spring and the suspension. Thus, we can write

$$T_{\text{appl}}(t) = J \frac{d^2\theta(t)}{dt^2} + D \frac{d\theta(t)}{dt} + \chi\theta(t)$$

where, J is the moment of inertia, D is the damping constant, χ is the stiffness or torsional constant, and $\theta(t)$ is the angular deflection at any instant t . Comparing this equation with Equation (4.18), we get

$$\begin{aligned} \text{Natural frequency } f_n &= \frac{\omega_n}{2\pi} = \frac{1}{2\pi} \sqrt{\frac{\chi}{J}} \\ \text{Damping ratio } \zeta &= \frac{\omega_n D}{2\chi} = \frac{D}{2\sqrt{\chi J}} \\ \text{Sensitivity } K &= \frac{C}{\chi} \end{aligned}$$

where $T_{\text{appl}} = C\theta_i$, C being a constant. From these equations, we get

$$\begin{aligned} \frac{(f_n)_1}{f_n} &= \sqrt{\frac{\chi_1}{\chi}} && \text{if } J \text{ is held constant and } \chi \text{ is changed to } \chi_1 \\ \frac{(f_n)_1}{f_n} &= \sqrt{\frac{J}{J_1}} && \text{if } \chi \text{ is held constant and } J \text{ is changed to } J_1 \\ \frac{\zeta_1}{\zeta} &= \sqrt{\frac{\chi}{\chi_1}} && \text{if } D \text{ and } J \text{ are held constant and } \chi \text{ is changed to } \chi_1 \\ \frac{\zeta_1}{\zeta} &= \sqrt{\frac{J}{J_1}} && \text{if } D \text{ and } \chi \text{ are held constant and } J \text{ is changed to } J_1 \\ \frac{K_1}{K} &= \frac{\chi}{\chi_1} && \text{if } C \text{ is held constant and } K \text{ is changed to } K_1 \end{aligned}$$

(a) Here, $K_1/K = 2$. Hence, $\chi_1 = \chi/2$

$$(f_n)_1 = \frac{f_n}{\sqrt{2}} = \frac{4}{\sqrt{2}} \simeq 2.83 \text{ Hz}$$

and $\zeta_1 = \sqrt{2}\zeta = \sqrt{2} \times 0.65 \simeq 0.92$

(b) Here, $f_n = 2.83 \text{ Hz}$, $\zeta = 0.92$ and $\zeta_1 = 0.65$. Hence

$$\sqrt{\frac{J}{J_1}} = \frac{\zeta_1}{\zeta} = \frac{0.65}{0.92} = \frac{1}{\sqrt{2}}$$

Then $(f_n)_1 = f_n \sqrt{\frac{J}{J_1}} = \frac{2.83}{\sqrt{2}} = 2 \text{ Hz}$

Dynamic Response of Second Order Instruments

We will study dynamic responses of a second order system with respect to standard inputs. Since these derivations are rather tedious than complicated, we show only one of them here, others being similarly done.

Another aspect has to be noted here as well. As we are already familiar with, the method of finding response to different inputs is to multiply the transfer function $G(s)$ by the Laplace transform of the input $Q_i(s)$ and to find the inverse Laplace transform of the product. While performing inverse Laplace transform, we have to factorise the product which, in turn, is related

to finding roots of the corresponding equation. Here in the second order transfer function, the denominator contains a quadratic expression which has two roots for the corresponding quadratic equation. These are

$$s_1, s_2 = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$$

Depending on the value of ζ , three cases present themselves:

1. $\zeta > 1$, called the *overdamped* system
2. $\zeta = 1$, called the *critically-damped* system
3. $\zeta < 1$, called the *underdamped* system

Consequently, when we study the response of second order systems to different standard inputs, we need to consider three cases, as above, for each input.

Step response

The response functions for the three cases are:

Overdamped system. In this case, the roots s_1 and s_2 are real and $s_1 \neq s_2$. Then,

$$\begin{aligned} Q_o(s) &= \frac{K\omega_n^2 A}{s(s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1})(s + \zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1})} \quad (4.22) \\ &\equiv K\omega_n^2 A \left(\frac{a}{s} + \frac{b}{s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1}} + \frac{c}{s + \zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1}} \right) \\ &= K\omega_n^2 A \left[\frac{a(s^2 + 2\zeta\omega_n s + \omega_n^2) + bs(s + \zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1}) + cs(s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1})}{s(s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1})(s + \zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1})} \right] \quad (4.23) \end{aligned}$$

where, a, b and c are arbitrary coefficients to be determined.

From Eq. (4.23) we get the following condition to determine a, b and c :

$$a(s^2 + 2\zeta\omega_n s + \omega_n^2) + bs(s + \zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1}) + cs(s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1}) = 1 \quad (4.24)$$

Comparing Eqs. (4.22) and (4.24), we get

$$\begin{aligned} a\omega_n^2 &= 1 \quad \text{or} \quad a = \frac{1}{\omega_n^2} \\ a + b + c &= 0 \quad \text{or} \quad (b + c) = -a = -\frac{1}{\omega_n^2} \\ (2a + b + c)\zeta\omega_n + (b - c)\omega_n\sqrt{\zeta^2 - 1} &= 0 \quad (4.25) \end{aligned}$$

On subtracting $a + b + c = 0$ from Eq. (4.25), we get

$$a\zeta + (b - c)\sqrt{\zeta^2 - 1} = 0 \quad \text{or} \quad (b - c) = -\frac{\zeta}{\omega_n^2\sqrt{\zeta^2 - 1}}$$

From relations of $(b + c)$ and $(b - c)$, we get

$$b = -\frac{\zeta + \sqrt{\zeta^2 - 1}}{2\omega_n^2 \sqrt{\zeta^2 - 1}}$$

$$c = -\frac{\zeta - \sqrt{\zeta^2 - 1}}{2\omega_n^2 \sqrt{\zeta^2 - 1}}$$

Thus,

$$Q_o(s) = KA \left[\frac{1}{s} - \frac{\left(1 + \zeta/\sqrt{\zeta^2 - 1}\right)/2}{s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1}} - \frac{\left(1 - \zeta/\sqrt{\zeta^2 - 1}\right)/2}{s + \zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1}} \right]$$

which on inverse Laplace transform yields

$$\frac{q_o(t)}{KA} = 1 - \frac{\zeta + \sqrt{\zeta^2 - 1}}{2\sqrt{\zeta^2 - 1}} \exp\left[\left(-\zeta + \sqrt{\zeta^2 - 1}\right)\omega_n t\right] + \frac{\zeta - \sqrt{\zeta^2 - 1}}{2\sqrt{\zeta^2 - 1}} \exp\left[\left(-\zeta - \sqrt{\zeta^2 - 1}\right)\omega_n t\right]$$

Critically-damped system. In this case the response function is

$$\frac{q_o(t)}{KA} = 1 - (1 + \omega_n t) \exp(-\omega_n t)$$

Underdamped system. Here, the response function is

$$\frac{q_o(t)}{KA} = 1 - \frac{\exp(-\zeta\omega_n t)}{\sqrt{1 - \zeta^2}} \sin\left(\sqrt{1 - \zeta^2}\omega_n t + \phi\right) \quad (4.26)$$

where $\phi = \cos^{-1} \zeta$.

We will presently see that this is the most important system of second order instruments, and therefore, we will study the response function in somewhat greater detail. Its graphical presentation for different values of ζ is given in Fig. 4.13.

Measurement error. The dynamic measurement error is

$$e_m = A \left[q_i(t) - \frac{q_o(t)}{KA} \right] = \frac{A \exp(-\zeta\omega_n t)}{\sqrt{1 - \zeta^2}} \sin\left(\sqrt{1 - \zeta^2}\omega_n t + \phi\right)$$

Steady-state error. The steady-state error is

$$e_{ss} = \lim_{t \rightarrow \infty} e_m = 0$$

A knowledge of the different parameters as depicted in Fig. 4.14 is helpful to minimise errors in measurement by second order instruments.

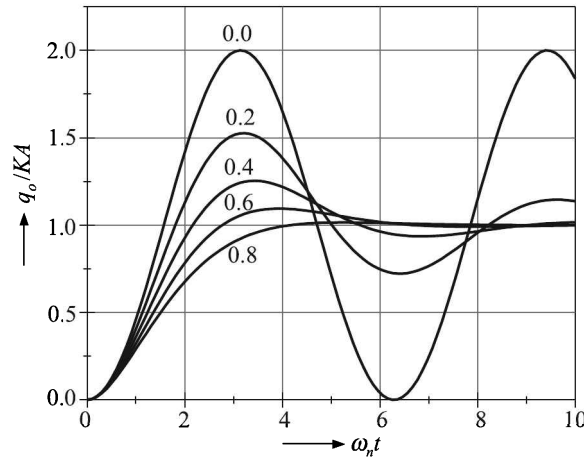


Fig. 4.13 Step response of a second order instrument for different damping ratios.

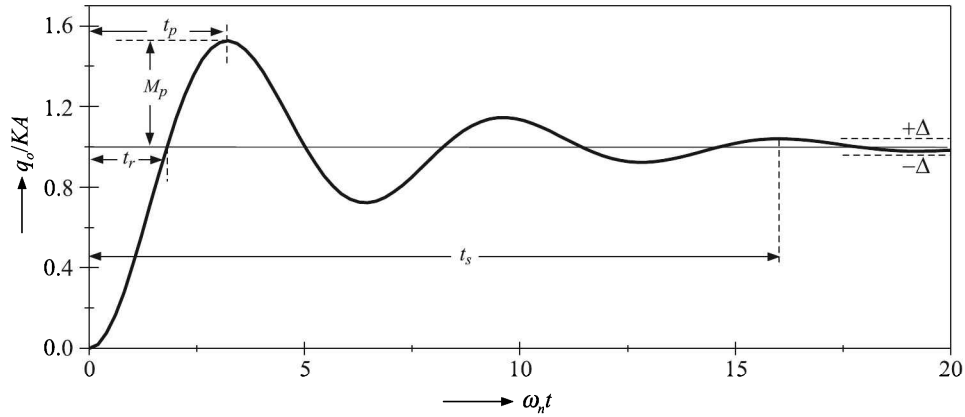


Fig. 4.14 Definition of different parameters for a second order underdamped system.

Rise time t_r is the time required to attain the steady state value for the first time. That means,

$$\left[\sin \left(\sqrt{1 - \zeta^2} \omega_n t + \cos^{-1} \zeta \right) \right]_{t=t_r} = 0 = \sin \pi$$

which gives

$$t_r = \frac{\pi - \cos^{-1} \zeta}{\omega_n \sqrt{1 - \zeta^2}}$$

Peak time t_p is the time required to reach the first peak. At this time the sine-term in Eq. (4.26) has to be negative, because only then the modifying terms to the amplitude A sum up. That, in turn, is achieved when the time-varying quantity in the sine term assumes the value of π radians. Thus,

$$(\sqrt{1 - \zeta^2} \omega_n t)_{t=t_p} = \pi$$

which gives

$$t_p = \frac{\pi}{\omega_n \sqrt{1 - \zeta^2}}$$

Peak overshoot M_p is the fractional value by which the response exceeds the input, or the steady-state response, at the peak time. Thus,

$$M_p = \frac{q_o(t_p) - q_o(\infty)}{q_o(\infty)} = \exp\left(-\frac{\pi\zeta}{\sqrt{1 - \zeta^2}}\right)$$

Settling time t_s is the time required for the output to reach and stay within a tolerance band $\pm\Delta$. This means,

$$\frac{\exp(-\zeta\omega_n t_s)}{\sqrt{1 - \zeta^2}} = \Delta$$

or

$$-\zeta\omega_n t_s = \ln\left(\Delta\sqrt{1 - \zeta^2}\right) \cong \ln\Delta - \frac{\zeta^2}{2}$$

or

$$t_s \cong \frac{\zeta}{2\omega_n} - \frac{\ln\Delta}{\zeta\omega_n} \quad (4.27)$$

- Note:* (a) The higher the value of ζ , the higher is the rise time. It means that it is better to design the instrument with a lower value of ζ .
- (b) But for a given Δ , say 10% or 5%, the settling time is higher for a lower value of ζ , because in Eq. (4.27) the second term dominates and it becomes positive owing to the negative value of the logarithm.
- (c) After the response reaches the steady-state value for the first time, a damped oscillation takes place and then the settling time determines when the output comes to stay within a tolerance band. Therefore, the combined value of $(\omega_n t_r + \omega_n t_s)$ determines what value of ζ to choose. The relevant plot (Fig. 4.15) shows that both for the 5% and 10% tolerance bands, a value of ζ between 0.6 and 0.7 gives the minimum combined settling time.

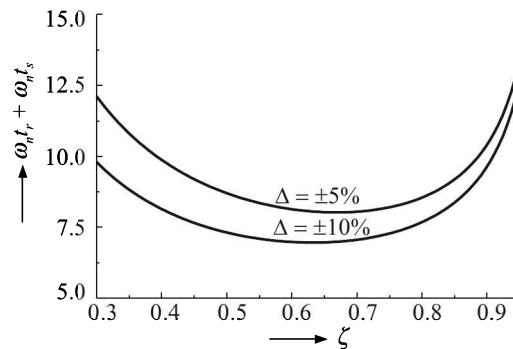


Fig. 4.15 Combined rise time and settling time for 5% and 10% tolerance bands.

This is one of the reasons why commercial second order instruments, such as PMMC galvanometers, are constructed to ensure that ζ lies between 0.6 and 0.7. Other reasons will be discussed while doing the frequency domain analysis.

Ramp response

We will consider only the underdamped system which is of importance to us.

Underdamped system. Here

$$\frac{q_o(t)}{KA} = t - \frac{2\zeta}{\omega_n} + \frac{\exp(-\zeta\omega_n t)}{\sqrt{1-\zeta^2}\omega_n} \sin(\sqrt{1-\zeta^2}\omega_n t + \phi)$$

where $\cos \phi = 1 - 2\zeta^2$.

Dynamic measurement error is

$$\begin{aligned} e_m &= A \left[q_i - \frac{q_o(t)}{KA} \right] \\ &= \frac{2\zeta}{\omega_n} A - \frac{A \exp(-\zeta\omega_n t)}{\sqrt{1-\zeta^2}\omega_n} \sin(\sqrt{1-\zeta^2}\omega_n t + \phi) \end{aligned}$$

and

Steady-state error is

$$e_{ss} = \lim_{t \rightarrow \infty} e_m = \frac{2\zeta}{\omega_n} A$$

Thus, the steady-state error can be reduced by lowering ζ and increasing ω_n . But the fact is that there is little control over ζ because its reduction results in larger oscillations.

Figure 4.16 shows the ramp response of a second order instrument.

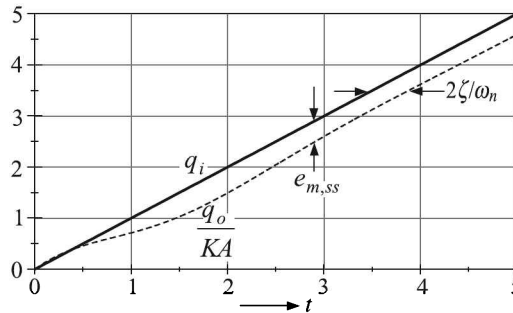


Fig. 4.16 Ramp response of second order instrument.

Like in the ramp response of first order instruments, there is a steady-state time lag in the response in this case as well, and this lag is given by $2\zeta A/\omega_n$.

Note: If the time constant is defined here as $\tau = 2\zeta/\omega_n$, the steady-state error and time lag here are similar to those of first order instruments.

Impulse response

Here, we will consider the critically-damped and underdamped systems which are of importance to us.

Critically-damped system. The response function here is

$$\frac{q_o(t)}{KA} = \omega_n^2 t \exp(-\omega_n t)$$

Underdamped system. In this case the response function can be written as

$$\frac{q_o(t)}{KA} = \frac{\omega_n \exp(-\zeta \omega_n t)}{\sqrt{1-\zeta^2}} \sin(\sqrt{1-\zeta^2} \omega_n t)$$

The plot of the response function for different values of the damping ratio is shown in dimensionless form in Fig. 4.17.

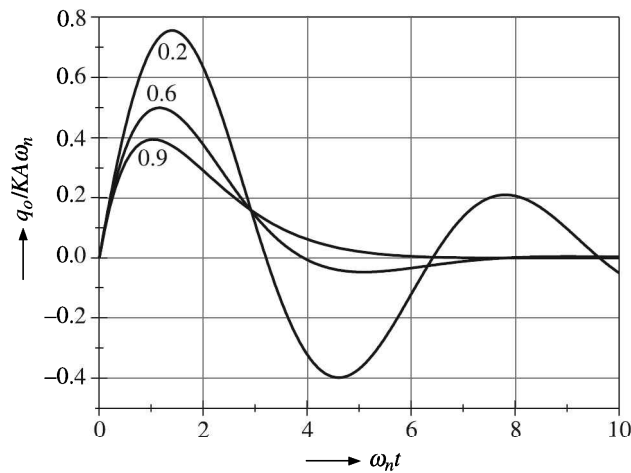


Fig. 4.17 Impulse response of a second order instrument for different values of ζ .

Frequency response

Here, the transfer function is given by

$$\begin{aligned} G(j\omega) &= \frac{K\omega_n^2}{(j\omega)^2 + 2\zeta\omega_n(j\omega) + \omega_n^2} = \frac{K\omega_n^2}{(\omega_n^2 - \omega^2) + 2j\zeta\omega_n\omega} \\ &\equiv \frac{K}{(1 - u^2) + 2j\zeta u} \end{aligned}$$

where, $u = \omega/\omega_n =$ normalised frequency. The amplitude ratio and phase angle are given by

$$M = \frac{1}{\sqrt{(1-u^2)^2 + (2\zeta u)^2}} \quad (4.28)$$

$$\phi = -\tan^{-1} \frac{2\zeta u}{1-u^2} \quad (4.29)$$

Equations (4.28) and (4.29) enable us to generate the following table:

| u | M | ϕ |
|----------|--------------------|------------------|
| 0 | 1 | 0 |
| 1 | $\frac{1}{2\zeta}$ | $-\frac{\pi}{2}$ |
| ∞ | 0 | $-\pi$ |

The plots of M and ϕ vs. normalised frequency u are shown in Figs. 4.18 and 4.19.

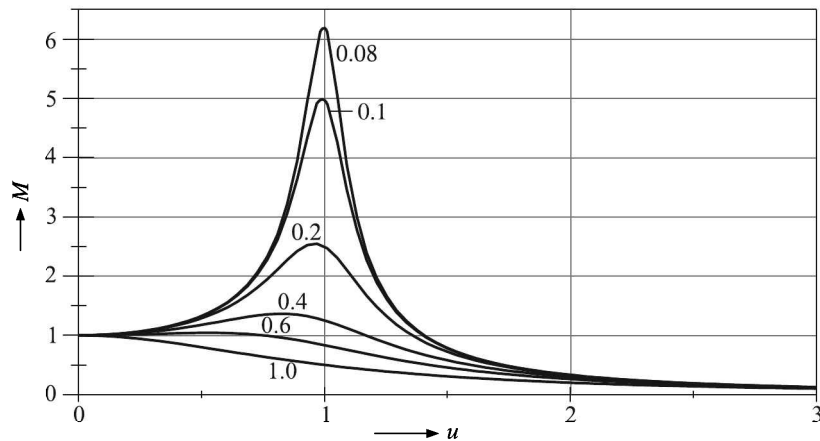


Fig. 4.18 Amplitude vs. normalised frequency plot of the second order instrument for different values of ζ .

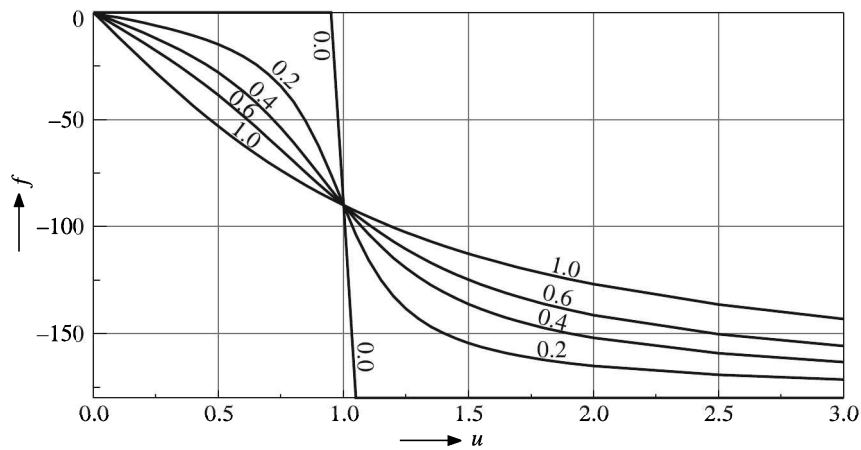


Fig. 4.19 Phase vs. normalised frequency plot of the second order instrument for different values of ζ .

The amplitude ratio plot is familiar to one who has studied the resonance phenomenon in sound, which is basically related to mechanical vibrations. The same phenomenon has an electrical analogue in the case of LCR circuits as well.

From these plots it is clear that

1. The widest flat amplitude ratio exists for ζ lying between 0.6 and 0.7.
2. The phase curves are nearly straight for the widest frequency range when ζ lies between 0.6 and 0.7.

These considerations, as well as the one already discussed (see *Note* at page 102) lead to the widely accepted choice of $\zeta = 0.6$ to 0.7 as the optimum value of damping for second order instruments.

Three terms and their values are important in this context.

Resonant frequency ω_r . is the frequency at which M has its maximum. It can be obtained by partially differentiating M with respect to u and equating the resultant to zero. It is given by

$$\omega_r = \omega_n \sqrt{1 - 2\zeta^2}$$

An interesting consequence offered by this equation is that for $\zeta = 1/\sqrt{2} = 0.707$, $\omega_r = 0$. Hence, for $\zeta \geq 0.707$ there is no resonant frequency.

Resonant peak M_r . is the maximum value of M when $\omega = \omega_r$ and is given by

$$M_r = \frac{1}{2\zeta\sqrt{1 - \zeta^2}}$$

Bandwidth ω_b . is the band of frequencies from zero to half-power point (or cut-off frequency) and is obtained by putting $M = 1/\sqrt{2}$ and taking the larger root of the resulting equation. It is related to natural frequency and damping ratio as follows:

$$\omega_b = \omega_n \sqrt{1 - 2\zeta^2 + \sqrt{2 - 4\zeta^2 + 4\zeta^4}}$$

Example 4.8

Figure 4.20 is a production line device for weighing masses of minimum and maximum values 0.9 kg and 1.1 kg respectively. The mass of the platform is 0.2 kg.

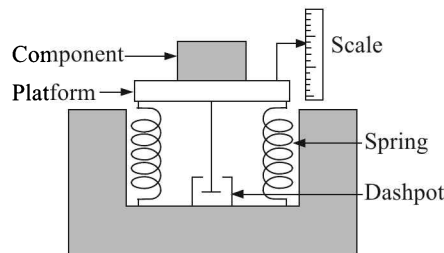


Fig. 4.20 The production line device.

- (a) What should be the spring stiffness if the scale deflection has to be within ± 10 mm over the range?
- (b) If we want to keep the damping ratio at 0.7, what should be the damping constant at the mean mass to be weighed?

- (c) For this damping constant what are the damping ratios corresponding to the minimum and maximum mass to be weighed?
- (d) What are the per cent overshoots when the minimum and maximum mass are suddenly placed on the platform?

Solution

This is a second order system, the relevant dynamics being governed by the equation

$$M \frac{d^2x}{dt^2} + D \frac{dx}{dt} + kx = f$$

where, M is the total mass, D is the damping constant, and k is the spring constant. Comparing this equation with Eq. (4.18) we get, $\zeta = D/2\sqrt{kM}$, and $\omega_n = \sqrt{k/M}$. With this background, we work out the problem as follows:

(a) $k\Delta x = g\Delta m$; Hence, $k = \frac{g\Delta m}{\Delta x} = \frac{9.81 \times 0.2}{0.02} \text{ N/m} = 98.1 \text{ N/m}$

(b) $D = 2\zeta\sqrt{kM} = 2 \times 0.7 \times \sqrt{98.1 \times 1.2} = 15.2 \text{ N}\cdot\text{s/m}$ (or kg/s)

(c) $\zeta_1 = \frac{15.2}{2\sqrt{98.1 \times 1.1}} = 0.73$ and $\zeta_2 = \frac{15.2}{2\sqrt{98.1 \times 1.3}} = 0.67$

(d) We know, $M_p = \exp\left(-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}\right)$. Hence,

$$M_{p1} = \exp\left(-\frac{\pi \times 0.73}{\sqrt{1-0.73^2}}\right) = 0.035 = 3.5\%$$

$$M_{p2} = \exp\left(-\frac{\pi \times 0.67}{\sqrt{1-0.67^2}}\right) = 0.059 = 5.9\%.$$

Example 4.9

A linear, second order, single degree of freedom system has a mass of 4 g and a stiffness of 1000 N/m. Calculate the natural frequency of the system. Determine the damping coefficient necessary to just prevent overshoot in response to a step input.

Solution

From Eqs. (4.20) and (4.21), we get

$$\text{Natural frequency } \omega_n = \sqrt{\frac{1000}{4 \times 10^{-3}}} = 500 \text{ rad/s}$$

$$\text{Damping ratio } \zeta = \frac{D}{2\sqrt{1000 \times 4 \times 10^{-3}}} = \frac{D}{4}$$

If the system has to just prevent the overshoot for a step input, it has to be critically damped, i.e. $\zeta = 1$. This gives, $D = 4 \text{ kg/s}$.

Review Questions

- 4.1 A temperature measuring system, with a time constant of 2 s, is used to measure the temperature of a heating medium which changes sinusoidally between 350°C and 300°C with a time period of 20 s. Find the maximum and minimum values of temperature as indicated by the measuring system and the time lag between the output and input signals.
- 4.2 When a step input of 6 V was applied to a $Y-t$ recorder, the final displacement was 6.0 cm and the final overshoot was at 6.2 cm. Determine the damping ratio of the instrument on this setting.
- 4.3 Show that a mercury-in-glass thermometer, to a reasonable approximation, behaves as a first order instrument.
- 4.4 Study the response of a first order instrument for (a) step, (b) ramp, and (c) impulse inputs.
- 4.5 Show, by graphs only, the response of an underdamped second-order instrument for (a) step, (b) ramp, and (c) impulse inputs.
- 4.6 Commercial second order instruments have a damping ratio between 0.6 and 0.7. Justify this practice from a study of the amplitude ratio and phase angle of the response with respect to the normalised input frequency.
- 4.7 Pressure is abruptly changed from 5 bar to 30 bar at $t = 0$. The transducer (being of first order) indicates a value of 20 bar after 30 seconds. Determine the time required to reach the pressure 95% of the final value.
- 4.8 A linear second order, single degree of freedom system has a mass of 8×10^{-3} kg and a stiffness of 1000 N/m. Calculate the natural frequency of the system. Determine the damping coefficient necessary to just prevent overshoot in response to a step input.
- 4.9 A temperature probe having a first order response with a time constant of 1 s is given a step input of 100°C from 0°C . Calculate the indicated temperature every 0.5 s interval up to 2 s.
- 4.10 The normalised transient response of a first order instrument shows an output of 33% with respect to the steady-state value in 120 ms. When subjected to a cycling input of 1.8 s time period, what per cent of the input magnitude will be indicated by the instrument? Determine also the phase shift introduced by the instrument.
If another similar instrument is cascaded to it, then what will be the natural frequency and damping ratio of the overall system?
- 4.11 A mercury thermometer has a capillary tube of 0.25 mm diameter. If the bulb is made of zero-expansion material, what volume must it have if a sensitivity of $1.5 \text{ mm}/^{\circ}\text{C}$ is desired? Assume volumetric coefficient for mercury = $1.8 \times 10^{-4} /^{\circ}\text{C}$.
- 4.12 A thermometer with a surface area $A = 10^{-3} \text{ m}^2$, specific heat $S = 500 \text{ J/deg C}$, heat transfer coefficient $C = 100 \text{ W/m}^2/\text{deg C}$ is inserted in a body whose temperature is rising from 0 deg C at a rate of 1°C/s .

-
- (a) Develop the dynamical equation connecting the temperature of the body T_i and the temperature indicated T_o .
- (b) Determine the true temperature when the thermometer monitors it as 800°C .
- 4.13 A thrust-producing device is tested in a laboratory by measuring the deflection of a spring element (spring constant $K = 750\text{ N/m}$) attached to the front end of the device. The mass of the device is 25 kg . Assuming that the thrust is idealised as step input to the system
- (a) Calculate the natural frequency of the system
- (b) Calculate the damped natural frequency of the system if the damping ratio is 0.7
- (c) Write the differential equation governing the measuring system.
- 4.14 Fill in the blanks:
- (a) The time constant of a first order instrument having an inductance L and resistance R is equal to ... and is also equal to the time taken to indicate ... per cent of step input.
- (b) Two first order systems are connected in series non-interactively. The response of the system is ... damped, and if more systems are connected in series non-interactively, then the response will even more ...
- (c) A unit step is applied at $t = 0$ to a first order system without time delay. The response has a value of 1.264 units at $t = 10\text{ min}$ and 2 units at steady state. The transfer function of the system is ...
- (d) The thermometer is initially at a temperature of 70°C and is suddenly placed in a liquid which is at 170°C . The thermometer indicates 133.2°C after a time interval of 3 s . The time constant of the thermometer is ...
- 4.15 Indicate the correct choice:
- (a) A temperature probe having a first order response with a time constant of 1 s is given a step input from 50 to 0°C . The temperature in degree centigrade after 0.6 s is
- (i) 18.4
- (ii) 25
- (iii) 27.4
- (iv) 45
- (b) A thermocouple temperature indicator with reference junction at room temperature has a time constant of 1 s . It is dipped in hot bath of 120°C . If the room temperature is 20°C , the thermocouple type temperature indicator will read after 1 s
- (i) 120°C
- (ii) 100°C
- (iii) 63.2°C
- (iv) 140°C

- (c) A first order system
- (i) gives only exponential response
 - (ii) may give oscillatory response for some excitation
 - (iii) always gives a constant value response
 - (iv) always gives a linearly increasing response
- (d) A second order system with zero damping ratio would
- (i) not oscillate at all
 - (ii) oscillate at natural frequency
 - (iii) oscillate with increasing amplitude
 - (iv) oscillate with decreasing amplitude
- (e) For an underdamped second order system subjected to step response, if the rise time is reduced
- (i) the maximum overshoot is also reduced
 - (ii) the maximum overshoot is increased
 - (iii) the maximum overshoot is unaffected
 - (iv) none of the above
- (f) The step response of a given system is

$$y = 1 - 10 \exp(-t)$$

The transfer function of this system is

- (i) $s(s - 9)/(s + 1)$
 - (ii) $(s - 9)/(s + 1)$
 - (iii) $10/(s + 1)$
 - (iv) $(1 - 9s)/(s + 1)$
- (g) The time constant for the following second order system

$$2 \frac{d^2 y}{dt^2} + 4 \frac{dy}{dt} + 8y = 8x$$

is equal to

- (i) 0.5
 - (ii) 1
 - (iii) 2
 - (iv) 3
- (h) The transfer function of a system is the Laplace transform of its
- (i) square wave response
 - (ii) step response
 - (iii) ramp response
 - (iv) impulse response
- (i) The transfer function of a system is given by

$$G(s) = \frac{8}{s^2 + 6s + 8}$$

Then the impulse response function is

- (i) $g(t) = -4 \exp(-4t) + 4 \exp(-2t), \quad t \geq 0$
- (ii) $g(t) = \text{constant}, \quad t \geq 0$
- (iii) $g(t) = -4 \exp(-4t) + 4t \exp(-2t), \quad t \geq 0$
- (iv) none of the above

(j) The Laplace transform of $\exp(-2t) \sin 2t$ is

- (i) $4/(s+2)^2 + 4$
- (ii) $4/(s^2 + 4)$
- (iii) $2/(s^2 + 4s + 8)$
- (iv) $2/(s^2 + 4)$

(k) The solution of the differential equation

$$\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + 2x = 1$$

- (i) is critically-damped
- (ii) is underdamped
- (iii) is overdamped
- (iv) is steady

(l) For the system given by

$$G(s) = \frac{25}{s^2 + 6s + 25}$$

the damping factor and damped frequency of oscillation will be

- (i) 0.6, 4
- (ii) 0.4, 6
- (iii) 0.5, 3
- (iv) 0.3, 5

(m) The steady-state error is determined as the difference between the reference input and the system output at

- (i) $t = t_p$
- (ii) $t = \infty$
- (iii) $t = 0$
- (iv) none of the above

(n) In a second order instrument, K is the spring constant, M is the mass of the instrument and B is the damping constant. The damping constant necessary to prevent the overshoot in the step response is

- (i) equal to \sqrt{KM}
- (ii) less than \sqrt{KM}
- (iii) greater than \sqrt{KM}
- (iv) equal to $2\sqrt{KM}$

- (o) A thermometer possessing a thermal time constant of 0.5 min is introduced in a bath where the temperature is increasing at a constant rate of $5^\circ\text{C}/\text{min}$. The steady state error in the thermometer reading is
- 10°C
 - 2.5°C
 - 0.1°C
 - 0.4°C
- (p) A thermocouple is suddenly immersed in a medium of high temperature bath. The approximate time taken by the thermocouple to reach 98% of the steady-state value is
- equal to the time constant of the thermocouple
 - equal to twice the value of the time constant of the thermocouple
 - equal to four times the value of the time constant of the thermocouple
 - independent of the time constant
- (q) A temperature measuring instrument is modelled as a first order system with a time constant of 5 s. The sensor of the instrument is placed inside an oil bath whose temperature has a sinusoidal variation with amplitude of 10°C and a period of 20 s around an average temperature of 200°C . The sinusoidal component at the output of the instrument will have an amplitude of
- 0°C
 - 5.37°C
 - 8.57°C
 - 10°C
- (r) For a first order instrument, a 5% settling time is equal to
- three times the time constant
 - two times the time constant
 - the time constant
 - time required for the output signal to reach 5% of the final value
- (s) A thermometer at room temperature 30°C is dipped suddenly into a bath of boiling water at 100°C . It takes 30 seconds to reach 96.5°C . The time required to reach a temperature of 98°C is
- 32.5 s
 - 34.6 s
 - 35.6 s
 - 38.6 s
- (t) A thermometer with time constant τ , initially at the ambient temperature, is used to measure the temperature of a liquid in a bath. The excess temperature of the thermometer and the liquid over the ambient are $\theta(t)$ and $\theta_l(t)$ respectively, where t denotes the time. If $\theta_l(t) = kt$, where k is a constant, the static error defined as $\lim_{t \rightarrow \infty} [\theta(t) - \theta_l(t)]$, is
- ∞
 - 0
 - $-k$
 - $-k\tau$

Transducers

Electrical quantities, such as current, voltage, etc. themselves produce electrical signals. Hence their measurements involve proper conditioning of the signals and displaying them in convenient ways. Transducers are seldom necessary in such measurements. Sometimes called *sensors*¹ or *detectors*, transducers more often than not constitute the first stage of an instrumentation set up for the measurement of non-electrical quantities.

A transducer is a device which receives energy in one form or state and transfers it to a convenient form or state. So, transduction is just not conversion of energy from one form to another, although sometimes it may be so. For example, a diaphragm will produce a displacement on the application of pressure. Now, pressure and displacement are both manifestations of mechanical energy, though from the measurement point of view the displacement is more convenient. So, a diaphragm is a pressure transducer although it does not convert energy from one form to another. Again, a junction of dissimilar metals—thermocouple—produces an electrical output with the change of temperature. Here, it is a case of conversion of heat energy to an electrical one, the latter being preferred from the standpoint of convenience of measurement. A thermocouple is, therefore, a temperature transducer.

The transducer, or the responding device can be mechanical, electrical, optical, acoustic, magnetic, thermal, nuclear, chemical or any of their combinations. But, of course, devices with electrical output are preferred for the following reasons:

1. The signal can be conditioned, i.e. modified, amplified, modulated, etc. as desired.
2. A remote operation as well as multiple readout is possible.
3. Devices, such as op-amps are available to ensure a minimal loading of the system.
4. Observer-independent data acquisition and minute control of the process with the help of microprocessors, or for that matter computers, are possible.

5.1 Classification of Transducers

Transducers can broadly be divided into the following categories:

1. Active and passive transducers
2. Analogue² and digital transducers
3. Primary and secondary transducers
4. Direct and inverse transducers

¹Some authors prefer to reserve this word for passive transducers only.

²We will use the British spelling instead of the American spelling *Analog*.

Active and Passive Transducers

Active transducers are self-generating devices, their functioning being based on conversion of energy from one form to another. And since they generate energy themselves, no external source of energy is necessary to excite them. For example, the thermocouple is an active transducer. Depending on their principles of operation, active transducers can be

1. Thermoelectric
2. Piezoelectric
3. Photovoltaic
4. Electromagnetic
5. Galvanic

Table 5.1 gives a rough idea of the use of different kinds of active transducers in the measurement of representative non-electrical properties.

Table 5.1 Active transducers

| <i>Property used</i> | <i>Device</i> | <i>Application in the measurement of</i> |
|---|--|---|
| Thermoelectricity generation | Thermocouple | Temperature |
| | Thermopile | Radiation pyrometry or temperature of distant objects |
| | Thermocouple gauge | Low pressure |
| Piezoelectricity generation | Piezoelectric transducer | Pressure |
| Photoelectricity generation | Photodiode in combination with a diaphragm | Pressure |
| Electricity generation by moving a coil in a magnetic field | Electromagnetic pick-up | Flow |

Passive transducers, on the other hand, do not generate any energy. They need be excited by the application of electrical energy from outside. The extracted energy from the measurand produces a change in their electrical state which can be measured. For example, a photoresistor can be excited by an emf from a cell and the voltage against the photoresistor can be measured. When exposed to a light of certain intensity (measurand) its resistance changes, thus changing the voltage across it.

Depending on their principles of operation, passive transducers can be

1. Resistive
2. Inductive
3. Capacitive
4. Magnetoresistive
5. Photoconductive
6. Thermoresistive
7. Elastoresistive
8. Hall effect-based.

Table 5.2 gives a rough idea of the use of different kinds of passive transducers in the measurement of representative non-electrical properties.

Table 5.2 Passive transducers

| <i>Property used</i> | <i>Device</i> | <i>Application in the measurement of</i> |
|-----------------------|--|---|
| Resistance variation | Potentiometer | Displacement |
| | Strain gauge | Small displacement useful in the measurement of strain, pressure, force, torque |
| | Pirani gauge | Low pressure |
| | Hot-wire anemometer | Flow |
| | Platinum resistance thermometer | Temperature |
| | Thermistor | Temperature |
| | Photoconductive cell or light-dependent-resistor (LDR) in combination with a diaphragm | Pressure |
| Inductance variation | Linear variable differential transformer (LVDT) | Displacement |
| | Synchro | Angular displacement |
| | Eddy-current gauge | Displacement |
| Capacitance variation | Capacitor gauge | Displacement, pressure |
| | Dielectric gauge | Liquid level, thickness (which are basically displacements) |

The lists are not exhaustive but representative. As discussed earlier, we are dealing with electrical transducers only because of their adaptability to instrumentation.

Analogue and Digital Transducers

An *analogue transducer*, such as a CdS cell³ might be wired into a circuit in a way that it will have an output that ranges from 0 volt to 5 volt. The value is continuous between 0 and 5 volt. An analogue signal is one that can assume any value in a range. It works like a tuner on an older radio. We could turn it up or down in a continuous motion. We could fine tune it by turning the knob ever so slightly. Transducers that we have discussed so far generate analogue outputs.

But *digital transducers* generate output in the discrete form. This means that there is a range of values that the sensor can output, but the values increase in steps. Discrete signals typically have a stair step appearance when they are graphed on chart. Consider a modern television set tuner. It allows us to change channels in steps. Or, consider a push button switch. This is one of the simplest forms of sensors. It has two discrete values. Either it is

³Cadmium Sulphide cells measure light intensity.

ON, or it is OFF. Other discrete transducers might provide us with a binary value. Digital displacement encoders⁴ belong to this category.

Primary and Secondary Transducers

A transducer is said to be a *primary transducer* when the applied signal is directly sensed by it. A transducer producing output in the electrical format may be the first element in an instrumentation system. Generally, such sensing elements are called primary transducers .

Sometimes, as for example in pressure measurement, a mechanical sensor senses the input and then another device converts the output of that sensor to an electrical format. There, the latter sensors are called *secondary transducers*.

Direct and Inverse Transducer

A *direct transducer* is a device which receives energy in one form or state and transfers it to an electrical signal. The sensing device can be mechanical, optical, acoustic, magnetic, thermal, nuclear, chemical or any of their combinations.

Inverse transducer is the transducer which converts electrical quantity into a non-electrical quantity. A current carrying coil moving in a magnetic field may be called an inverse transducer because the current carried by it is converted to a force which causes translational or rotational displacement. Many data indicating and recording devices are practically inverse transducers. For example, an analogue ammeter or voltmeter converts current to the mechanical rotation of a pointer, or a speaker in a public address system converts voltage to vibration of air which produces sound.

5.2 A Few Phenomena

Now we will consider a few not so well-known phenomena based on which transducers are constructed. They are:

1. Magnetic effects
2. Piezoelectricity
3. Piezoresistivity
4. Surface acoustic wave
5. Optical effects

Magnetic Effects

All the magnetic effects that are of importance for production of transducers are given in Table 5.3.

Of these effects, magnetoelastic effects—namely, Joule effect, Villari effect, Wiedemann effect and Matteucci effect—and Hall effect are finding more and more use in so-called smart sensors. So we discuss these effects in a little more detail here.

⁴See Section 6.6 at page 216.

Table 5.3 Magnetic effects used in transducers

| <i>Effect</i> | <i>Year of discovery</i> | <i>What it is</i> | <i>Application</i> |
|------------------------------------|--------------------------|---|---|
| Faraday effect | 1831 | Generation of electricity in a coil with the change in the ambient magnetic field | Reluctance based transducers |
| Joule effect (Magnetostriction) | 1842 | Change in shape of a ferromagnetic body with magnetisation | In combination with piezoelectric elements for magnetometers and potentiometers |
| ΔE effect | 1846 | Change in Young's modulus with magnetisation | Acoustic delay line components for magnetic field measurement |
| Matteucci effect | 1847 | Torsion of a ferromagnetic rod in a longitudinal field changes magnetisation | Magnetoelastic sensors |
| Thomson effect | 1856 | Change in resistance with magnetic field | Magneto resistive sensors |
| Wiedemann effect | 1858 | A torsion is produced in a current carrying ferromagnetic rod when subjected to a longitudinal field | Torque and force measurement Displacement measurement Level measurement |
| Villari effect | 1865 | Effect on magnetisation by tensile or compressive stress | Magnetoelastic sensors |
| Hall effect | 1879 | A current carrying crystal produces a transverse voltage when subjected to a magnetic field vertical to its surface | Magnetogalvanic sensors |
| Skin effect | 1903 | Displacement of current from the interior of material to surface layer due to eddy currents | Distance and proximity sensors |
| Josephson effect | 1962 | Quantum tunnelling between two superconducting materials with an extremely thin separating layer | SQUID magnetometers |

Magnetoelastic effects

Various aspects of the coupling between the magnetisation of the ferromagnetic materials and their elasticity can be employed to sense parameters of interest. Several effects which have application for sensing are

| <i>Direct effect</i> | <i>Inverse effect</i> |
|----------------------|-----------------------|
| Joule effect | Villari effect |
| Wiedemann effect | Matteucci effect |

We discuss these effects briefly here.

Joule⁵ effect. The Joule effect, the first of the magnetoelastic effects discovered in 1842, is a change in length due to an applied magnetic field. A transverse change in length and the associated volumetric change are also observed.

The change in the shape of a material due to a change in its magnetisation is also called *Magnetostriction*.

The mechanism of magnetostriction at an atomic level is relatively a complex subject but on a macroscopic level may be segregated into two distinct processes:

1. The first process is dominated by the migration of domain walls within the material in response to external magnetic fields.
2. The second is the rotation of the domains.

These two mechanisms allow the material to change the domain orientation which in turn causes a dimensional change. Since the deformation is isochoric⁶ there is an opposite dimensional change in the orthogonal direction. Although there may be many mechanisms to the reorientation of the domains, the basic idea, represented in Fig. 5.1, remains that the rotation and movement of magnetic domains cause a physical length change in the material.

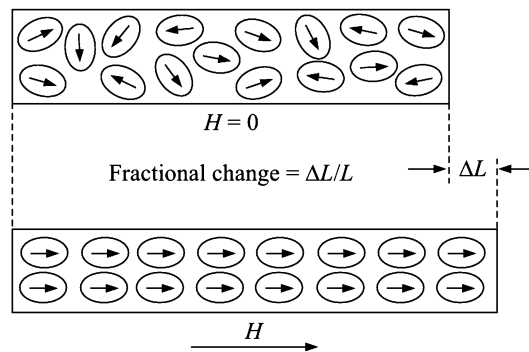


Fig. 5.1 Alignment of magnetic domains in a material due to a magnetic field H that causes a change in its length ΔL (magnetostriction).

We hear a humming sound emitted from a transformer or a fluorescent tube choke. This is caused by magnetostriction. 50 Hz ac generates magnetic fields in transformers causing the core to change the maximum length twice per cycle thus producing the familiar and sometimes annoying 100 Hz (or higher harmonics) hum.

Villari⁷ effect. The Joule effect has an important inverse effect known as the *Villari effect*. A stress induced in the material causes a change in the magnetisation. This change in magnetisation can be sensed, and once calibrated, used to measure the applied stress or force.

The *Villari reversal* is the change in sign of the magnetostriction coefficient as it happens in iron (crystal, 100 axis) from positive to negative when exposed to magnetic fields of approximately 40,000 A-turn/m (500 oersted).

⁵ James Prescott Joule (1818–1889) was an English physicist.

⁶ An *isochoric* process, aka a *constant-volume* process, or an *isovolumetric* process, is a thermodynamic process during which the volume of the closed system undergoing such a process remains constant.

⁷ Named after an Italian physicist E Villari (1836–1904).

The Joule effect and the Villari effect are both utilised in producing magnetostrictive displacement sensors (see Section 6.5) and level sensors (see Section 12.1).

Wiedemann⁸ effect. A wire made of a magnetostrictive material exhibits an important characteristic known as the *Wiedemann effect* (Fig. 5.2). When an axial magnetic field is applied to a magnetostrictive wire, and a current is passed through the wire, a twisting occurs at the location of the axial magnetic field. The twisting is caused by the interaction of the axial magnetic field, usually from a permanent magnet, with the magnetic field along the magnetostrictive wire, which is present due to the current in the wire.

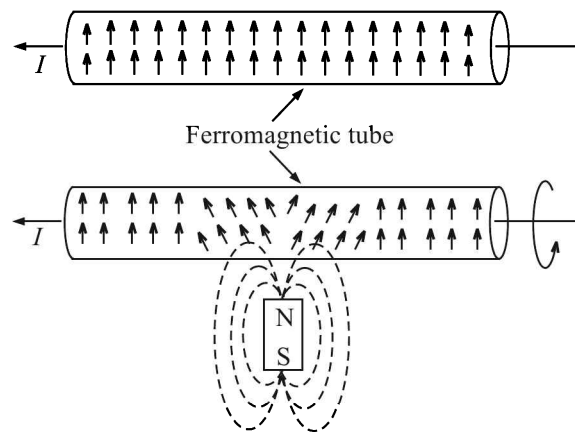


Fig. 5.2 Twisting of a current carrying ferromagnetic tube when subjected to an axial field of a bar magnet.

Matteucci⁹ effect. When a wire made of a magnetostrictive material or a magnetised wire is twisted, its magnetisation (i.e. magnetic permeability) changes. The change in magnetisation can be measured and related to the external torque that twisted it. This inverse of the Wiedemann effect, known as the *Matteucci effect*, is used for magnetoelastic torque sensors.

Magnetostriction coefficient. The magnetostriction coefficient λ is defined as the fractional change in length as the magnetisation increases from zero to its saturation value. Thus,

$$\lambda = \frac{\Delta L}{L} \Big|_{\text{saturated } \mathbf{B}}$$

where, L is the original length of the material and ΔL is the change in length.

If the material expands, the magnetostriction is considered *positive* and if it contracts, it is *negative*. While an iron bar shows positive magnetostriction, a nickel bar shows negative magnetostriction.

Thus, the coefficient λ may be positive or negative and is usually on the order of 10^{-5} . The coefficients λ of common ferromagnetic materials are given in Table 5.4.

⁸Gustav Heinrich Wiedemann (1826–1899) was a German physicist. He was also a litterateur.

⁹Carlo Matteucci (1811–1868) was an Italian physicist and neurophysiologist who was a pioneer in the study of bioelectricity.

Table 5.4 Magnetostriction coefficients for different materials

| <i>Material</i> | <i>Crystal axis</i> | <i>Magnetostriction coefficient</i> $\lambda (\times 10^{-5})$ |
|---|---------------------|---|
| Iron | 100 | +1.1 to +2.0 |
| | 111 | -1.3 to -2.0 |
| | Polycrystalline | -0.8 |
| Nickel | 100 | -5.0 to -5.2 |
| | 111 | -2.7 |
| | Polycrystalline | -2.5 to -4.7 |
| Cobalt | Polycrystalline | -5.0 to -6.0 |
| Terfenol-D ($\text{Tb}_x\text{Dy}_{1-x}\text{Fe}_y$) | Polycrystalline | 2000 |

It is seen from Table 5.4 that cobalt exhibits the largest room temperature magnetostriction of a pure element at 60 microstrain. Among alloys, the highest known magnetostriction is exhibited by Terfenol-D¹⁰. It exhibits about 2000 microstrains in a field of 2 kOe (160 kA-turn/m) at room temperature and is the most commonly used engineering magnetostrictive material.

In the case of a single stress σ applied on a single magnetic domain, the magnetic strain energy density E_σ can be expressed as:

$$E_\sigma = \frac{3}{2}\lambda_s\sigma \sin^2\theta$$

where, λ_s is the magnetostrictive coefficient at saturation

θ is the angle between the saturation magnetisation and the stress direction

For λ_s and $\sigma > 0$ (like in iron under tension), E_σ is minimum for $\theta = 0$ i.e., when the tension is aligned with the saturation magnetisation. Consequently, the magnetisation is increased by tension.

This elastic strain energy associated with the deformation, leads to dissipation of energy in transformer cores in the form of sound.

Applications. The existence of both direct and reciprocal Joule and Wiedemann effects leads to two modes of operation for magnetostrictive transducers:

1. Transferring magnetic energy to mechanical energy
2. Transferring mechanical energy to magnetic energy

The first mode is used to design

1. Actuators for generating motion and/or force
2. Sensors for detecting states of magnetic field

¹⁰Ter for terbium, Fe for iron, NOL for Naval Ordnance Laboratory (USA), and D for dysprosium.

The second mode is used to design

1. Sensors for detecting motion and/or force
2. Devices for inducing change in the magnetic state of a material
3. Passive damping devices, which dissipate mechanical energy as magnetically and/or electrically induced thermal losses

As with many other transducer technologies such as electromagnetic (moving coils) and piezoelectricity, a magnetostrictive transducer has the ability to both actuate and sense simultaneously. Applications such as the telephone, scanning sonar and others make use of this dual mode. For example, a Terfenol-D sonar transducer can be used as either a transmitter or a receiver or both at the same time. Another potential use of dual mode operation is in active vibration and acoustic control. One transducer can be used to sense deleterious structural vibrations and provide the actuation force to suppress them.

It is also utilised to produce ultrasonic vibrations either as a sound source or as ultrasonic waves in liquids which can act as a cleansing agent in ultrasonic cleaning devices.

Hall effect

When a current-carrying conductor is placed into a magnetic field \mathbf{B} , a voltage V_H is generated perpendicular to both the current \mathbf{I} and the field. This phenomenon is known as the *Hall effect*. Written mathematically,

$$V_H \propto \mathbf{I} \times \mathbf{B}$$

The phenomenon originates from the action of the Lorentz¹¹ force on the moving charge carriers in the conductor.

If e is the electronic charge

\mathbf{v} is the velocity of carriers

\mathbf{E} is the electric field

then the Lorentz force \mathbf{F} experienced by charge carriers due to the combined electric and magnetic fields is given by

$$\mathbf{F} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (5.1)$$

The second factor on the RHS of Eq. (5.1) is responsible for the generation of Hall voltage.

Figure 5.3 illustrates the basic principle of the Hall effect. It shows a thin sheet of semiconducting material (Hall element) through which a current is passed. The output connections are perpendicular to the direction of current.

The current consists of the movement of many small charge carriers— typically electrons, holes, ions or all three. When no magnetic field is present [Fig. 5.3(a)], the charges follow approximately straight paths between collisions with impurities, phonons¹², etc. As a result, the current distribution through the material is uniform and no potential difference is seen across the output. However, when a perpendicular magnetic field is applied, moving charges experience a Lorentz force. As a result, their paths between collisions are curved as shown in Fig. 5.3(b). So, the moving charges accumulate on one face of the material. This leaves

¹¹Named after the Dutch physicist Hendrik Antoon Lorentz (1853—1928) who first formulated it. He was awarded the Nobel Prize in physics in 1902.

¹²Quantised lattice vibrations.

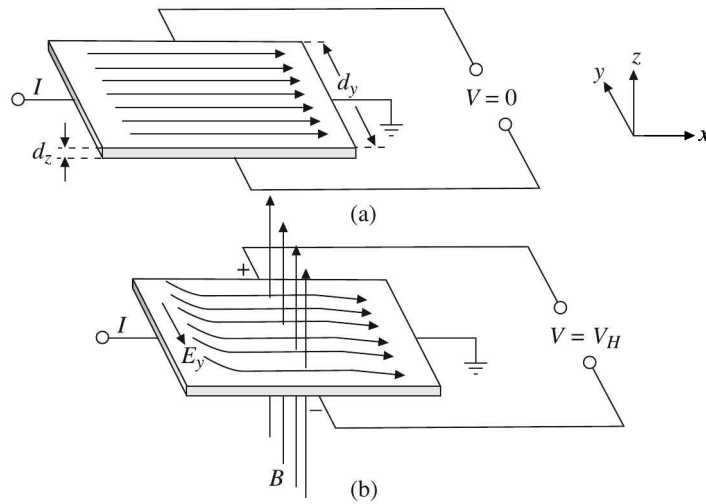


Fig. 5.3 Hall voltage generation principle: (a) no magnetic field and (b) magnetic field present.

equal and opposite charges exposed on the other face, where there is a scarcity of mobile charges. The result is an asymmetric distribution of charge density across the Hall element that is perpendicular to both the current path and the applied magnetic field. The separation of charge establishes an electric field that impedes the migration of further charge so that a steady electrical potential difference builds up for as long as the charge is flowing. This potential difference or voltage is the *Hall*¹³ voltage (V_H).

Hall effect in metals. For a simple metal, where only electrons are the charge carriers, the Hall voltage is given by

$$V_H = -\frac{IB}{ned_z} \quad (5.2)$$

where n is the charge carrier density and d_z is the thickness of the Hall element. The Hall coefficient is defined as

$$R_H = \frac{E_y}{J_x B} \quad (5.3)$$

where J_x is the current density in the x -direction of the carrier electrons and E_y is the generated Hall electric field. Now,

$$\begin{aligned} J_x &= \frac{I}{d_y d_z} \\ E_y &= \frac{V_H}{d_z} \end{aligned} \quad (5.4)$$

where d_y and d_z are dimensions of the Hall element in y and z directions respectively. Therefore, we have from Eqs. (5.2), (5.3) and (5.4)

$$R_H = \frac{V_H d_z}{IB} = -\frac{1}{ne} \quad (5.5)$$

¹³Named after its discoverer Edwin Herbert Hall (1855–1938), an American Physicist.

But then, Hall coefficients for metals are too low to serve any useful purpose. A few representative values are given in Table 5.5.

Table 5.5 Hall coefficients at room temperature for metals

| <i>Metal</i> | <i>Hall coefficient</i> (m^3/C) ^a |
|--------------|--|
| Gold | -0.72 |
| Copper | -0.55 |
| Aluminium | -0.30 |
| Magnesium | -0.94 |
| Tin | -0.04 |

^a Source: *American Institute of Physics Handbook*, New York (1985).

Therefore although discovered in 1879, the Hall effect found its first applications with the advent of semiconducting materials in the 1950s. The Hall voltage in semiconductors is appreciable.

Hall effect in semiconductors. In a semiconductor, there are both negative and positive charge carriers namely, electrons and holes. Let us consider a semiconducting Hall element.

If n is the concentration of electrons

p is the concentration of holes

μ_e is the drift mobility of electrons

μ_h is the drift mobility of holes

E is the electrostatic field

then,

$$\begin{aligned} \text{the drift velocity of electrons } v_e &= \mu_e E \\ \text{the drift velocity of holes } v_h &= \mu_h E \end{aligned} \quad (5.6)$$

Now, the net electrostatic force F acting on a single electron is given by

$$F = eE \quad (5.7)$$

With the help of Eqs. (5.6) and (5.7), we get

$$\left. \begin{aligned} v_e &= \frac{\mu_e}{e} F \\ v_h &= \frac{\mu_h}{e} F \end{aligned} \right\} \quad (5.8)$$

We note that since both holes and electrons are present in the sample, both charges experience a Lorentz force in the same direction because they will be drifting in opposite directions as shown in Fig. 5.4.

Thus, both electrons and holes tend to accumulate near the bottom surface though the magnitudes of Lorentz forces will be different because drift mobilities, and hence drift velocities, will be different for electrons and holes. In equilibrium, there will be no current flowing in

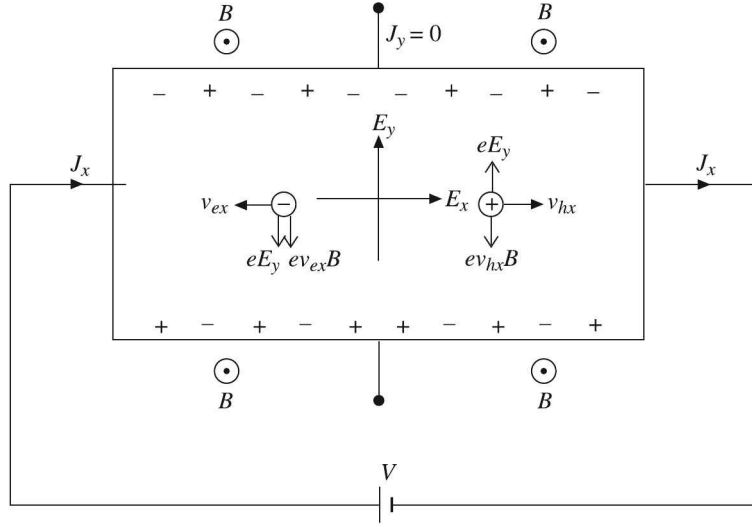


Fig. 5.4 Schematic of drift velocities and forces experienced by holes \oplus and electrons \ominus in an ambipolar Hall element. The magnetic field B is out from the plane of the paper.

the y -direction as we have an open output circuit. Let us assume that more holes accumulate near the bottom surface so that an electric field E_y builds up thereby. Since there exists no current in the y -direction, we have

$$J_y = J_h + J_e = epv_{hy} + env_{ey} = 0 \quad (5.9)$$

where v_{hy} and v_{ey} indicate drift velocities of holes and electrons in the y -direction respectively. Eq. (5.9) yields

$$p v_{hy} = -n v_{ey} \quad (5.10)$$

We note that the electrons and holes experience the following Lorentz forces:

$$\left. \begin{aligned} F_{ey} &= -eE_y - ev_{ex}B \\ F_{hy} &= eE_y - ev_{hx}B \end{aligned} \right\} \quad (5.11)$$

where v_{ex} and v_{hx} indicate drift velocities of electrons and holes in the x -direction respectively. Now, from Eq. (5.8), we can write

$$\left. \begin{aligned} F_{ey} &= -\frac{ev_{ey}}{\mu_e} \\ F_{hy} &= \frac{ev_{hy}}{\mu_h} \end{aligned} \right\} \quad (5.12)$$

Combining Eqs. (5.11) and (5.12), we get

$$\left. \begin{aligned} \frac{ev_{ey}}{\mu_e} &= eE_y + ev_{ex}B \\ \frac{ev_{hy}}{\mu_h} &= eE_y - ev_{hx}B \end{aligned} \right\} \quad (5.13)$$

Substituting $v_{ex} = \mu_e E_x$ and $v_{hx} = \mu_h E_x$, we get from Eq. (5.13)

$$\begin{aligned} \frac{v_{ey}}{\mu_e} &= E_y + \mu_e E_x B \\ \Rightarrow v_{ey} &= \mu_e E_y + \mu_e^2 E_x B \end{aligned} \quad (5.14)$$

$$\begin{aligned} \frac{v_{hy}}{\mu_h} &= E_y - \mu_h E_x B \\ \Rightarrow v_{hy} &= \mu_h E_y - \mu_h^2 E_x B \end{aligned} \quad (5.15)$$

Substituting for v_{ey} and v_{hy} from Eqs. (5.14) and (5.15) in Eq. (5.10), we obtain

$$\begin{aligned} p \mu_h E_y - p \mu_h^2 E_x B &= -n \mu_e E_y - n \mu_e^2 E_x B \\ \Rightarrow E_y (p \mu_h + n \mu_e) &= B E_x (p \mu_h^2 - n \mu_e^2) \\ \Rightarrow E_x &= \frac{E_y}{B} \frac{p \mu_h + n \mu_e}{p \mu_h^2 - n \mu_e^2} \end{aligned} \quad (5.16)$$

Now, let us consider the current flow in the x -direction. Here the total current density is finite and is given by the following equation:

$$J_x = ep v_{hx} + env_{ex} = (p \mu_h + n \mu_e) e E_x \quad (5.17)$$

Substituting the value of E_x from Eq. (5.16) in Eq. (5.17), we get after a little algebraic manipulation

$$e E_y (p \mu_h + n \mu_e)^2 = B J_x (p \mu_h^2 - n \mu_e^2)$$

So, from the definition of Hall coefficient [Eq. (5.3)] we get for ambipolar semiconductors

$$R_H = \frac{p \mu_h^2 - n \mu_e^2}{e (p \mu_h + n \mu_e)^2} = \frac{p - nb^2}{e(p + nb)^2} \quad (5.18)$$

where $b = \mu_e/\mu_h$. The following conclusions can be drawn from Eq. (5.18):

1. R_H depends on both the drift mobility ratio and the concentrations of holes and electrons
2. R_H is positive for $p > nb^2$
3. R_H is negative for $p < nb^2$
4. Equation (5.5) for metals is obtained by substituting $p = 0$

Let us now work out an example to see what is the value of the Hall coefficient of a semiconductor.

Example 5.1

The following are the data for the intrinsic Si: $n = p = n_i = 1.5 \times 10^{10} \text{ cm}^{-3}$, $\mu_h = 450 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$, and $\mu_e = 1350 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$. Calculate R_H .

Solution

From the given data we have

$$\begin{aligned} n = p &= 1.5 \times 10^{16} \text{ m}^{-3} \\ b &= \frac{\mu_e}{\mu_h} = \frac{1350 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}}{450 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}} = 3 \end{aligned}$$

Therefore,

$$R_H = \frac{(1.5 \times 10^{16}) - (1.5 \times 10^{16})(3)^2}{(1.6 \times 10^{-19})[(1.5 \times 10^{16}) + (1.5 \times 10^{16})(3)^2]} \text{ m}^3/\text{C}$$

$$= -208.3 \text{ m}^3/\text{C}$$

Thus, it is evident that the Hall coefficient of a semiconductor is orders of magnitude higher than that for typical metals. This is why all Hall devices use a semiconductor rather than a metal element.

Basic Hall Effect Sensor. Hall effect sensors convert a magnetic field to a useful electrical signal. When physical quantities, like position, speed, temperature, etc. other than a magnetic field are sensed, the magnetic system converts this physical quantity to a magnetic field which, in turn, can be sensed by Hall effect sensors. The block diagram in Fig. 5.5 illustrates this concept.

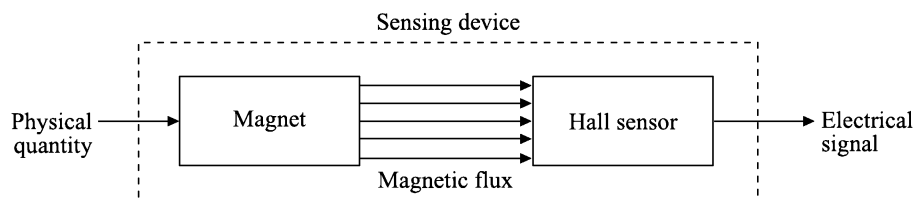


Fig. 5.5 Basic Hall effect sensor.

Many physical parameters can be measured by inducing the motion of a magnet. For example, both temperature and pressure can be sensed through the expansion and contraction of a bellows to which a magnet is attached. We will discuss them in detail in appropriate chapters.

Elimination of piezoresistivity. Silicon, the basic semiconductor which is used to construct Hall elements, exhibits piezoresistivity¹⁴. This interferes with the Hall sensing. It is necessary to minimise this effect in a Hall sensor. This is accomplished by using multiple Hall elements and orienting them on the IC to minimise the effect of stress. Figure 5.6 shows two Hall elements located in close proximity on an IC. They are positioned in this manner so that they may both experience the same packaging stress. The first Hall element has its excitation applied along the vertical axis and the second along the horizontal axis. Summing the two outputs eliminates the signal due to stress. Micro-switch Hall ICs use two or four elements.

Signal conditioning. The Hall element, which is the basic magnetic field sensor, requires signal conditioning to make the output usable for most applications. The signal conditioning electronics comprises

1. An amplifier stage
2. A temperature compensator
3. A voltage regulator when operating from an unregulated supply

¹⁴See Section 5.2 at page 150.

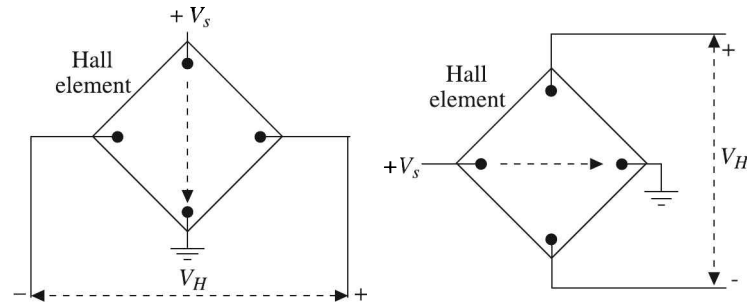


Fig. 5.6 Orientation of Hall elements.

Figure 5.7 illustrates the signal conditioning of a basic Hall effect sensor.

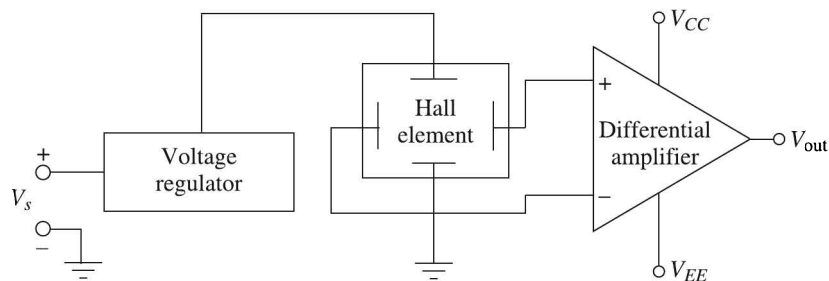


Fig. 5.7 Schematic diagram of a basic Hall effect sensor.

We know that the Hall voltage is zero when no magnetic field is present. However, if voltage at each output terminal is measured with respect to ground, a non-zero voltage will appear. This is the common mode voltage, and is the same at each output terminal. Obviously the *potential difference* is zero. The amplifier shown in Fig. 5.7 is thus a differential amplifier which amplifies only the potential difference, i.e. the Hall voltage.

The generated Hall voltage is on the order of 30 microvolts when subjected to a magnetic field of one gauss. This low-level output requires an amplifier with low noise, high input impedance and moderate gain. An op-amp differential amplifier with these characteristics can be readily integrated with the Hall element. Temperature compensation is also easily integrated (not shown in the diagram).

We know that the Hall voltage is a function of the input current. The purpose of the voltage regulator in Fig. 5.7 is to hold this current constant so that the output of the sensor only reflects the intensity of the magnetic field. Since the regulated power supply is available in many places, some Hall effect sensors do not include an internal regulator.

Zero elevation. According to the orientation of the sensed magnetic field, the output of the amplifier may be driven either positive or negative, thus requiring both plus and minus power supplies. To obviate the requirement for two power supplies, a fixed *offset* or bias is introduced into the differential amplifier. The bias value appears on the output when no magnetic field is present and is referred to as a null voltage. When a positive magnetic field is sensed, the output increases above the null voltage. Conversely, when a negative magnetic field is sensed,

the output decreases below the null voltage, but remains positive. This is akin to the *zero elevation* used in pressure measurement¹⁵.

To further increase the interface flexibility of the device, an open emitter, open collector, or push-pull transistor is added to the output of the differential amplifier. Analogue output sensors are commercially available in ranges of 4.5 to 10.5 V, 4.5 to 12 V, or 6.6 to 12.6 V dc. They typically require a regulated supply voltage to operate accurately.

Transfer characteristic of a Hall sensor. The transfer characteristic of a device relates its output to its input. It can be expressed in the form of either an equation or a graph.

For analogue output Hall effect sensors, the transfer characteristic expresses the relationship between the magnetic field (gauss) input and the voltage output. For a typical analogue output sensor, it is illustrated in Fig. 5.8.

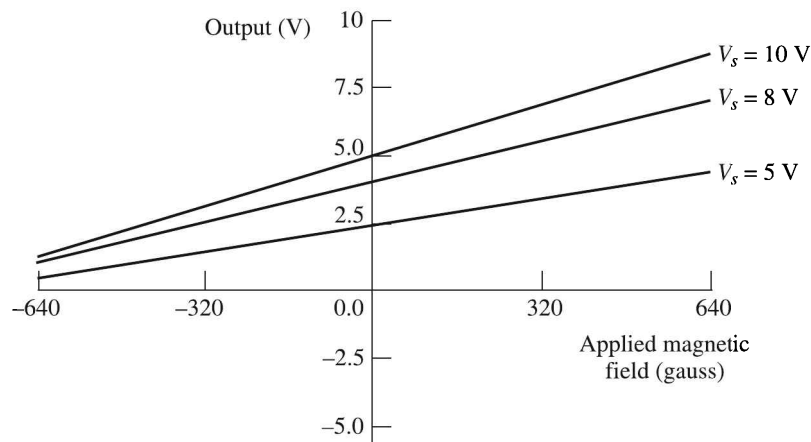


Fig. 5.8 Transfer characteristics of a typical Hall device.

For $-640 < B$ (gauss) < 640 , the transfer characteristic equation for this particular sensor is given by

$$V_{\text{out}} \text{ (volt)} = 6.25 \times 10^{-4} V_s B + 0.5 V_s \quad (5.19)$$

where, B is the magnetic field strength and V_s , the supply voltage.

The transfer characteristic of a Hall sensor is specified by the following properties:

1. Sensitivity
2. Null offset
3. Span

Sensitivity. By definition, the sensitivity is the ratio of the change in output resulting from a given change in input. In this case, it is

$$S = \left| \frac{\Delta V_{\text{out}}}{\Delta B} \right| = 6.25 \times 10^{-4} V_s \text{ volt/gauss) [From Eq. (5.19)]}$$

¹⁵See Section 12.1 at page 502.

Here, we have considered the Hall sensor as a whole comprising the Hall element and the electronics.

Another term, *cross* or *secondary sensitivity* is often mentioned. That indicates the sensitivity of the Hall element with respect to the variation of parameters like temperature or pressure.

Null offset. The null offset is the output from a sensor with no magnetic field excitation. Therefore, substituting $B = 0$ in Eq. (5.19), we get

$$\text{Null offset} = 0.5V_s$$

The imperfection in the fabrication process of the sensor may give rise to the null offset.

Span. The span, defined as the difference between the maximum output and the minimum output, is

$$\text{Span} = V_{\text{out}}|_{\max B} - V_{\text{out}}|_{\min B}$$

for a given supply voltage V_s .

Magnetoresistance

The magnetoresistance effect is closely associated with Hall effect transducers.

Suppose, in Fig. 5.3 if d_y of the Hall element is made much shorter than d_x , the Hall voltage can be almost short-circuited. As a consequence, the charge carriers move at the Hall angle to the x -direction. The increase in path length for the carriers causes an increase in resistance of the device. This increase in resistance of the Hall element owing to the application of a magnetic field is known as the geometrical *magnetoresistance effect*.

A magnetic field applied to a current-carrying conductor causes deviation of some charge carriers from their path. So, when a Hall voltage grows, there is a current decrease, which results in an increased electric resistance. In most conductors this magnetoresistive effect is of a second order when compared to the Hall effect. But in anisotropic materials, such as ferromagnetics, their resistance depends on their state of magnetisation. Then the effect of an external applied magnetic field is more pronounced and the resistance varies from 2 to 5%. The relation between change in resistance and the magnetic field intensity is not linear but quadratic; however, it is possible to linearise it by using biasing methods. If we ignore this need for linearisation and their thermal dependence, magnetoresistors offer the following advantages as compared to other magnetic sensors:

1. A magnetoresistor is a zero order system while inductive sensors are first order systems because their response depends on the time derivative of the magnetic flux density.
2. Hall effect sensors also are zero order systems. But magnetoresistors show increased sensitivity, temperature range, and frequency passband (from dc to several megahertz) compared with 25 kHz for Hall effect sensors.

Construction. Magnetoresistors are usually manufactured from permalloy, which is an alloy of approximately 20% iron and 80% nickel. Also Ni-Fe-Co and Ni-Fe-Mo alloys have been tried.

Application. Depending upon the principle, the applications can be divided into two groups as shown in Table 5.6.

Table 5.6 Applications of magnetoresistive effect

| <i>Principle</i> | <i>Application</i> |
|---|--|
| Direct measurement of magnetic fields | Magnetic audio recording Reading machines for credit cards, magnetically coded price tags |
| Measuring magnetic field variation ^a | Measurement of linear and angular displacement Proximity switches Position measurement Angular velocity of ferrous gear wheels |

^a To accomplish this, it must be either a metallic object or an object with a metallic coating or an identifier placed in a constant magnetic field, or the moving element to be detected must incorporate a permanent magnet.

Piezoelectricity

Certain materials, especially the crystalline ones, produce an emf when deformed by an application of pressure along the specific axes. The phenomenon is known as *piezoelectricity*¹⁶ or *piezoelectric effect* and is widely used for the construction of many transducers that involve the measurement of dynamic pressure.

Origin of piezoelectricity

In most crystals, the unit cell (the basic repeating unit) is symmetrical; in piezoelectric crystals, it is not. Normally, piezoelectric crystals are electrically neutral—the atoms inside them may not be symmetrically arranged, but their electrical charges, are perfectly balanced. A quartz (SiO_2) tetrahedron is shown in Fig. 5.9. When a pressure is applied to the tetrahedron (or a macroscopic crystal element) a displacement of the positive ion charge towards the centre of the negative ion charges occurs. Hence, the outer faces of such a piezoelectric element get charged under this pressure.

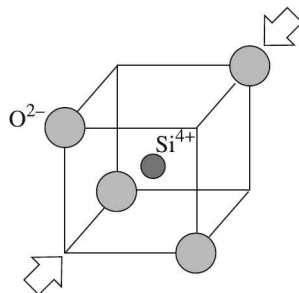


Fig. 5.9 Piezoelectricity generation in quartz. Arrows indicate the direction of application of pressure.

¹⁶Pronounced as piē'zō or pēā'zō (Webster's Universal Collegiate Dictionary, 1997). The prefix is a Greek word meaning *squeeze*.

Conversely, when an electric field is applied to a piezoelectric crystal, a mechanical strain is produced in it. This is sometimes called the *inverse piezoelectric effect*. If an alternating field is applied to such a crystal, the strain also varies periodically—but generally there is a phase lag between the applied field and the resulting strain, depending on the frequency of the applied field. At the natural frequency of vibration of the crystal, called the *resonance frequency*, the two are exactly in phase. This effect is utilised to construct resonant transducers and also to stabilise frequency in electronic clocks.

Piezoelectric materials

Generally, piezoelectric materials are classified into the following four categories:

| <i>Category</i> | <i>Examples</i> |
|--|---|
| Naturally occurring single crystals | Quartz Tourmaline Topaz Cane Sugar Rochelle salt (potassium sodium tartrate tetrahydrate, $\text{KNaC}_4\text{H}_4\text{O}_6, 4\text{H}_2\text{O}$) |
| Man-made crystals | Gallium Orthophosphate (GaPO_4) Langasite ($\text{La}_3\text{Ga}_5\text{SiO}_{14}$)— both quartz analogous crystals |
| Man-made polycrystalline ceramic materials | Barium Titanate (BaTiO_3) Lead Zirconate Titanate ($\text{Pb}[\text{Zr}_x\text{Ti}_{1-x}]\text{O}_3$ where $0 < x < 1$)—more commonly known as PZT Lead Titanate (PbTiO_3) Potassium Niobate (KNbO_3) Lithium Niobate (LiNbO_3) Lithium Tantalate (LiTaO_3) Sodium Tungstate (NaWO_3) |
| Man-made polymers | PolyVinylidene Fluoride (PVDF) |

PZT is the most common piezoelectric ceramic in use today. Among the naturally occurring crystals, quartz is inexpensive. Tourmaline, a naturally occurring semi-precious form of quartz, has sub-microsecond response time and, therefore, very useful in the measurement of rapid transients.

PVDF exhibits piezoelectricity several times greater than quartz. Unlike ceramics, where the crystal structure of the material creates the piezoelectric effect, in polymers the intertwined long-chain molecules attract and repel each other when an electric field is applied.

The so-called natural crystals are already polarised and the piezoelectric element is usually a cut from the crystal in the direction of any of the electrical axes (called *X*-axes) or mechanical axes (called *Y*-axes)[Fig. 5.10(a)]. Figures 5.10(b) and (c) show how an *X*-cut piece of the hexagonal quartz crystal can be obtained.

Synthetic polycrystalline ceramic materials have to be baked under a strong dc electric field to provide polarisation. Thus, they have the advantage of being moulded into any shape or size.

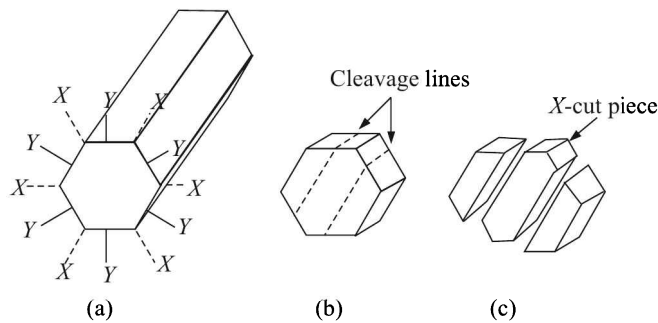


Fig. 5.10 Quartz crystal: (a) X and Y axes, (b) and (c) X -cutting

Curie temperature. The Curie temperature T_C is the temperature at which the piezoelectric material changes to a non-piezoelectric form. Before polarisation or above Curie temperature¹⁷, PZT crystallites have symmetric cubic unit cells [Fig. 5.11(a)]. Below the Curie temperature, the lattice structure becomes deformed and asymmetric. The unit cells then exhibit spontaneous polarisation [Fig. 5.11(b)], i.e. the individual PZT crystallites become piezoelectric.

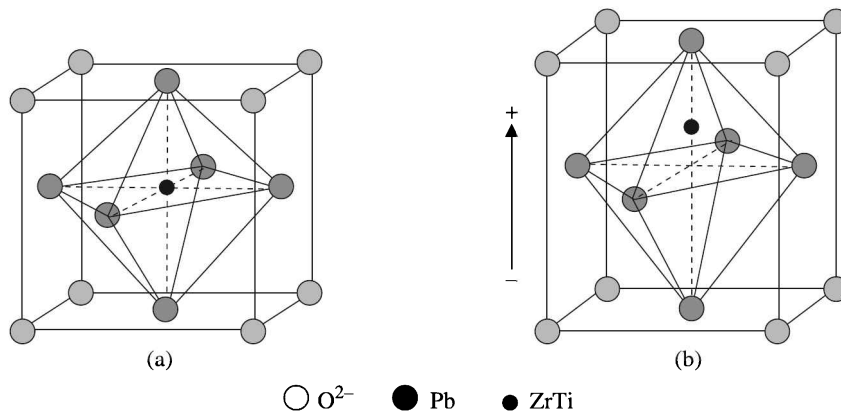


Fig. 5.11 PZT unit cell: (a) Perovskite-type PZT unit cell in the symmetric cubic state above the Curie temperature, and (b) tetragonally distorted unit cell below the Curie temperature.

Domains. Groups of unit cells with the same orientation of polarisation are akin to *Weiss domains* of ferromagnetism. The random distribution of the domain orientations in the ceramic material manifests no macroscopic piezoelectric behaviour [Fig. 5.12(a)]. Due to the ferroelectric¹⁸ nature of the material, it is possible to force permanent alignment of the different domains using a strong electric field. This process is called *poling* [Fig. 5.12(b)]. Some PZT ceramics must be poled at an elevated temperature to acquire a remnant

¹⁷Named after brothers Pierre and Jacques Curie of France who discovered piezoelectricity in 1880.

¹⁸Dielectrics which show hysteresis effect for applied field and polarisation are called *ferroelectrics*. A ferroelectric is spontaneously polarised, i.e. it is polarised in the absence of an electric field. Since the dielectric behaviour of these materials is in many respects analogous to the magnetic behaviour of ferromagnetic materials, they are called *ferroelectric solids*.

polarisation. The ceramic then exhibits piezoelectric properties [Fig. 5.12(c)]. It will also change dimensions when an electric potential is applied (inverse piezoelectricity).

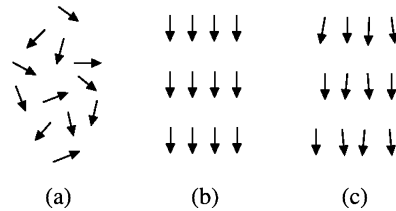


Fig. 5.12 Electric dipoles in domains: (a) unpoled ferroelectric ceramic, (b) during poling and (c) after poling (piezoelectric ceramic).

Modes of utilising piezoelectricity

In piezoelectric sensors, many modes of stressing the piezoelectric material can be used. Acting as precision springs, the different element configurations shown in Fig. 5.13 offer various advantages and disadvantages as detailed in Table 5.7.

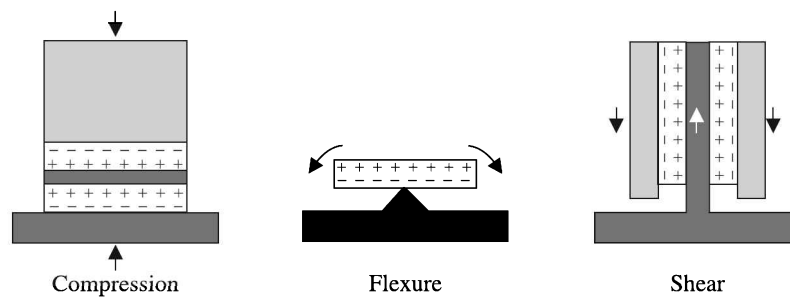


Fig. 5.13 Different modes of stressing the piezoelectric material. The white represents the piezoelectric crystals, while the arrows indicate how the material is stressed. Compression and shear modes typically have a seismic mass, which is represented by the grey colour.

Table 5.7 Advantages and disadvantages of different configurations

| <i>Configuration</i> | <i>Advantages</i> | <i>Disadvantages</i> |
|----------------------|--|--|
| Compression | High rigidity, making it useful for implementation in high frequency pressure and force sensors | Somewhat sensitive to thermal transients |
| Flexure | Simplicity of design | Narrow frequency range and low overshock survivability |
| Shear | Offers a well balanced blend of wide frequency range, low off-axis sensitivity, low sensitivity to base strain and low sensitivity to thermal inputs | Rather complicated design |

Piezoelectric coefficients

Because of the anisotropic nature of piezoelectric ceramics, piezoelectric effects are dependent on direction. To identify directions, the axes 1, 2, and 3 are introduced, corresponding to X , Y and Z of the classical right-handed orthogonal axis set. The axes 4, 5 and 6 identify rotations (shear)—23, 31, 12. Figure 5.14 illustrates them.

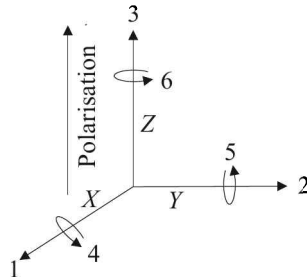


Fig. 5.14 Orthogonal system describing the properties of a poled piezoelectric ceramic. Axis 3 is the poling direction.

The direction of polarisation (axis 3) is established during manufacturing process by a strong dc field applied between two electrodes. For linear actuators¹⁹ which involve translation, the piezo properties along the poling axis, along which the largest deflection generally takes place, are the most important.

Piezoelectric materials are characterised by d , g , h , e coefficients as well as a coupling parameter k . We will discuss them after we talk about the notations used in defining them.

Apart from the piezoelectric coefficients, piezoelectricity is also affected by

1. Electric properties like permittivity and pyroelectricity²⁰
2. Elastic property like the Young's modulus
3. Thermal property like the Curie temperature

Notations. Piezoelectric constants are generally expressed with double subscripts. The subscripts link electrical and mechanical quantities. The first subscript indicates the direction of the stimulus while the second, the direction of the reaction of the system.

For example, d_{33} applies when the electric field is along the polarisation axis (direction 3) and the strain (deflection) is along the same axis. d_{31} applies if the electric field is in the same direction as before, but the deflection of interest is that along axis 1 (perpendicular to the polarisation axis).

In addition, piezoceramic material constants may be written with a *superscript* which specifies either a mechanical or electrical boundary condition. The superscripts are T , E , D and S are explained in Table 5.8.

¹⁹See at page 146.

²⁰Pyroelectric materials are those which produce electric charge as they undergo a temperature change. Piezoelectric materials are also pyroelectric. When their temperature is increased, they develop a voltage that has the same orientation as the polarisation voltage. When their temperature is decreased, they develop a voltage having an orientation opposite to the polarisation voltage. This creates a depolarising field with the potential to degrade the state of polarisation of the part.

Table 5.8 Significance of superscripts used to specify piezoelectric material constants

| <i>Superscript</i> | <i>Implication</i> | <i>Meaning</i> |
|--------------------|--|----------------------|
| <i>T</i> | Stress = constant | Mechanically free |
| <i>E</i> | Electric field = 0 | Short circuited |
| <i>D</i> | Charge displacement (i.e. current) = 0 | Open circuit |
| <i>S</i> | Strain = constant | Mechanically clamped |

- Note:*
1. We use here *S* for strain and *T* for stress rather than the conventional symbols ϵ and σ only to avoid confusion with the permittivity symbol ϵ .
 2. In a dielectric material the presence of an electric field \mathbf{E} causes the bound charges in the material (atomic nuclei and their electrons) to slightly separate, inducing a local electric dipole moment. The electric displacement field \mathbf{D} is defined as

$$\mathbf{D} \equiv \epsilon_0 \mathbf{E} + \mathbf{P}$$

where ϵ_0 is the vacuum permittivity (also called permittivity of free space), and \mathbf{P} is the (macroscopic) density of the permanent and induced electric dipole moments in the material, called the *polarisation density*.

3. In a linear, homogeneous, isotropic dielectric with instantaneous response to changes in the electric field, \mathbf{P} depends linearly on the electric field, giving rise to the relation

$$\mathbf{P} = \chi \epsilon_0 \mathbf{E}$$

where the constant of proportionality χ is called the *electric susceptibility* of the material. Thus

$$\mathbf{D} = \epsilon_0(1 + \chi)\mathbf{E} = \epsilon\mathbf{E}$$

where ϵ ($= \epsilon_0\epsilon_r$) is the permittivity and ϵ_r ($= 1 + \chi$) is the relative permittivity of the material.

4. In linear, homogeneous, isotropic media ϵ is a constant. However, in linear anisotropic media it is a matrix.

Now, let us define the different coefficients we have talked about.

***d* coefficient.** The piezoelectric charge coefficient (aka *charge constant*), d_{ij} , is defined as follows:

Direct effect

$$d_{ij} = \left. \frac{\text{Charge density developed in } i\text{-direction}}{\text{Applied stress in } j\text{-direction}} \right|_{E=0} \quad \text{C/N} \quad \left[\text{from } \frac{\text{C/m}^2}{\text{N/m}^2} \right] \quad (5.20)$$

Inverse effect

$$d_{ij} = \left. \frac{\text{Developed strain in } j\text{-direction}}{\text{Applied electric field in } i\text{-direction}} \right|_{T=\text{const.}} \quad \text{m/V} \quad \left[\text{from } \frac{\text{m/m}}{\text{V/m}} \right] \quad (5.21)$$

Note: The directions i and j are inverted in the inverse effect—the j -direction is in the numerator in this case.

Equations (5.20) and (5.21) may be written in the following form:

$$d_{ij} = \left(\frac{\partial D_i}{\partial T_j} \right)^E = \left(\frac{\partial S_j}{\partial E_i} \right)^T$$

Because for the inverse piezoelectric effect, the strain induced in a piezoelectric material by an applied electric field is the product of the value for the electric field and the value for d_{ij} , it is an important indicator of a material's suitability for strain-dependent (actuator) applications. The larger the value of d_{ij} , the larger the mechanical displacement which is usually sought in motional transducer devices.

g coefficient. The piezoelectric voltage coefficient (aka *voltage constant* or *voltage sensitivity*), g_{ij} , is defined as

Direct effect

$$g_{ij} = - \frac{\text{Developed electric field in } i\text{-direction}}{\text{Applied stress in } j\text{-direction}} \Big|_{D=0} \quad \text{Vm/N} \quad \left[\text{from } \frac{\text{V/m}}{\text{N/m}^2} \right] \quad (5.22)$$

Inverse effect

$$g_{ij} = \frac{\text{Strain developed in } j\text{-direction}}{\text{Applied charge density in } i\text{-direction}} \Big|_{T=\text{constant}} \quad \text{m}^2/\text{C} \quad \left[\text{from } \frac{\text{m/m}}{\text{C/m}^2} \right] \quad (5.23)$$

Combining Eqs. (5.22) and (5.23), we can write

$$g_{ij} = - \left(\frac{\partial E_i}{\partial T_j} \right)^D = \left(\frac{\partial S_j}{\partial D_i} \right)^T \quad (5.24)$$

Because the strength of the induced electric field produced by a piezoelectric material in response to an applied physical stress is the product of the value for the applied stress and the value for g_{ij} , high g_{ij} constants favour large voltage output, and therefore, are sought after for sensor applications.

h coefficient. The third coefficient h_{ij} is defined as

$$h_{ij} = - \left(\frac{\partial E_i}{\partial S_j} \right)^D = - \left(\frac{\partial T_j}{\partial D_i} \right)^S$$

It can be interpreted as negative of the voltage gradient per unit strain when the displacement is constant for the direct effect, or negative of the stress gradient per unit charge displacement when the strain is constant for the inverse effect.

e coefficient. The fourth coefficient e_{ij} is defined as

$$e_{ij} = \left(\frac{\partial D_i}{\partial S_j} \right)^T = - \left(\frac{\partial T_j}{\partial E_i} \right)^S$$

Coupling factor. The electromechanical coupling factor, k , is an indicator of the *effectiveness* with which a piezoelectric material converts electrical energy into mechanical energy, or vice versa. It is defined as

For an electrically stressed component

$$k_{ij}^2 = \frac{\text{Mechanical energy stored}}{\text{Electrical energy applied}}$$

For a mechanically stressed component

$$k_{ij}^2 = \frac{\text{Electrical energy stored}}{\text{Mechanical energy applied}}$$

Obviously, k ($0 \leq k < 1$) is a dimensionless quantity. It can be associated with the vibratory modes of certain simple transducer shapes. The first subscript to k denotes the direction along which the electrodes are applied and the second denotes the direction along which the mechanical energy is applied/stored.

A high k usually is desirable for efficient energy conversion, but k does not account for dielectric losses or mechanical losses, nor for recovery of unconverted energy. The accurate measure of *efficiency* is the ratio of converted, usable energy delivered by the piezoelectric element to the total energy taken up by the element. By this measure, piezoelectric ceramic elements in well designed systems can exhibit efficiencies that exceed 90%.

Useful relations

Piezoelectricity is a combined effect of the electric and elastic behaviour of the material.

Electric behaviour. The electrical condition of an unstressed medium placed under the influence of an electric field is defined by two quantities: the field strength, \mathbf{E} and the dielectric displacement, \mathbf{D} . Their relationship is expressed as:

$$\mathbf{D} = \varepsilon \mathbf{E}$$

where ε is the permittivity of the medium.

Elastic behaviour. The elastic behaviour is of the same medium at zero electric field strength is defined by two mechanical quantities: the applied stress, \mathbf{T} and the strain, \mathbf{S} . These two quantities are related by the well known *Hooke's law*

$$\mathbf{S} = \mathbf{s} \mathbf{T}$$

where \mathbf{s} denotes the compliance of the medium. The compliance is the inverse of the Young's modulus. Piezoelectricity involves the interaction between these two behaviours of the medium. To a good approximation, this interaction can be described by eight linear equations which relate the different electrical and mechanical variables. Three of them are as follows:

$$\mathbf{D} = \mathbf{d} \mathbf{T} + \varepsilon^T \mathbf{E} \quad (5.25)$$

$$\mathbf{S} = \mathbf{s}^E \mathbf{T} + \mathbf{d}^t \mathbf{E} \quad (5.26)$$

$$\mathbf{E} = -\mathbf{g} \mathbf{T} + \mathbf{D} / \varepsilon^T \quad (5.27)$$

where \mathbf{d}^t indicates the transpose of \mathbf{d} .

The reason for writing the quantities in bold letters is that they are not scalar quantities. In fact, \mathbf{E} , \mathbf{D} are vectors (tensors of rank 1)

- \mathbf{T} , \mathbf{S} are tensors of rank 2, but converted to 6-dimensional vectors
- \mathbf{s} is a 6×6 symmetric matrix
- $\boldsymbol{\varepsilon}$ is a 3×3 symmetric matrix
- \mathbf{d} , \mathbf{g} are 3×6 piezoelectric matrices

When written in their matrix forms, these equations relate the properties to the crystallographic directions. For ceramics and other crystals, the piezoelectric constants are anisotropic. For example, Eq. (5.25) can be written explicitly as follows:

$$\begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11}^T & \varepsilon_{12}^T & \varepsilon_{13}^T \\ \varepsilon_{21}^T & \varepsilon_{22}^T & \varepsilon_{23}^T \\ \varepsilon_{31}^T & \varepsilon_{32}^T & \varepsilon_{33}^T \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \quad (5.28)$$

Here T_1 , T_2 , T_3 denote longitudinal stress in 1, 2 and 3 directions while T_4 , T_5 , T_6 indicate shear stress along 23, 31, 12 directions as defined earlier.

Although Eq. (5.28) looks formidable, it simplifies a lot when the symmetry considerations of the crystal structures are invoked. For example, for the poled piezoelectric ceramic PZT, which possesses tetragonal symmetry, it simplifies to the following:

$$\begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{24} & 0 & 0 \\ d_{31} & d_{32} & d_{33} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11}^T & 0 & 0 \\ 0 & \varepsilon_{22}^T & 0 \\ 0 & 0 & \varepsilon_{33}^T \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix}$$

Relation between \mathbf{d} and \mathbf{g} constants. Utilising Eqs. (5.25) and (5.27) we can establish the relation between \mathbf{d} and \mathbf{g} as follows:

$$\begin{aligned} \mathbf{D} &= \mathbf{d}\mathbf{T} + \boldsymbol{\varepsilon}^T \mathbf{E} \\ &= \mathbf{d}\mathbf{T} + \boldsymbol{\varepsilon}^T \left(-\mathbf{g}\mathbf{T} + \frac{\mathbf{D}}{\boldsymbol{\varepsilon}^T} \right) \\ &= \mathbf{d}\mathbf{T} - \boldsymbol{\varepsilon}^T \mathbf{g}\mathbf{T} + \mathbf{D} \\ \Rightarrow \quad \mathbf{d} &= \boldsymbol{\varepsilon}^T \mathbf{g} \end{aligned}$$

Alternatively, we know from Eq. (5.20) that d_{33} denotes the ratio of charge per unit area perpendicular to the 3-direction to the stress applied in the 3-direction when the electrodes are open circuited. Therefore, d_{33} is given by the following expression:

$$d_{33} = \frac{Q_3/A}{F_3/A} = \frac{Q_3}{F_3} = \frac{CV_3}{F_3} \quad (5.29)$$

where A is the area stressed by a force F_3 , C is the capacitance between the electrodes of the piezoelectric and V_3 is the voltage generated. Now, the capacitance of the piezoelectric of thickness t and plate area A can be written as

$$C = \frac{\varepsilon_0 \varepsilon_r A}{t} \quad (5.30)$$

where $\varepsilon_0 = 8.85 \times 10^{-12}$ F/m.

Substituting the value of C from Eq. (5.30) in Eq. (5.29), we get

$$\begin{aligned} d_{33} &= \frac{\varepsilon_0 \varepsilon_r A V_3}{F_3 t} = \varepsilon_0 \varepsilon_r \frac{V_3/t}{F_3/A} \\ &= \varepsilon_0 \varepsilon_r g_{33} \quad [\text{from Eq. (5.22)}] \end{aligned}$$

Other useful relations are:

$$\begin{aligned} \mathbf{d} &= \mathbf{e}\mathbf{s}^E \\ \mathbf{e} &= \boldsymbol{\varepsilon}^s \mathbf{h} = \mathbf{d}\mathbf{Y}^E \\ \mathbf{g} &= \mathbf{h}\mathbf{s}^D \\ \mathbf{h} &= \mathbf{g}\mathbf{Y}^D \end{aligned}$$

Table 5.9 gives the useful data of a few piezoelectric materials.

Table 5.9 Useful data of piezoelectric materials

| <i>Material</i> | <i>d</i> (C/N) $\times 10^{-12}$ | ε_r | <i>Young's modulus</i> (N/m) $\times 10^9$ | <i>Max. Temp.</i> (°C) | <i>Humidity</i> range (%) |
|-------------------------------|--|-----------------|--|---------------------------|---------------------------------|
| Quartz | 2.3 | 4.5 | 80 | 550 | 0–100 |
| Tourmaline | 1.9 | 6.6 | 160 | 1000 | 0–100 |
| Rochelle salt | 550 | 350 | 19 | 45 | 40–70 |
| Lithium sulphate | 13.5 | 10.3 | 46 | 75 | 0–95 |
| Ammonium dihydrogen phosphate | 48 | 15.3 | 19.3 | 125 | 0–94 |
| PZT | 356 | 1750 | 59 | 285 | |
| Barium titanate | 150 | 1412 | 86 | 100 | |

Of all piezoelectric transducer materials, quartz is the most suitable for many applications. It has a lower temperature sensitivity and a higher resistivity, thus giving an inherently long time-constant which permits static calibration. Further, it exhibits good linearity with very low hysteresis over a wide range of pressure. Piezoelectric ceramics, though possessing higher sensitivity and wider adaptability in the form of shapes and sizes, have a poor temperature characteristic. Rochelle salt is hygroscopic and can be used only up to 45°C, while quartz devices can work between –200°C and 550°C.

With the advent of microelectronics and field effect transistors (FET), the piezoelectric transducer design has undergone a considerable change. Nowadays isolation amplifier and signal conditioning circuitry are packaged with the transducer. These integrated circuit piezoelectric transducers, called *smart sensors*²¹ operate over a simple 2-wire cable and are commercially available for all kinds of measurements where piezoelectricity helps.

Circuit analysis

The piezoelectric material is an insulator. Therefore to apply voltage to it or to extract voltage from it, metal electrodes have to be plated on the selected faces of the material. This configuration gives it the shape of a capacitor.

The piezoelectric transducer is also a charge generator. The charge is generated whenever it is stressed but it slowly dissipates through the piezoelectric material when left alone. So, the circuit constitutes a voltage source, a capacitor and a leakage resistor. The leakage resistance is very high, generally around $10^{11} \Omega$. Since the leakage resistance is high, the corresponding time constant will be considerable so as to make the decay a slow process. To measure such voltage, one needs to connect it to an op-amp with either a resistor in the feedback path (voltage amplifier configuration, see page 143) or a capacitor in the feedback path (charge amplifier configuration, mentioned later in this chapter, see page 144).

The connecting cables introduce some stray capacitance to the circuit. As a result, the equivalent circuit looks like Fig. 5.15(a) where the subscript p refers to the piezoelectric transducer.

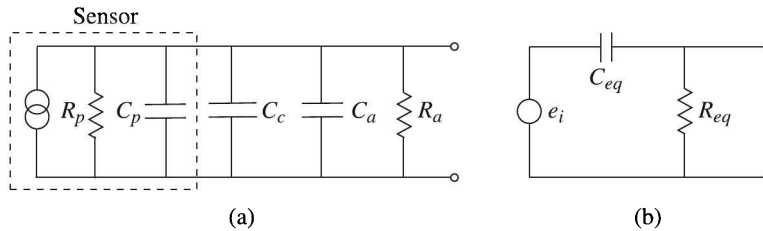


Fig. 5.15 (a) Equivalent circuit of a piezoelectric transducer installation and (b) its reduced form. C_c indicates cable capacitance, and C_a , R_a indicate amplifier capacitance and resistance respectively.

The circuit of Fig. 5.15(a) can be replaced by that of Fig. 5.15(b) where $R_{eq} \equiv R_p \parallel R_a$ and $C_{eq} \equiv C_p + C_c + C_a$. This equivalent circuit is similar to the one for a capacitive displacement transducer (see Section 6.2 at page 198) where E_b is replaced by a current generator.

In the Laplace transformed form, the input-output relation can be written as

$$E_o(s) = \frac{R_{eq}}{R_{eq} + \frac{1}{sC_{eq}}} E_i(s) = \frac{sR_{eq}C_{eq}}{sR_{eq}C_{eq} + 1} E_i(s)$$

$$\Rightarrow G(s) \equiv \frac{E_o(s)}{E_i(s)} = \frac{s\tau}{s\tau + 1} \quad (5.31)$$

where $\tau (= R_{eq}C_{eq})$ is the time constant.

²¹See Section 5.4 at page 166.

Now, the input voltage e_i , generated by the piezoelectric transducer, can be written as

$$e_i = \frac{Q}{C_{\text{eq}}} = \frac{dF}{C_{\text{eq}}} \quad (5.32)$$

where F is the applied force. Combining Eqs. (5.31) and (5.32), we get

$$\frac{E_o(s)}{F(s)} = \frac{d}{C_{\text{eq}}} \cdot \frac{s\tau}{s\tau + 1} \quad (5.33)$$

We know, $F = \eta x$ where η is the force constant and x is the displacement. Incorporating this in Eq. (5.33), we get

$$\frac{E_o(s)}{X(s)} = \frac{d\eta}{C_{\text{eq}}} \cdot \frac{s\tau}{s\tau + 1} = K \frac{s\tau}{s\tau + 1} \quad (5.34)$$

where $K \equiv \frac{d\eta}{C_{\text{eq}}}$ is the gain.

Impulse response For a unit impulse input, $X(s) = 1$. Therefore, the impulse response can be worked out from Eq. (5.34) as follows

$$\begin{aligned} E_o(s) &= K \frac{s\tau}{s\tau + 1} = K \left[1 - \frac{1}{s\tau + 1} \right] \\ &= K \left[1 - \frac{1}{\tau} \cdot \frac{1}{s + (1/\tau)} \right] \\ \Rightarrow e_o(t) &= K \left[\delta(t) - \frac{\exp(-t/\tau)}{\tau} \right] \end{aligned} \quad (5.35)$$

It is interesting to observe from Eq. (5.35) that for an input of $\delta(t)$, the output generates two factors—a $\delta(t)$ which dies out in a short time ε ($\varepsilon \rightarrow 0$) and another negative quantity that is slowly driven to a zero value.

To visualise what happens within time ε , let us consider it as a step function of amplitude $1/\varepsilon$ and duration ε . The Laplace transform of the input is then $1/(\varepsilon s)$. So, the response is as follows:

$$\begin{aligned} E_o(s) &= K \cdot \frac{s\tau}{s\tau + 1} \cdot \frac{1}{\varepsilon s} \\ &= \frac{K}{\varepsilon} \cdot \frac{1}{1 + (1/\tau)} \end{aligned} \quad (5.36)$$

$$\Rightarrow e_o(t) = \frac{K}{\varepsilon} \exp(-t/\tau) \quad (5.37)$$

Plots of Eqs. (5.37) and (5.35) for several values of τ are shown in Figs. 5.16(a) and (b).

It is clear from Fig. 5.16 that for a linear response, a large time constant is desirable. Now, the time constant can be increased by increasing either R or C . The latter can be increased, by adding a capacitor in parallel to the transducer, at the expense of the gain K which varies inversely with C_{eq} . Thus, a higher C reduces the sensitivity. Therefore, the connection of an R in series before feeding the transducer output to an amplifier is a better choice.

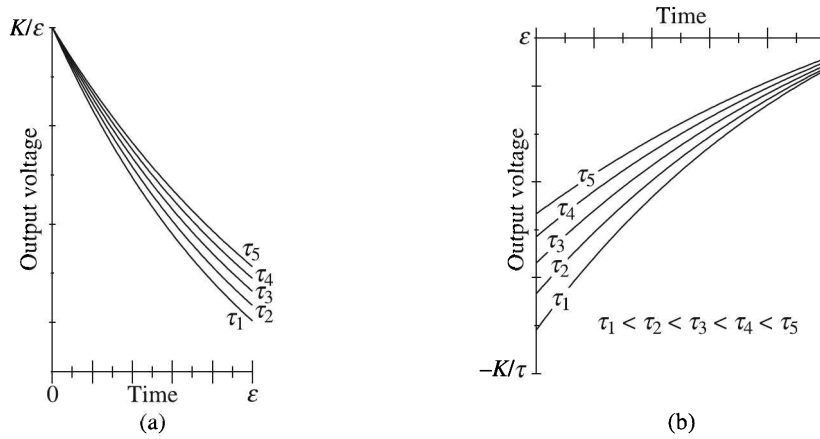


Fig. 5.16 Impulse response of piezoelectric transducer: (a) initial response for duration ϵ s [input assumed a step function of short duration] and (b) final response.

Frequency response In the frequency domain, the input-output relation can be written as

$$\begin{aligned}
 e_o(j\omega) &= \frac{R_{eq}}{R_{eq} + (1/j\omega C_{eq})} e_i(j\omega) = \frac{j\omega R_{eq} C_{eq}}{j\omega R_{eq} C_{eq} + 1} e_i(j\omega) \\
 &= \frac{j\omega\tau}{j\omega\tau + 1} e_i(j\omega)
 \end{aligned} \tag{5.38}$$

Thus
$$\frac{e_o(j\omega)}{x(j\omega)} = K \frac{j\omega\tau}{j\omega\tau + 1} \equiv K \frac{1}{\sqrt{1 + (1/\omega\tau)^2}} \angle\phi \tag{5.39}$$

where
$$\phi = \frac{\pi}{2} - \tan^{-1} \omega\tau \tag{5.40}$$

The frequency response is shown in Fig. 5.17. Equations (5.39) and (5.40), and Fig. 5.17 help us conclude that

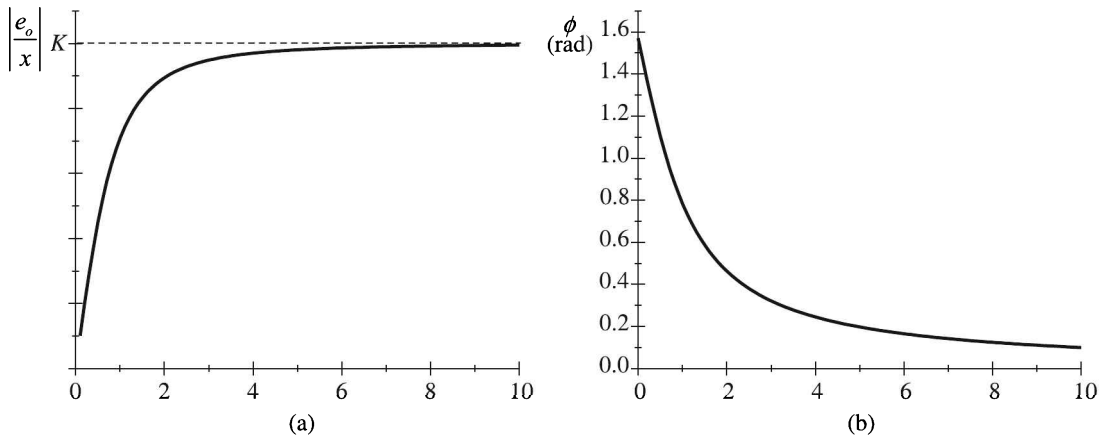


Fig. 5.17 Frequency response of a piezoelectric transducer: (a) amplitude and (b) phase.

1. For $\omega\tau = 0$, $e_o = 0$. That means, for a static pressure, the transducer produces no steady voltage.
2. For $\omega\tau \rightarrow \infty$, $e_o/x = K$, $\phi = 0^\circ$. That means, the response is independent of frequency, but dependent on C_{eq} at high frequencies. This property makes piezoelectric transducers suitable for hi-fidelity low power sound reproduction speakers. There, the inverse piezoelectric effect is utilised.

Signal conditioning

Two circuits are used for signal conditioning, namely

1. Voltage mode amplifier circuit
2. Charge mode amplifier circuit

Voltage mode amplifier. The voltage mode amplification is used when the amplifier is pretty close to the sensor.

The circuit and its frequency response are shown in Fig. 5.18. Here, the output depends on the amount of capacitance seen by the sensor.

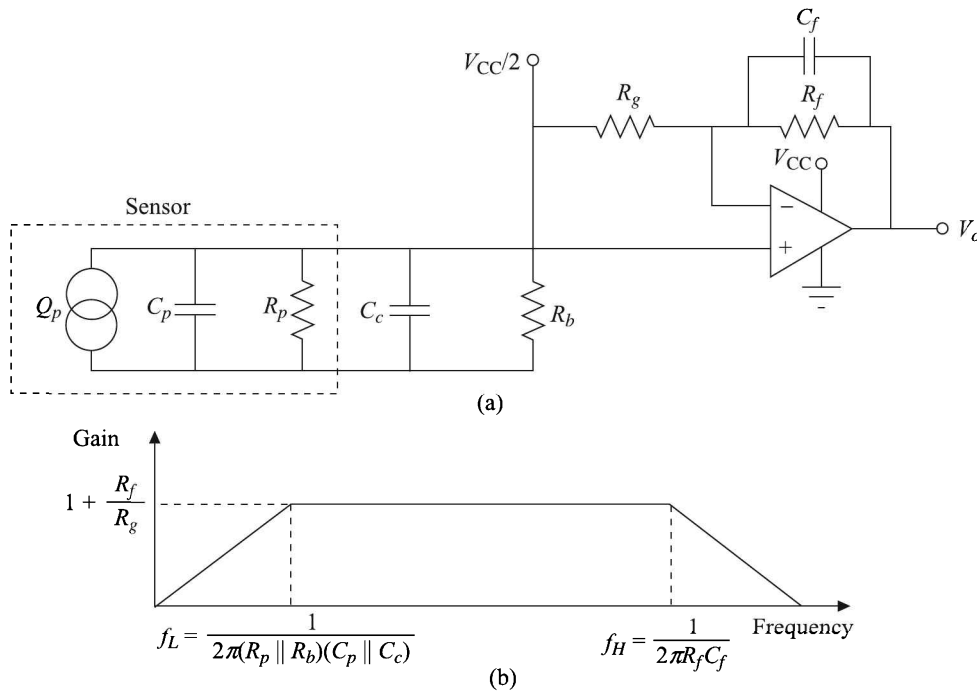


Fig. 5.18 Voltage mode amplifier: (a) Circuit and (b) frequency response. Output voltage, $V_o = \frac{Q_p}{C_p + C_c} \left[1 + \frac{R_f}{R_g} \right] + \frac{V_{cc}}{2}$.

The capacitance of the interface cable, C_c , will affect the output voltage. If the cable is moved or replaced, variations in C_c may cause problems. The resistor R_b provides a dc bias path for the amplifier input stage. The choice of R_f and C_f sets the upper cutoff frequency. The lower cutoff frequency is calculated from the relation:

$$f_L = \frac{1}{2\pi(R_p \parallel R_b)(C_p \parallel C_c)}$$

R_b should be high and the interface cable length low. The biasing will put the output voltage at $V_{cc}/2$ with no input. The output is given by

$$V_o = \frac{Q_p}{C_p + C_c} \left[1 + \frac{R_f}{R_g} \right] + \frac{V_{cc}}{2} \quad (5.41)$$

where Q_p is the charge developed on the piezoelectric
 C_p is the capacitance of the piezoelectric
 C_c is the capacitance of the connecting cable
 R_f is the feedback resistor
 C_f is the feedback capacitor
 R_g is the regulating resistor

It is obvious from Eq. (5.41) that V_o will swing above and below the dc bias $V_{cc}/2$.

Charge mode amplifier. The charge mode amplification is resorted to when the amplifier is remote to the sensor.

The charge mode amplifier circuit and its frequency response are shown in Fig. 5.19.

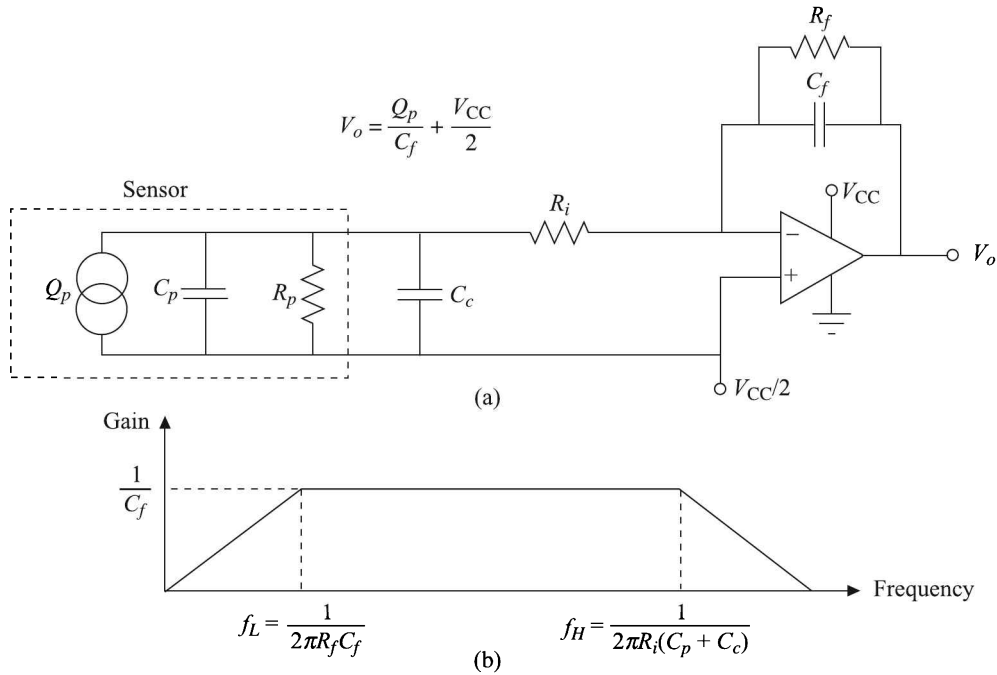


Fig. 5.19 Charge mode amplifier: (a) Circuit and (b) frequency response.

The amplifier will balance the charge injected into the negative input by charging the feedback capacitor C_f . The resistor R_f does two functions, namely

1. It bleeds the charge off the capacitor C_f at a low rate to prevent the amplifier from drifting into saturation.
2. It also provides a dc bias path for the negative input.

The values of R_f and C_f set the lower cutoff frequency of the amplifier given by

$$f_L = \frac{1}{2\pi R_f C_f}$$

The amplifier action maintains a 0 V across its input terminals so that the stray capacitance of the connecting cable poses no problem. The resistor R_i provides ESD²² protection as well as it combines with the capacitors C_p and C_c to set the higher cutoff frequency given by

$$f_H = \frac{1}{2\pi R_i (C_p + C_c)}$$

The output is given by

$$V_o = -\frac{Q_p}{C_f} + \frac{V_{cc}}{2}$$

Thus, for no input the biasing puts the output voltage at $V_{cc}/2$. Which means, the output will swing around this dc level.

Example 5.2

Determine the pressure sensitivity of a quartz piezoelectric transducer of thickness 2.5 mm. Voltage sensitivity of quartz is 50×10^{-3} Vm/N.

Solution

Given $g = E_o/t_p = 50 \times 10^{-3}$ Vm/N and $t = 2.5$ mm = 2.5×10^{-3} m. Therefore, pressure sensitivity = $E_o/p = gt = 125$ mV/kPa.

Example 5.3

A quartz piezoelectric transducer has the following specifications: area = 1 cm², thickness = 1 mm, Young's modulus = 9×10^{10} Pa, charge sensitivity = 2 pC/N, relative permittivity = 5 and resistivity = 10^{14} Ω-cm. A 20 pF capacitor and a 100 MΩ resistor are connected in parallel across the electrodes of the piezoelectric transducer. If a force $F = 0.02 \sin(10^3 t)$ N is applied, calculate

- (a) the peak-to-peak voltage generated across the electrodes
- (b) the maximum change in crystal thickness

Solution

Given:

Area of the piezoelectric transducer, $A = 1$ cm² = 10^{-4} m²

Thickness, $t = 1$ mm = 10^{-3} m

Young's modulus, $Y = 9 \times 10^{10}$ Pa

Charge sensitivity, $d = 2$ pC/N = 2×10^{-10} C/N

Relative permittivity, $\epsilon_r = 5$, therefore, $\epsilon = \epsilon_0 \epsilon_r = 4.405 \times 10^{-11}$ F/m

Resistivity, $\rho = 10^{14}$ Ω-cm = 10^{10} Ω-m

Parallel resistance, $R = 100$ MΩ = 10^8 Ω

Parallel capacitance, $C = 20$ pF = 20×10^{-10} F

²²ElectroStatic Discharge.

Therefore,

Resistance of the piezoelectric transducer, $R_p = \rho A/t = 10^{11} \Omega$

Capacitance of the piezoelectric transducer, $C_p = \epsilon A/t = 4.425 \times 10^{-12} \text{ F}$

Equivalent resistance, $R_{\text{eq}} = R_p \parallel R \simeq 10^8 \Omega$

Equivalent capacitance, $C_{\text{eq}} = C_p + C = 24.425 \times 10^{-12} \text{ F}$

Time constant, $\tau = R_{\text{eq}} C_{\text{eq}} = 24.425 \times 10^{-4} \text{ s}$

The applied force is sinusoidal with an amplitude of 0.02 N, i.e. with a peak-to-peak value, $(F)_{\text{p-p}} = 0.04 \text{ N}$ and its angular frequency, $\omega = 10^3 \text{ rad}$.

(a) Therefore, from Eq. (5.40), we get

$$\begin{aligned} (e_o)_{\text{p-p}} &= \frac{d}{C_{\text{eq}}} \frac{1}{\sqrt{1 + (1/\omega\tau)^2}} (F)_{\text{p-p}} \\ &= \frac{2 \times 10^{-12}}{24.425 \times 10^{-12}} \times \frac{1}{\sqrt{1 + 1/(10^3 \times 24.425 \times 10^{-4})^2}} \times 0.04 \text{ V} \\ &\simeq 2.8 \text{ mV} \end{aligned}$$

(b) Since, Young's modulus, $Y = \frac{\text{longitudinal stress}}{\text{longitudinal strain}} = \frac{F/A}{\Delta t/t}$, we have,

$$\begin{aligned} (\Delta t)_{\text{p-p}} &= \frac{(F)_{\text{p-p}} t}{AY} \\ &= \frac{0.04 \times 10^{-3}}{10^{-4} \times 9 \times 10^{10}} \text{ m} \\ &\simeq 4.4 \times 10^{-12} \text{ m} = 4.4 \text{ pm} \end{aligned}$$

Piezoelectric actuator

An actuator is a mechanical device for moving or controlling a mechanism or system. It is operated by a source of energy, usually in the form of an electric current.

In instrumentation, actuators are a subdivision of transducers. They are devices which transform an input signal (mainly an electrical signal) into motion. Specific examples include: electrical motors, pneumatic actuators, hydraulic actuators, linear actuators etc. Piezoelectric actuators are also used because piezoelectrics deform linearly with an applied electric field.

Commonly used stack actuators achieve a relative displacement of up to 0.2%. Displacement of piezoceramic actuators is primarily a function of the applied electric field strength E , the length L of the actuator, the forces applied to it and the properties of the piezoelectric material used. The material properties can be described by the piezoelectric charge constant d_{ij} . We know that this constant describes the relationship between the applied electric field and the mechanical strain produced.

The change in length, ΔL , of an unloaded single-layer piezo actuator can be estimated by the following equation:

$$\Delta L = S \cdot L \approx \pm E d_{ij} L$$

where S is the strain $= \Delta L/L$.

Because strains are so small, piezoelectric actuators are mainly used in speakers or precision micro-positioning applications where small, precise motion is needed. However, deflection amplification methods make piezoelectrics possible actuators in other applications including micro-robotics.

Unimorph. One method of amplification is using the unimorph design shown in Fig. 5.20. When a voltage is applied across the ceramic and metal plate, the unimorph bends. It bends in the other direction if the voltage is reversed.

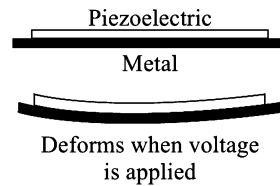


Fig. 5.20 Schematic diagram of unimorph.

This device relies on the d_{31} piezoelectric constant. This relates the change in strain induced perpendicular to the electric field. The value of d_{31} is typically half of d_{33} value. However, a motion of 0.875 inch can be produced by a unimorph of approximately one inch in diameter and 0.02 inch thick. This design is typically found in loudspeakers.

Bimorph. The bimorph uses two piezoelectric plates that amplify the deflection as shown in Fig. 5.21. Since piezoelectricity appears only on the surface, it is easy to understand why two layers, instead of a thicker piezoelectric plate, is used for this purpose.

A piezo bimorph operates like a bimetallic strip in a thermostat. When the ceramic is energised, the substrate is deflected with a motion proportional to the applied voltage. Bimorph actuators providing motion up to 1000 μm are available and greater travel range is possible.

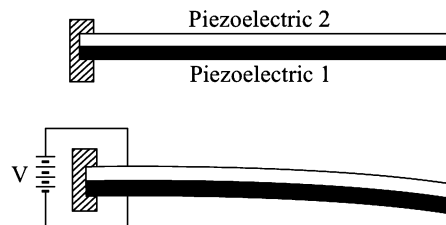


Fig. 5.21 Schematic diagram of bimorph.

Two basic versions are available:

1. Two electrode bimorph, i.e. serial bimorph [Fig. 5.22(a)]
2. Three electrode bimorph, i.e. parallel bimorph [Fig. 5.22(b)]

Serial bimorph operates with the two ceramic plates in the same direction of polarisation. To avoid depolarisation in the middle, the maximum electric field is limited to a few hundred volts per millimetre. Serial bimorph benders are widely used as force sensors because of halved capacity and higher output voltage.

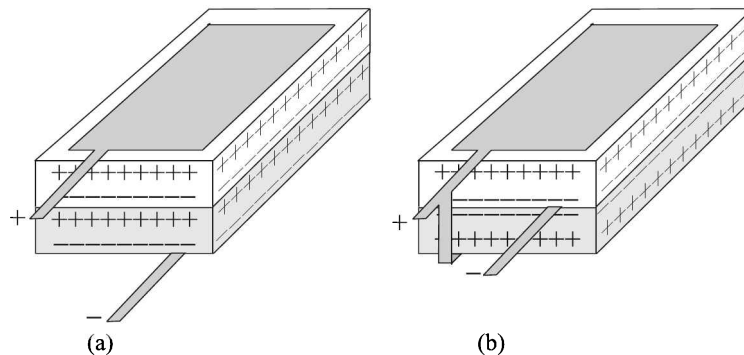


Fig. 5.22 Bimorphs: (a) serial and (b) parallel.

Parallel bimorph produces twice the capacitance as that of a series connection and in a sender-type transducer admits the full excitation voltage across each plate.

In either case the device relies on the d_{31} piezoelectric constant and that the strain is proportional to the square of the applied voltage.

Multimorph. Instead of two plates, monolithic, multi-layer type piezo benders, known as *multimorphs*, are available too. Similar to multilayer stack actuators, they run on a low operating voltage (60 to 100 V).

Bender type actuators provide large motion in a small package at the expense of stiffness, force and speed.

Advantages and disadvantages

From its discovery by the Curies in 1880, it took about 70 years before the piezoelectric effect was used for industrial sensing applications. Since then, its utilisation has experienced a constant growth and can nowadays be regarded as a mature technology with an outstanding inherent reliability. It has been successfully used in various critical applications like in medical, aerospace and nuclear instrumentation.

Advantages. The high modulus of elasticity of many piezoelectric materials is comparable to that of many metals and the maximum stress that they can withstand can be as high as $105 \times 10^6 \text{ N/m}^2$. Even though piezoelectric sensors are electromechanical systems that react on compression, the sensing elements show almost zero deformation. The piezoelectric sensors are

1. Rugged
2. Have an extremely high natural frequency
3. An excellent linearity over a wide amplitude range
4. Insensitive to electromagnetic fields and radiation, enabling measurements under harsh conditions
5. Materials like gallium phosphate or tourmaline have an extreme stability over temperature enabling sensors to have a working range of 1000°C

Table 5.10 gives an idea about the relative standing of piezoelectric transducers vis-à-vis the strain sensitivity of other transducers.

Table 5.10 Comparison of sensitivity of sensing principles

| <i>Principle</i> | <i>Sensitivity (V/μm)</i> | <i>Resolution (μm)</i> | <i>Dynamic range (dB)</i> |
|------------------|---------------------------|------------------------|---------------------------|
| Piezoelectric | 5.0 | 0.00001 | 160 |
| Piezoresistive | 0.0001 | 0.0001 | 128 |
| Inductive | 0.001 | 0.0005 | 126 |
| Capacitive | 0.005 | 0.0001 | 117 |

Disadvantages. In comparison to the advantages of piezoelectric transducers, disadvantages are only a few, namely

1. The major disadvantage is that they cannot be used for true static measurements. A static force will generate a fixed amount of charge on the piezoelectric material. Working with conventional electronics, not perfect insulating materials, and reduction in internal sensor resistance will result in a constant loss of charge, thus yielding an inaccurate signal.
2. Elevated temperatures cause an additional drop in internal resistance. Therefore, at higher temperatures, only piezoelectric materials can be used that maintain a high internal resistance.

Applications

The piezo materials are available in a variety of shapes and sizes such as discs, plates, bars, rings, rods, tubes, etc. Some of their typical applications as transducers are as follows:

1. Vibration and shock measurement
2. Accelerometers
3. Ultrasound flow meters
4. Dynamic force and pressure measurement
5. NDT (non-destructive testing) transducers²³

Other applications include

1. Stable oscillation frequency generators
2. High voltage generators for gas lighters
3. Fuses for explosives
4. Nebulisers
5. SONAR
6. Deepwater hydrophones²⁴
7. Actuators/translators
8. Ultrasonic cleaners, welders

²³Ultrasound waves are passed through a material and received at different speeds relative to the density and elastic properties of the material, producing data that can be utilised to create a cross-sectional image.

²⁴Device for converting sound waves into electrical signals, similar in operation to a microphone but used primarily for detecting sound waves from an underwater source, such as a submarine.

We would like to mention in this context that many creatures make an interesting use of piezoelectricity. Bones are piezoelectric materials and they act as force sensors. Once loaded, bones produce charges proportional to the resulting internal strain. Those charges stimulate and drive the development of new bone material. This leads to the strengthening of structures where the internal displacements are the greatest. Thus, with time weaker structures increase their strength and stability as material is laid down proportional to the forces affecting the bone.

Piezoresistivity

The piezoresistive effect is the change of electric resistivity of the material caused by an applied mechanical stress. Many materials change their resistance when stressed, but the piezoresistive effect is the most pronounced in semiconductors. Semiconductor piezoresistive sensing elements, or piezoresistors, are typically used as pressure and force sensors, where the applied mechanical load is converted into a proportional electric signal.

Origin of piezoresistivity

It is apparent that piezoresistivity has nothing to do with piezoelectricity, though some piezoresistors are piezoelectric as well for reasons different altogether.

When a semiconducting material is stressed, the interatomic spacings within the material change. This eventually changes the bandgaps in each atom making it easier (or harder depending on the material and strain) for electrons to be raised into the conduction band. A higher or lower electron population in the conduction band results in a change in resistivity of the semiconductor.

We know that the resistance R of a conductor is given by

$$R = \rho \frac{l}{A}$$

where ρ is the resistivity of the material of the conductor

l is the length of the conductor

A is the area of cross-section of the conductor.

For metals, ρ is more or less a constant at a given temperature because their conduction bands are already sufficiently populated with electrons. But the conduction bands of semiconductors are not so populated normally and as already discussed, at a given temperature ρ varies for semiconductors when they are stressed. For them the piezoresistivity ρ_σ is defined by

$$\rho_\sigma = \frac{d\rho/\rho}{\varepsilon}$$

where $d\rho$ is the change in resistivity

ρ is the original resistivity

ε is the strain

Now, when a semiconductor is strained, its length and area of cross-section will eventually change with a consequent change in its resistance. But its piezoresistive change can be several

orders of magnitudes larger than the geometrical effect. This effect is conspicuous in materials like germanium, polycrystalline silicon, amorphous silicon, silicon carbide, and single crystal silicon.

Piezoresistors

Piezoresistors are fabricated using a wide variety of piezoresistive materials. The simplest form of silicon piezoresistors are diffused resistors. They consist of a simple two contact diffused n or p -well within a p or n -substrate. The typical square resistances of these devices are in the range of several hundred ohms. This necessitates additional p^+ or n^+ diffusions to facilitate ohmic contacts to the device (Fig. 5.23).

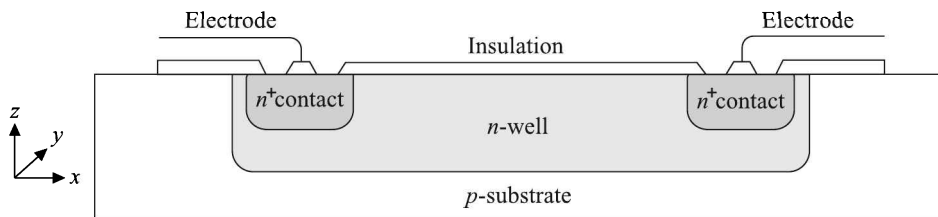


Fig. 5.23 Schematic cross-section of an elementary silicon n -well piezoresistor.

For typical stress values in the order of mPa, the piezoresistivity can be written as

$$\rho_{\sigma} = \frac{d\rho/\rho}{\varepsilon} = \pi Y$$

where π is the piezoresistive coefficient and Y is the Young's modulus. Both π and Y depend on

1. Basic material, now mostly silicon
2. Majority carriers, i.e. p or n
3. Crystal orientation given by Miller indices like (100) or (111)
4. Angle between the current and stress; the stress may be tensile, shear or volume compression
5. Degree of doping indicated by the room-temperature resistivity ρ_0
6. Size and shape of the resistor

In general, both the stress and the current are along the length of the piezoresistor. Then, the relation for the longitudinal piezoresistance coefficient is given by

$$\pi_1 = \pi_{11} - 2(\pi_{11} - \pi_{12} - \pi_{44})(\alpha_1^2\beta_1^2 + \alpha_1^2\gamma_1^2 + \beta_1^2\gamma_1^2)$$

where π_{11} , π_{12} , π_{44} are the fundamental piezoresistive coefficients, the subscripts referring to the current and stress directions

α_1 , β_1 , γ_1 are the direction cosines of the current with respect to the crystallographic axes

For a shear stress perpendicular to the current direction, the relation for the transverse piezoresistive coefficient is given by

$$\pi_1 = \pi_{12} + (\pi_{11} - \pi_{12} - \pi_{44})(\alpha_1^2 \alpha_2^2 + \beta_1^2 \beta_2^2 + \gamma_1^2 \gamma_2^2)$$

Table 5.11 gives an idea about the characteristics of lightly doped silicon and germanium at moderate strain at the room temperature. In general, the sensitivity and thermal coefficient of doped semiconductors decrease with higher doping.

Table 5.11 Relevant piezoresistive characteristics of lightly doped silicon and germanium at moderate strain at the room temperature. Crystal orientations are (111) for all, except *n*-Si for which it is (100)

| Characteristic (unit) | <i>p</i> -Si | <i>n</i> -Si | <i>p</i> -Ge | <i>n</i> -Ge |
|---|--------------|--------------|--------------|--------------|
| Unstrained resistivity ($\times 10^{-3} \Omega\text{-m}$) | 78 | 117 | 150 | 166 |
| π_{11} ($\times 10^{-12} \text{ m}^2/\text{N}$) | 66 | -1022 | -106 | -52 |
| π_{12} ($\times 10^{-12} \text{ m}^2/\text{N}$) | -11 | 534 | 50 | 55 |
| π_{44} ($\times 10^{-12} \text{ m}^2/\text{N}$) | 1381 | -136 | 986 | -1387 |
| ρ_σ | 175 | -133 | 102 | -157 |
| Young's modulus ($\times 10^9 \text{ N/m}^2$) | 187 | 130 | 155 | 155 |
| Poisson's ratio | 0.180 | 0.278 | 0.156 | 0.156 |

Nonlinearity. The variation of piezoresistivity with strain at moderate doping is far from linear. For example, the fractional resistivity variation of lightly doped *p*-Si at moderate tensile stress and at the room temperature can be written as

$$\frac{d\rho}{\rho} = 175\varepsilon + 72625\varepsilon^2$$

At higher stress levels, the nonlinearity can be even higher and more temperature dependent. Of course, the nonlinearity as well as temperature dependence can be lowered by higher doping but at the expense of sensitivity. Figure 5.24 shows a qualitative picture of the variation.

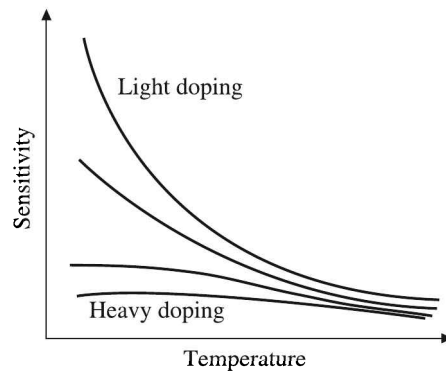


Fig. 5.24 Sensitivity vs. temperature plots for *p*-Si at different degrees of doping.

Despite the fairly large stress sensitivity of simple resistors, they are preferably used in more complex configurations eliminating certain cross sensitivities and thermal effects.

Typical applications of piezoresistors are as strain gauges which are used in pressure transducers and accelerometers.

Surface Acoustic Waves

In 1887, Lord Rayleigh²⁵ discovered the surface acoustic wave (SAW) mode of propagation. Named after their discoverer, Rayleigh waves have a longitudinal and a vertical shear component that can couple with a medium in contact with the surface of the device (Fig. 5.25). Such coupling strongly affects the amplitude and velocity of the wave. This feature enables SAW sensors to directly sense mass and mechanical properties.

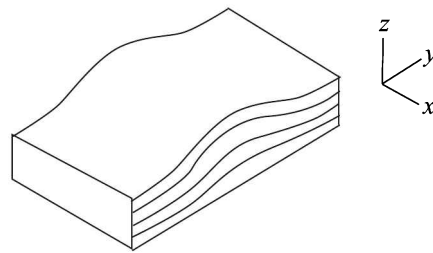


Fig. 5.25 Surface acoustic wave propagation, shown in an exaggerated way.

The wave has a velocity that is nearly 5 orders of magnitude less than the corresponding electromagnetic wave, making Rayleigh surface waves among the slowest to propagate in solids. The wave amplitudes are typically around 10 \AA and the wavelengths range from $1\text{--}100 \mu\text{m}$.

Fabrication of the SAW sensor

The SAW sensors are made by a photolithographic process. Initially a piezoelectric substrate is carefully polished and cleaned. Then a metal, usually aluminium, is deposited uniformly on the substrate. Next it is coated with a photoresist, baked to harden it and then exposed to UV light through a mask with opaque areas corresponding to the areas to be metallised on the final device. The exposed areas undergo a chemical change that allows them to be removed with a developing solution. Finally, the remaining photoresist is removed. The pattern of metal remaining on the device is called an *interdigital*²⁶ transducer, or IDT. By adjusting the length, width, position, and thickness of the IDT, the performance of the SAW sensor can be maximised. The diagram of a SAW sensor is shown in Fig. 5.26.

Among the piezoelectric materials chosen for the substrate, the most common are quartz (SiO_2), lithium tantalate (LiTaO_3), and, sometimes, lithium niobate (LiNbO_3).

Principle of operation

The input IDT of the sensor provides the electric field necessary to generate an oscillation in the piezoelectric substrate through the *inverse piezoelectric effect*. Thus a travelling acoustic wave

²⁵Lord Rayleigh (John William Strutt), a British Physicist (1842–1919) worked on the theory of waves. He became the Cavendish Professor of Physics at Cambridge and was awarded the Nobel prize in Physics (1904) for his discovery of the gas Argon.

²⁶The word *digit* literally means *a finger of a human being*. So, *interdigital* signifies a pattern that resembles the interwoven fingers of both hands.

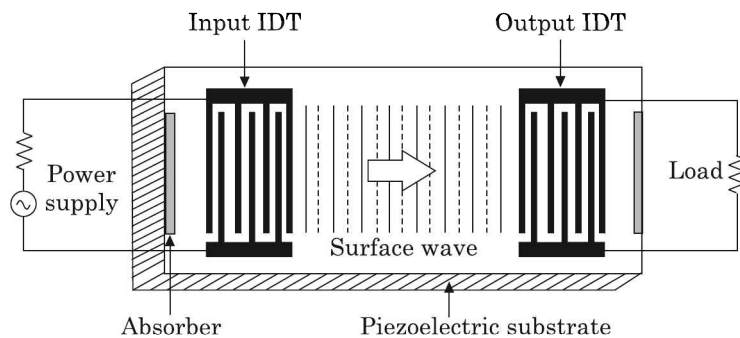


Fig. 5.26 Diagram of a surface acoustic wave sensor.

is formed. The wave propagates to the other end of the substrate, where it is converted back to an electric field by the output IDT. Since this is a longitudinal wave, alternate compression and stretching of the substrate take place in the transverse direction. These produce alternate polarisations in the opposite directions on the surface of the substrate. The fingers of the output IDT are so spaced that they are able to capture these charges to produce a voltage.

Applications

The SAW sensors generally operate from 25–500 MHz. They are sensitive, to varying degrees, to perturbations from many different physical parameters. The range of phenomena that can be detected by them can be greatly expanded by coating the devices with materials that undergo changes in their mass, elasticity, or conductivity upon exposure to some physical or chemical stimulus. For example, these sensors become

- *Pressure, torque, shock, and force detectors* under an applied stress that changes the dynamics of the propagating medium.
- *Mass, or gravimetric, sensors* when particles are allowed to come in contact with the propagation medium thus changing the stress on it.
- *Vapour sensors* when a coating is applied that absorbs only specific chemical vapours and changes the mass of the coating.
- *Biosensors*, if the coating absorbs specific vapours of biological fluids.

One disadvantage of these devices is that the Rayleigh waves are surface-normal waves, making them unsuitable for liquid sensing. When a SAW sensor is in contact with a liquid, the resulting compressional waves cause an excessive attenuation of the surface wave.

Two interesting applications of the SAW sensor are as band-pass filters in mobile telephones and sensing device in touch-screen displays.

Optical Effects

Detectors based on optical effects can be classified as shown in the tree diagram of Fig. 5.27.

Photographic film, photopolymers, etc. can be called chemical detectors. They do not give a signal output in the usual sense as do the other types.

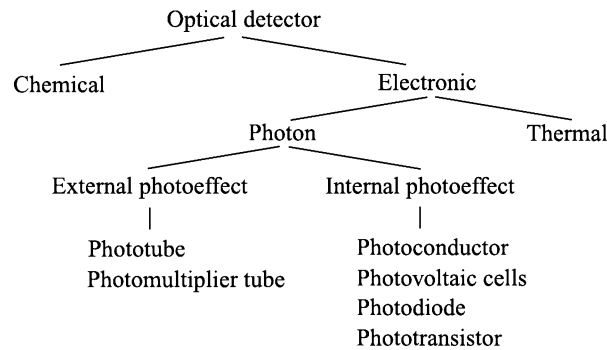


Fig. 5.27 Optical detectors tree.

In thermal detectors, the absorption of light raises the temperature of the device and this, in turn, results in changes in some temperature-dependent parameter (e.g. electrical conductivity). Some of the better known thermal detectors are the thermocouple, the bolometer and pyroelectric detectors. The last type can be made with response times in the nanosecond region and with a wavelength response up to 100 μm . They have proved very useful as low cost, robust infrared detectors.

The operation of photon detectors is based on the photoeffect, in which the absorption of photons by some materials results directly in an electronic transition to a higher energy level and the generation of mobile charge carriers. The photoeffect takes two forms:

- *External photoeffect*: The process involves photoelectric emission in which the photo-generated electrons escape from the material (the photocathode) as free electrons.
- *Internal photoeffect*: In the internal photoeffect, the photoexcited carriers (electrons and holes) remain within the material.

We will consider the external photoeffect, which can be called *photoemissive effect*, first.

External photoeffect: Photoemission

When irradiated by electromagnetic radiation of very short wavelength, such as visible or ultraviolet light, electrons are emitted from matter—metals and non-metallic solids, liquids or gases—as a consequence of their absorption of energy. This phenomenon is called *the photoelectric effect* or *photoemissive effect* (Fig. 5.28). Emitted electrons are referred to as *photoelectrons*. The effect was first observed by Heinrich Hertz²⁷ in 1887.

The photoelectric effect requires photons with energies from a few electron volts to over 1 MeV in high atomic number elements. In photoelectric emission, the photo-generated electrons escape from the material (the photocathode) as free electrons with a maximum kinetic energy given by the photoelectric equation of Einstein:

$$E_{\text{max}} = h\nu - \varphi$$

where the work function φ is the energy difference between the Fermi level and the continuum of the material. The study of the photoelectric effect led to important consequences in

²⁷Heinrich Rudolf Hertz (1857–1894) was a German physicist who carried out important experiments to prove the electromagnetic nature of light.

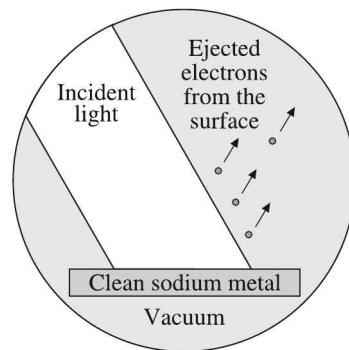


Fig. 5.28 Schematic diagram showing emission of photoelectrons.

understanding the quantum nature of light and electrons, and led to the formation of the concept of wave–particle duality of matter.

Observations. The remarkable aspects of the photoelectric effect are:

1. The electrons are emitted immediately. There is no time lag between the irradiation of the substance and the ejection of electrons from it.
2. If the intensity of the light is increased, the number of photoelectrons also increases, but not their maximum kinetic energy.
3. An impinging red light ($\lambda = 700 \text{ nm}$) will not cause the ejection of electrons, no matter whatever its intensity is. A green ($\lambda = 550 \text{ nm}$) or violet ($\lambda = 400 \text{ nm}$) light will eject electrons. But their maximum velocities are greater the shorter the wavelength (Fig. 5.29).

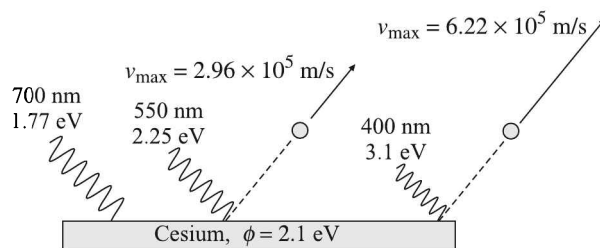


Fig. 5.29 Schematic representation of emission of photoelectrons for incidence of different wavelengths of light. ϕ indicates the value of the work function for potassium.

Theoretical explanation. In 1905 Einstein²⁸ gave a very simple interpretation of the photoelectricity phenomenon. He assumed that the incoming radiation should be thought of as quanta of energy $h\nu$, with ν the frequency. In photoemission, one such quantum is absorbed by one electron. If the electron is deep inside the material, some energy will be lost as it moves towards the surface. This is usually called the *work function*, ϕ . The most

²⁸Albert Einstein (1879–1955) was a German-born theoretical physicist who is often regarded as the father of modern physics. He received the 1921 Nobel Prize in Physics “for his services to theoretical physics, and especially for his discovery of the law of the photoelectric effect”.

energetic electrons emitted will be those very close to the surface, and they will leave the surface with kinetic energy

$$E_{\max} = h\nu - \varphi$$

where h is the Planck constant and ν is the frequency of the incident photon. The work function satisfies the relation

$$\varphi = h\nu_0$$

where ν_0 is the threshold frequency for the metal. The maximum kinetic energy of an ejected electron is then

$$E_{\max} = h(\nu - \nu_0)$$

Since kinetic energy is a positive quantity, we must have $\nu > \nu_0$ for the photoelectric effect to occur.

Stopping potential. Let a light source illuminates a plate X , and another plate electrode Y collects the emitted electrons. We apply a voltage between X and Y , change it slowly and measure the current flowing in the external circuit between the two plates.

If the frequency and the intensity of the incident radiation are fixed, the photoelectric current increases gradually with an increase in the voltage until the photoelectric current attains a saturation value and does not increase further whatever the voltage. The saturation current depends on the intensity of illumination, but not its wavelength.

If we now apply a negative voltage to Y with respect to X and gradually increase it, the photoelectric current decreases until it is zero, at a certain negative voltage. The minimum negative voltage at which the photoelectric current becomes zero is called *stopping potential* or *cut off potential*. It may be observed that for a given frequency of the incident radiation, the stopping potential V_0 is

1. Independent of its intensity.
2. Related to the maximum kinetic energy of the photoelectron that is just stopped from reaching Y . If m is the mass and v_{\max} is the maximum velocity of photoelectron emitted, then

$$E_{\max} = \frac{1}{2}mv_{\max}^2$$

If e is the charge on the electron, then the work done by the retarding potential in stopping the electron = eV_0 , which gives

$$\frac{1}{2}mv_{\max}^2 = eV_0 \quad (5.42)$$

Eq. (5.42) shows that the maximum velocity of the emitted photoelectron is independent of the intensity of the incident light. Hence

$$E_{\max} = eV_0 = e\varphi = \frac{hc}{\lambda_0}$$

$$\Rightarrow \lambda_0 = \frac{hc}{e\varphi} = \frac{1.2431}{\varphi} \mu\text{m}$$

Photoemissive materials. Many metals are photoemissive when irradiated with photons of moderate energy. The work functions and corresponding threshold wavelengths of a few common elements are listed in Table 5.12.

Table 5.12 Work functions and threshold wavelengths for a few common elements

| <i>Element</i> | φ (eV) | λ_0 (μm) | <i>Element</i> | φ (eV) | λ_0 (μm) |
|----------------|----------------|-------------------------------|----------------|----------------|-------------------------------|
| Aluminum | 4.08 | 0.3025 | Magnesium | 3.68 | 0.3354 |
| Beryllium | 5.0 | 0.2468 | Mercury | 4.5 | 0.2742 |
| Cadmium | 4.07 | 0.3032 | Nickel | 5.01 | 0.2463 |
| Calcium | 2.9 | 0.4256 | Niobium | 4.3 | 0.2870 |
| Carbon | 4.81 | 0.2566 | Potassium | 2.3 | 0.5366 |
| Cesium | 2.1 | 0.5877 | Platinum | 6.35 | 0.1943 |
| Cobalt | 5.0 | 0.2468 | Selenium | 5.11 | 0.2415 |
| Copper | 4.7 | 0.2626 | Silver | 4.73 | 0.2609 |
| Gold | 5.1 | 0.2420 | Sodium | 2.28 | 0.5413 |
| Iron | 4.5 | 0.2742 | Uranium | 3.6 | 0.3428 |
| Lead | 4.14 | 0.2981 | Zinc | 4.3 | 0.2870 |

Photoemissive transducers. The photoemissive transducers that we will consider are

1. Phototube
2. Photomultiplier tube

Phototube. Photoemissive devices usually take the form of vacuum tubes called *phototubes* (Fig. 5.30). It works on the basic principle of light striking the cathode which causes the emission of electrons. Emitted electrons travel to the anode which is kept at 70–180 V. As a result, an electric current proportional to the photon flux incident on the cathode is created in the circuit.

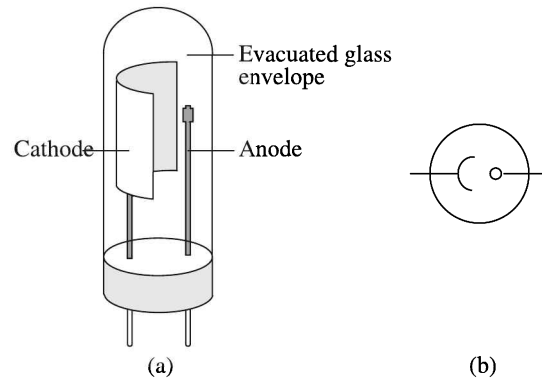


Fig. 5.30 (a) Schematic diagram of a phototube, (b) symbolic representation of a phototube.

The *quantum yield* K_λ is defined as

$$K_\lambda = \frac{\text{Number of electrons released}}{\text{Number of photons absorbed}}$$

The value of K_λ lies between 0 and 0.5. Pure metals are rarely used as cathodes since they have low quantum yields (~ 0.1) and high work functions ($\varphi = 2.1$ eV for Cs) which makes them useful only in the visible and ultraviolet regions of the spectrum. However, semiconductors can operate with higher quantum yields and lower work functions corresponding to wavelengths up to about $1.1 \mu\text{m}$. Typically cathodes are made of materials like Cs_3Sb , NaO , AgOCs .

The *responsivity* R_λ is defined as

$$R_\lambda = \frac{X_\lambda}{\Phi_\lambda}$$

where X_λ is the output signal like voltage, current, charge

Φ_λ is the incident flux in W

The generated current i_a is given by the relation

$$i_a = \eta e h \nu \Phi_\lambda R_\lambda$$

where η is the anode collection efficiency (0-1) which is dependent on its design and bias

e is the electronic charge

The responsivity vs. λ is called the *spectral response*. The relative spectral response of various photoelectric transducers is shown in Fig. 5.31.

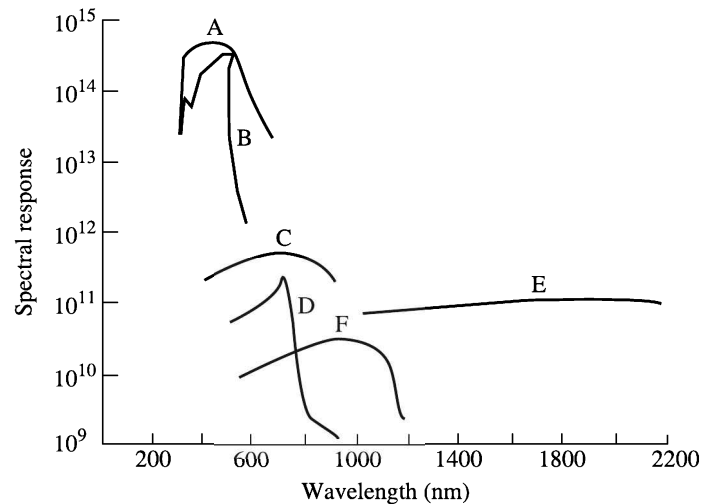


Fig. 5.31 Relative spectral response of various photon transducers. A: photomultiplier tube, B: CdS photoconductive cell, C: GaAs photovoltaic cell, D: CdSe photoconductive cell, E: PbS photoconductive cell, F: Si photodiode.

Photomultiplier tube. The *photomultiplier tube* (Fig. 5.32) makes use of the photoelectric effect to convert small intensities of light into electrical current.

Electrons dislodged by the photoelectric effect from the photoemissive cathode travel down a special tube consisting of secondary electron generating dynodes. Dynodes are usually made of materials like MgO or GaP . Each dynode is biased on the order of 100 V more positive

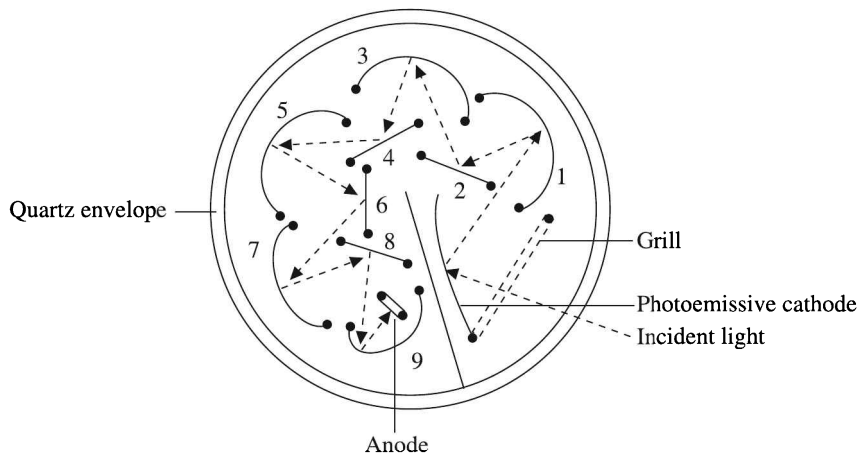


Fig. 5.32 Diagram of a photomultiplier tube. The numbers 1, 2, 3, ... indicate dynodes.

than the previous to accelerate electrons from dynode to dynode (Fig. 5.33). The gain per dynode is typically 2–5 and the total gain is 10^6 – 10^8 . So, by the time it gets to the end, a single electron can gather a hundred million other electrons. The charge pulse at anode is a few nanosecond wide.

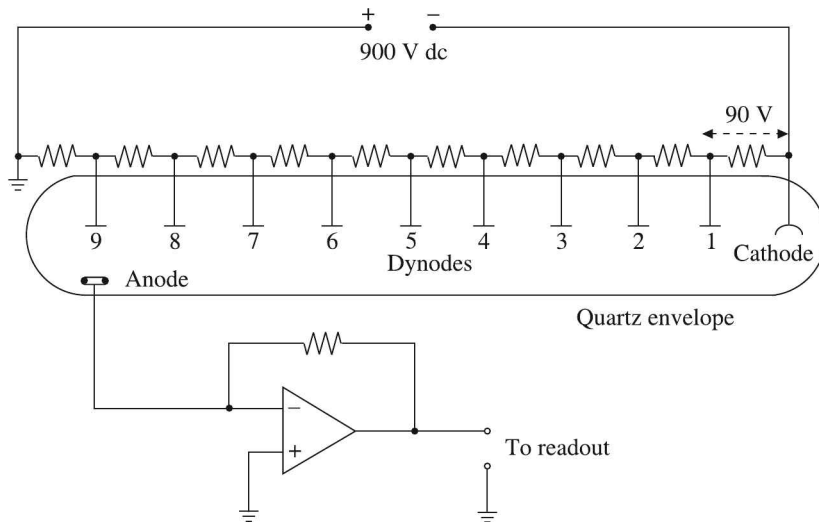


Fig. 5.33 Biasing of dynodes and output from a photomultiplier tube. The numbers 1, 2, 3, ... indicate dynodes.

Internal photoeffect

Among the transducers that utilise internal photoeffect, we will consider the following:

1. Photoconductor
2. Photovoltaic transducer

3. Photodiode
4. Phototransistor

Photoconductor. Photoconductivity is one of the internal photoeffects in which a material becomes more electrically conductive when irradiated by light of suitable waveband such as visible, ultraviolet, infrared or gamma radiation. Such a device is called a *photoresistor*, or a *photoconductor* or sometimes a *light dependent resistor* (LDR).

Construction. Photoconductors are typically made by depositing thin films of photoconductive substances on a glass or plastic substrate. Electrodes are deposited on the photoconductive surface and are made of materials which give an ohmic contact, but with low resistance compared with that of the photoconductor. Gold is typically used because its Fermi level²⁹ matches those of intrinsic semiconductors and therefore, no rectification of signals takes place at the points of contact. The electrodes are usually interdigital, i.e. in the form of interlocked fingers, as shown in Fig. 5.34.

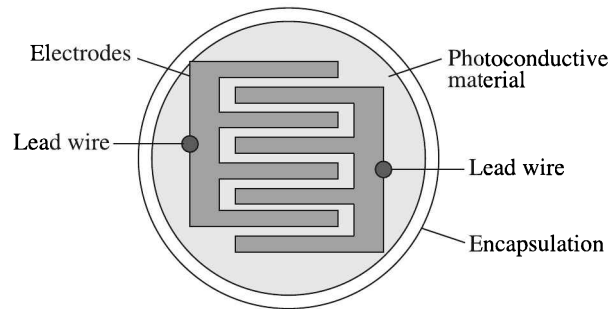


Fig. 5.34 Photoconductor.

The design of the electrode system affects the resistance and voltage ratings of the cell. Different resistances and voltage ratings can be achieved on a film of a given size. A device with a small number of widely spaced electrodes will have a higher resistance and voltage rating than a device using the same film with a large number of closely spaced electrodes. Devices can be made for end on or side illumination. The encapsulation is hermetically sealed so that it is resistant to humidity and corrosive atmosphere.

Principle of operation. Photoconductor detectors rely directly on the light-induced increase in the conductivity, an effect exhibited by almost all semiconductors. The absorption of a photon results in the generation of a free electron excited from the valence band to the conduction band, and a hole is generated in the valence band. An external voltage source connected to the material causes the electrons and holes to move, resulting in a detectable electric current. The detector operates by registering either the current (which is proportional to the photon flux) or the voltage drop across a series resistor. Unlike the quantum efficiency for the photoelectric effect, for example, the gain in a photoconductor may be larger than unity. The semiconducting material may take the form of a slab or a thin film.

²⁹Top filled electron energy level at 0 K.

Photoconductive materials. Photoresistors are available in many different types. Inexpensive cadmium sulphide (CdS) ones can be found in many consumer items like camera light meters, clock radios, burglar alarms and automatic ON/OFF street lights.

CdS LDRs have a long life, typically up to 10,000 hours. Their response is temporarily impaired by exposure to strong light, but they recover by themselves and are not damaged. Damage may be caused by electrical overload. So, the applied voltage and current for the required illumination must be known. If in doubt a series resistor can be incorporated to limit the current. Overheating can cause damage, but the device is usually vibration resistant.

A few classic examples of photoconductive materials are listed in Table 5.13

Table 5.13 A few classic photoconductive materials and their uses

| <i>Material</i> | <i>Comments</i> |
|---------------------------------|---|
| Polyvinylcarbazole ^a | Conductive polymer used extensively in photocopying (xerography). |
| Lead sulphide | Used in infrared detection applications, such as heat-seeking missiles <i>Sidewinder</i> (USA) and <i>Atoll</i> (Russia). |
| Selenium | Employed in early television and photocopying. |
| Ge:Cu | Among the best far-infrared detectors available. They are used for infrared astronomy and infrared spectroscopy. |

^a Commonly called OPC (organic photoconductor).

The photoconductive cells usually have high gains (10^3 – 10^4) but poor response times (~ 50 ms).

Photovoltaic transducer. The *photovoltaic* cell is another transducer which utilises the internal photoeffect.

Principle of operation. The photovoltaic cell is a *p-n* junction structure where photons absorbed in the depletion layer generate electron-hole pairs which are subject to the local electric field within that layer. Because of this field, the two carriers drift in opposite directions and an electric current is induced in the external circuit.

Construction. Photovoltaic cells are fabricated from thin layer of crystalline semiconductor, e.g. Se, Si, Cu₂O as well as from ternary and quaternary compound semiconductors such as InGaAs, HgCdTe and InGaAsP sandwiched between two different metal electrodes. Pure selenium, a *p*-type semiconductor, is coated on a metal base such as aluminium or brass. Then cadmium is diffused into selenium to form a *p-n* junction as cadmium oxide forms the *n*-layer. This layer is sometimes called the barrier layer. On top of it is coated a thin layer of silver or gold which forms the metal layer as well as an electrode. The entire cell is encapsulated in a plastic case. Devices are often constructed in such a way that the light impinges normally on the *p-n* junction instead of parallel to it. A typical construction is shown in Fig. 5.35.

Solar cell. A basic photovoltaic cell is also called a *solar cell*. For solar cells, a thin semiconductor wafer is specially treated to form an electric field, positive on one side and negative on the other. A number of solar cells electrically connected to each other and mounted in a support structure or frame is called a *photovoltaic module*. Modules are

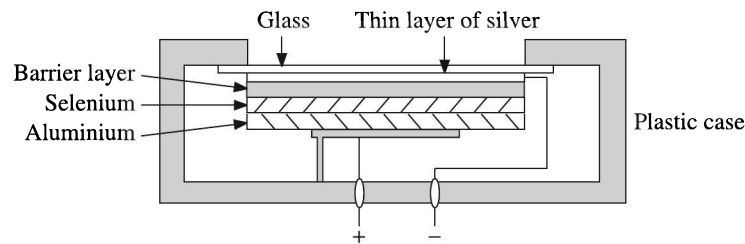


Fig. 5.35 Schematic diagram of a photovoltaic cell.

designed to supply electricity at a certain voltage, such as a common 12 volt system. The current produced is directly dependent on how much light strikes the module. Multiple modules can be wired together to form an *array*. In general, the larger the area of a module or array, the more electricity will be produced.

Multijunction solar cell. In a single-junction photovoltaic cell, only photons whose energy is equal to or greater than the band gap of the junction can free an electron-hole pair for an electric circuit. In other words, the photovoltaic response of single-junction cells is limited to the portion of the sun's spectrum whose energy is above the band gap of the absorbing material, and lower-energy photons are not used.

One way to get around this limitation is to use two (or more) different cells, with more than one band gap and more than one junction, to generate a voltage. These are referred to as *multijunction* cells (aka *cascade* or *tandem* cells). Multijunction devices can achieve a higher total conversion efficiency because they can convert more of the energy spectrum of light to electricity.

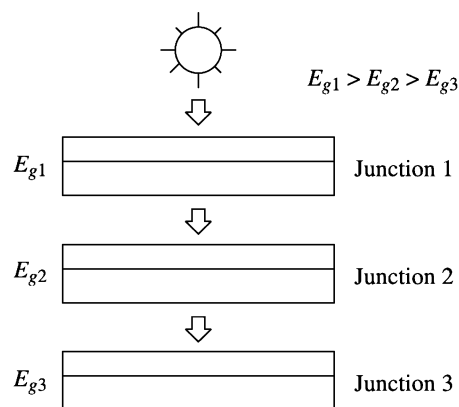


Fig. 5.36 Functioning of a multijunction cell.

As shown in Fig. 5.36, a multijunction device is a stack of individual single-junction cells in descending order of band gap (E_g). The top cell captures the high-energy photons and passes the rest of the photons on to be absorbed by lower band gap cells.

Multijunction materials. Much of today's research in multijunction cells focusses on gallium arsenide as one (or all) of the component cells. Such cells have reached efficiencies of around 35% under concentrated sunlight. Other materials studied for multijunction devices have been amorphous silicon and copper indium diselenide.

Photodiode. A *photodiode* is sort of a miniature solar cell that consists of an active $p-n$ junction which is operated in reverse bias [Fig. 5.37(a)]. The light incident on the junction generates a reverse current which is proportional to the intensity of illumination.

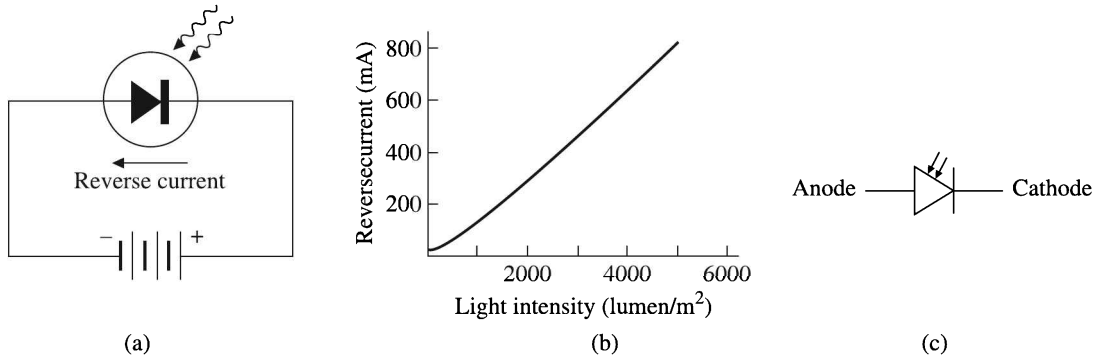


Fig. 5.37 Photodiode: (a) operation, (b) response curve, and (c) symbol.

Its linear response [Fig. 5.37(b)] to light makes it a useful photodetector for some applications. It is also used as the active element in light-activated switches. The symbol of a photodiode is shown in Fig. 5.37(c).

The photodiode response is fast—on the order of nanoseconds. But it is not as sensitive as a phototransistor. However, its linearity of response can be utilised to construct simple light meters.

Phototransistor. Phototransistors are designed specifically to take advantage of the fact that like diodes, all transistors are light-sensitive. The most common variant is an $n-p-n$ bipolar transistor with an exposed base region. When light strikes the base, a voltage is applied to the base. So, a phototransistor amplifies variations in the light striking it. A phototransistor may or may not have a base lead. If it does, the base lead allows us to bias the light response of the phototransistor.

Of course, photodiodes also can provide a similar function, but with much lower gain. Which means, photodiodes allow much less current to flow than do phototransistors.

The illustration given in Fig. 5.38, where both circuits are equivalent, helps us to understand the difference between a photodiode and a phototransistor. It suggests that the phototransistor is basically a combination of a photodiode and a transistor which not only detects the light intensity like a photodiode, but also amplifies the generated current.

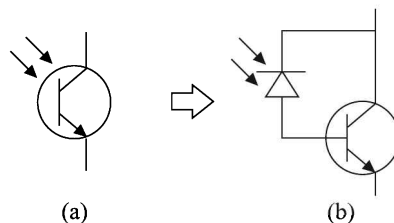


Fig. 5.38 (a) Phototransistor and (b) an equivalent circuit with a photodiode.

Table 5.14 offers a comparison between the characteristics of different photon detectors.

Table 5.14 Typical characteristics of photon detectors

| <i>Type</i> | D^* ^a (cm Hz ^{1/2} W ⁻¹) | R_λ ^b | <i>Linear range</i> (decades) | <i>Spectral range</i> (nm) | <i>Rise time</i> ^c (ns) | <i>Output</i> |
|----------------------|---|---|----------------------------------|-------------------------------|---------------------------------------|-------------------|
| Phototube | 10 ⁸ –10 ¹⁰ | 0.001–0.1 ^d | 5 | 200–1000 | 1–10 | Current |
| Photomultiplier tube | 10 ¹² –10 ¹⁷ | 10–10 ⁵ ^d | 6 | 110–1000 | 1–10 | Current, charge |
| Photoconductive cell | 10 ⁹ –10 ¹² | 10 ⁴ –10 ⁶ ^e | 5 | 750–6000 | 50–10 ⁶ | Resistance change |
| Photovoltaic cell | 10 ⁸ –10 ¹¹ | 100–10 ⁶ ^e | 3 | 400–5000 | 10 ³ | Current, voltage |
| Si Photodiode | 10 ¹⁰ –10 ¹² | 0.05–0.5 ^d | 5–7 | 250–1100 | 1–10 | Current |

^a Measure of minimum detectability: $D = 1/\Phi_n$; D^* is normalised D for area A (cm²) and bandwidth Δf (Hz) [$DA^{1/2}(\Delta f)^{1/2}$].

^b Values indicate range for several different types.

^c Time for output to rise from 10–90% of final value for instantaneous increase in radiant power.

^d A/W

^e V/W

5.3 Selection of Transducers

It is important to remember that since transducers constitute the sensing element in an instrumentation system, the precision of the data produced by the instrumentation system depends, in most of the cases, on the capability of the transducer. For example, a precise temperature control can hardly be achieved if the transducer used is a simple bi-metallic strip. Therefore, the selection of a proper transducer is important from the standpoint of required precision.

The following points should be considered while selecting a transducer:

Fundamental parameters. These include

- (a) Type of measurand
- (b) Range of measurement
- (c) Required precision, which includes
 - (i) Allowable nonlinearity effects
 - (ii) Allowable dead-zone effects
 - (iii) Frequency response
 - (iv) Resolution.

Environment. This includes consideration of

- (a) Ambient temperature
- (b) Corrosive or non-corrosive atmosphere
- (c) Shock and vibration to withstand.

Physical conditions. These are

- (a) Room or available space to mount the transducer
- (b) Whether the measurement is static or dynamic
- (c) How much energy can be extracted from the measurand to carry-out measurement without much loading.

Compatibility with the next stage. Normally, some standard signal conditioner and display devices are used with a transducer, unless they are custom-built to suit the requirements of the transducer. In the former case, the transducer should be so chosen as to meet the requirements of the next stage, such as

- (a) Impedance matching
- (b) Excitation voltage matching
- (c) Sensitivity tolerance matching.

These criteria, of course, are not exhaustive but they may offer some guidance as regards selection of a suitable transducer.

Transducers can be constructed from various materials and in many designs. But to gain acceptance in the field of instrumentation they must conform to the following six cardinal requirements:

1. Ruggedness to withstand overloads
2. Linearity
3. Repeatability
4. Stability and reliability
5. Good dynamic response
6. Convenient instrumentation

5.4 Smart Sensors and IEEE 1451 Standard

Ordinary sensors or transducers help sense and/or control process parameters like temperature, pressure, strain, flow, pH, etc. Smart sensors provide functions beyond those. A smart sensor

1. Acquires data
2. Conditions the signal
3. Converts the measurement into the attribute's units
4. Transmits the data to a network by wireline or wireless method

An example will perhaps make the concept clear. Suppose, we are measuring the temperature of a measurand with the help of thermocouple that generates a particular voltage corresponding to a particular temperature. If the thermocouple is connected to network, the network controller needs to convert it to represent the data in degrees Celsius or Fahrenheit ('attribute's units'). A smart sensor incorporates a look-up table or transducer electronic data sheet (TEDS) that performs the conversion and presents the data in appropriate unit of temperature to the network controller. To perform the latter task, the smart sensor also possesses a built-in digital interface that provides a communication channel with the network control. A functional smart sensor system thus consists of two main components, namely

1. Transmitter interface module (TIM) that contains physical transducer and data acquisition system, and
2. Network capable application processor (NCAP) where control and data correction take place.

Diagrammatically the system can be represented as shown in Fig. 5.39.

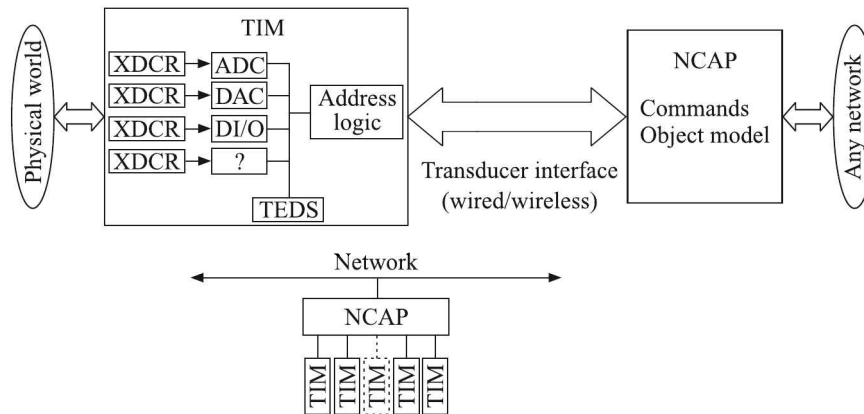


Fig. 5.39 Smart sensor system. XDCR \equiv transducer.

The TIM and NCAP perform the following functions:

| <i>TIM</i> | <i>NCAP</i> |
|------------------------------------|---|
| 1. Analogue signal conditioning | 1. Message encoding and decoding |
| 2. Triggering | 2. Detection and control of TIMs |
| 3. Analogue to digital conversion | 3. Correction and interpretation of TEDS data |
| 4. Command processing | 4. Message routing and interface control |
| 5. TEDS storage | |
| 6. Data transfer and communication | |

With the advancement of digital communication, increasing demand for networked configuration to connect sensors and actuators and proliferation of smart sensor manufacturers, the IEEE³⁰ introduced a standard, called IEEE 1451, to address the following requirements:

1. Network and vendor independent plug-and-play TIMs without having to add special drivers or profiles. The features that enable plug-and-play operation are TEDS and the basic command set.
2. Analogue or digital interfaces so that sensors or actuators can be easily connected by either wireline or wireless method.
3. Installation, upgradation, replacement and/or movement of sensors with minimum effort.
4. Elimination of manual data entry and system configuration steps which are error prone.

³⁰Institute of Electrical and Electronics Engineers (USA).

Be it mentioned here that IEEE 1451 is *not* another field network. It is an open standard that may be used with multiple networks.

Smart sensors, though very elegant, are finding it hard to make inroads to existing industries owing to (i) the great technological leap and (ii) the cost associated with the replacement of entire existing software and hardware infrastructures.

The new standard, referred to as IEEE P1451.4, provides many of the features of IEEE 1451 such as automatic detection, configuration and calibration while retaining the existing measurement architectures. It is backward compatible to traditional sensors and measurement architecture while allowing integration of smart sensors.

In the following chapters we will discuss a few transducers and associated methods of measurements and other relevant matters. For our convenience we will deal with the measurement of a few physical quantities because the techniques involved in these measurements are of representative character.

Review Questions

- 5.1 What is transducer? What is the difference between sensor and transducer? Name some of the active transducers which are used in the measurement of temperature.
- 5.2 What is 'transducer'? Define active and passive transducers with examples and state the role of each in measuring system.
- 5.3 Match the following:

| <i>Phenomenon</i> | <i>Name</i> |
|---------------------------------------|-----------------------------------|
| (a) Applied force causes emf | (i) Seebeck effect |
| (b) Applied voltage causes vibrations | (ii) Hall effect |
| (c) Temperature difference causes emf | (iii) Photoelectric effect |
| (d) Light causes current | (iv) Magnetostriction |
| | (v) Piezoelectric effect |
| | (vi) Inverse piezoelectric effect |

- 5.4 Match the devices with the sensors:

| | |
|------------------------------|---------------------------|
| (a) High frequency vibration | (e) Strain gauge |
| (b) Load cell | (f) Hall element |
| (c) Gauss meter | (g) Potentiometer |
| (d) Large displacement | (h) Piezoelectric crystal |

- 5.5 Match the devices with quantities measured:

| | |
|------------------------------|-----------------------|
| (a) Rotameter | (e) Relative humidity |
| (b) Hydrometer | (f) Density |
| (c) Sling psychrometer | (g) Dynamic pressure |
| (d) Piezoelectric transducer | (h) Flow rate |

5.6 Match the transducers with the materials used:

- | | |
|---------------------|----------------------|
| (a) Permalloy | (e) Piezoelectric |
| (b) Advance | (f) Magnetostrictive |
| (c) Phosphor bronze | (g) Strain gauge |
| (d) Quartz | (h) Spring |

5.7 In the context of transducers, indicate the correct choice:

- (a) A photoconductive transducer works on the principle that when a light beam strikes
- the material, its resistance decreases, which is sensed by an external circuit
 - the barrier between transparent metal layer and a semiconductor material, a voltage is generated
 - the barrier between transparent metal layer and a semiconductor material, a current is generated in the external circuit
 - the cathode, it releases electrons, which are attracted towards the anode, thereby producing electric current in the external circuit
- (b) Identify the correct set of matches:
- | | |
|-------------------------------|-------------------------|
| (a) Mean free path | (p) Optical pyrometer |
| (b) Humidity | (q) Knudsen gauge |
| (c) Heat transfer coefficient | (r) Sling psychrometer |
| (d) Intensity of radiation | (s) Hot wire anemometer |
- (a)→ (p), (b)→ (q), (c)→ (r), (d)→ (s)
 - (a)→ (q), (b)→ (p), (c)→ (s), (d)→ (r)
 - (a)→ (r), (b)→ (s), (c)→ (p), (d)→ (q)
 - (a)→ (q), (b)→ (r), (c)→ (s), (d)→ (p)
- (c) Identify the correct set of matches:
- | | |
|--------------------------|--------------------------------|
| (p) Thermocouple | (1) DC bridge |
| (q) Strain gauge | (2) Phase sensitive detector |
| (r) Piezoelectric sensor | (3) Charge amplifier |
| (s) LVDT | (4) Cold junction compensation |
| | (5) Instrumentation amplifier |
- (p)→ (2), (q)→ (3), (r)→ (5), (s)→ (1)
 - (p)→ (1), (q)→ (5), (r)→ (2), (s)→ (3)
 - (p)→ (4), (q)→ (1), (r)→ (2), (s)→ (5)
 - (p)→ (4), (q)→ (1), (r)→ (3), (s)→ (2)

Displacement Measurement

It is sometimes said that the measurement of displacement—linear or angular—is fundamental to all measurements. Many measurements, such as force, strain, pressure, temperature, level etc. boil down to the measurement of displacement in the ultimate analysis.

The displacement transducers can be broadly classified into the following categories:

1. Pneumatic
2. Electrical
3. Optical
4. Ultrasonic
5. Magnetostrictive
6. Digital

Of course, the common method of measuring displacements is using mechanical devices like scales—simple or Vernier, measuring tapes, micrometers, spherometers, etc. But these self-sufficient devices are so common that it is pointless to discuss them here. We will discuss only those transducers which can be used as components in an instrumentation system.

6.1 Pneumatic Transducers

The mostly used pneumatic transducer, called the *nozzle-flapper transducer* is an important pneumatic transducer that finds many applications in small displacement measurement. Though it can be used to measure small displacements, it is generally used to determine the null position in a servosystem.

Nozzle-flapper Transducer

Consider the arrangement shown in Fig. 6.1. A gas at a fixed pressure p_s flows through a nozzle past a restriction in the tube. An obstruction, called flapper¹ is placed close to the nozzle. Owing to the presence of the flapper, there will be a back pressure that will alter the pressure of the gas in the volume between the restriction and the nozzle, to p_o . A pressure transducer, such as a piezoelectric device² is attached to the volume to monitor the pressure there.

¹Meaning, 'a piece of something attached on one side only, that covers an opening', *Pocket Oxford Dictionary*, Oxford University Press (2004).

²See Section 5.2 at page 130.

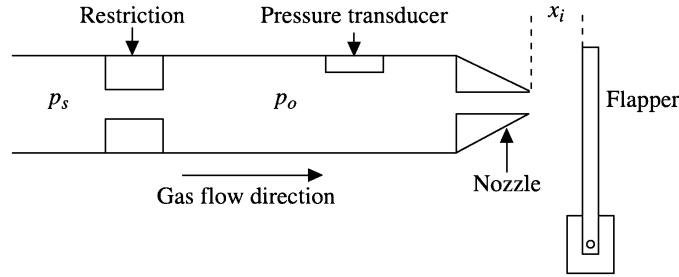


Fig. 6.1 Schematic diagram of nozzle-flapper transducer.

Clearly, the distance between the nozzle and the flapper will determine the pressure p_o which means, p_o can be calibrated in terms of the displacement x_i of the flapper. An approximate relation between p_o and x_i can be derived as follows.

If C_d is the discharge coefficient (see Section 11.2 at page 447 for definition)

d_s is the diameter of the supply orifice

ρ is the density of the fluid

then, assuming that the fluid is incompressible, the mass flow rate G_s through the supply orifice is given by

$$G_s = C_d \left(\frac{\pi d_s^2}{4} \right) \sqrt{2\rho(p_s - p_o)} \quad (6.1)$$

The flow from the nozzle spreads over a cylindrical volume of height x_i and diameter d_n which is the nozzle diameter. Therefore, the mass supply-rate through the nozzle G_n is given by

$$G_n = C_d(\pi d_n x_i) \sqrt{2\rho p_o} \quad (6.2)$$

neglecting the ambient pressure which is small in comparison to p_o and assuming the same discharge coefficient for the flow through the nozzle. Under equilibrium condition $G_s = G_n$. Then from Eqs. (6.1) and (6.2), we have

$$\begin{aligned} \frac{d_s^2}{4} \sqrt{p_s - p_o} &= d_n x_i \sqrt{p_o} \\ \Rightarrow \frac{p_o}{p_s} &= \frac{1}{1 + (16d_n^2 x_i^2 / d_s^4)} \end{aligned} \quad (6.3)$$

If we put $p_N = \frac{p_o}{p_s}$ and $x_N = \frac{d_n x_i}{d_s^2}$, Eq. (6.3) turns out to be

$$p_N = \frac{1}{1 + 16x_N^2} \quad (6.4)$$

A plot of p_N vs. x_N is shown in Fig. 6.2.

The operating point is chosen to keep the output pressure the same for equal displacement of the flapper on either side of its main position. Normally, the variation of p_N is nearly linear between 0.15 and 0.75. In industries, the supply pressure p_s is usually 20 psig (1.33 kg/cm²). Which means, p_o varies between 3 and 15 psig.

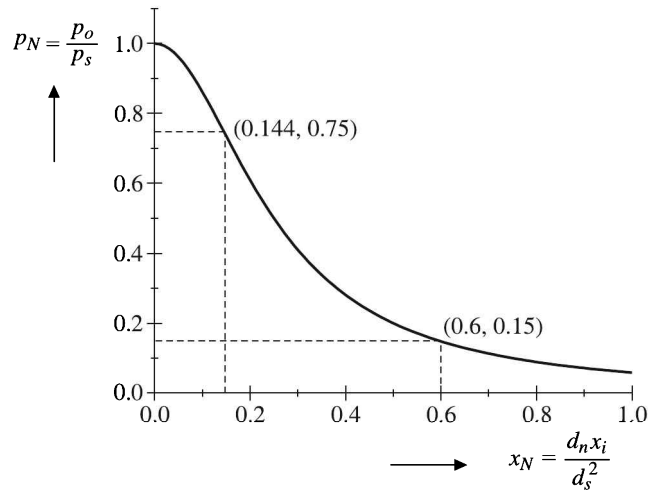


Fig. 6.2 Plot of p_N vs. x_N as given by Eq. (6.4). The operating area is between points (0.144, 0.75) and (0.6, 0.15) as shown in the diagram.

As low a displacement as 0.1 mm of the flapper produces an appreciable change in the output pressure p_o as may be seen from Example 6.1. This high sensitivity has made it rather popular in mechanical instrumentation. However, because of the approximately linear range of the transducer, it finds more application as a sensitive null detector in a servosystem rather than a final readout device.

Example 6.1

Calculate the variation of the output pressure for a nozzle-flapper transducer when the supply pressure is 20 psig, restriction and nozzle diameters are both 0.5 mm and the flapper movement is ± 0.05 mm.

Solution

Given: $d_n = d_s = 0.05$ cm, $p_s = 20$ psig and $x_i = \pm 0.005$ cm. We know, $p_N|_{\max} = 0.75$ and $p_N|_{\min} = 0.15$. Also

$$x_N = \pm \frac{(0.05)(0.005)}{(0.05)^2} = \pm 0.1$$

Therefore, from Eq. (6.4) the output pressure variation is given by

$$\begin{aligned} \Delta p_o &= (p_N|_{\max} - p_N|_{\min}) \cdot \frac{p_s}{1 + 16x_N^2} \\ &= (0.75 - 0.15) \cdot \frac{20}{1 + 16(0.1)^2} = 10.34 \text{ psig} \end{aligned}$$

Example 6.2

A pneumatic displacement gauge shown in Fig. 6.3 operates on the principle that the flow through the orifices of diameters D_1 and D_2 is governed by the separation distance x between the outlet and the workpiece.

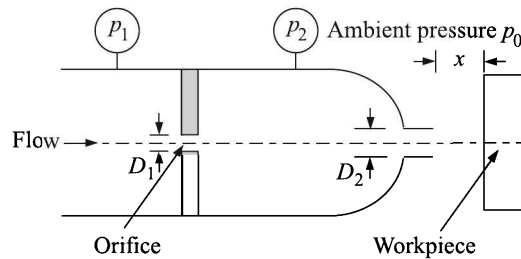


Fig. 6.3 Example 6.2

- (a) Obtain an expression for the pressure ratio $R = p_2/p_1$ as a function of the diameter ratio of the two orifices. Assume, ambient pressure $p_0 = 0$ and the discharge coefficients for the orifices to be equal.
- (b) Find the displacements for the pressure ratios of $R = 0.4$ and $R = 0.9$, if the orifice diameters are $D_1 = 0.5$ mm and $D_2 = 1.0$ mm.

Solution

(a) Modifying Eq. (6.3) suitably, the relation is given by

$$R = \frac{p_2}{p_1} = \frac{D_1^4}{D_1^4 + 16D_2^2x^2}$$

(b) From the above equation, we get the expression for the required displacement as

$$x = \sqrt{\frac{(1-R)D_1^4}{16RD_2^2x^2}} = \sqrt{\frac{1-R}{R}} \cdot \frac{D_1^2}{4D_2}$$

$$\text{For } R = 0.4 \quad x = \sqrt{\frac{0.6}{0.4}} \cdot \frac{(0.5)^2}{4} = 0.076 \text{ mm}$$

$$\text{For } R = 0.9 \quad x = \sqrt{\frac{0.1}{0.9}} \cdot \frac{(0.5)^2}{4} = 0.021 \text{ mm}$$

6.2 Electrical Transducers

Here our aim is to convert displacement to an electrical format. An electrical circuit consists basically of three variable passive components, namely resistance, inductance and capacitance. All three of them can be utilised to construct devices for transducing displacement.

Resistive Transducer: Potentiometer

The familiar potentiometer, or *pot* in common parlance, is widely used as a transducer. Basically, it consists of a resistance element provided with a movable contact. The motion of the contact can be translational, rotational, or helical which is a combination of the two former motions (Fig. 6.4).

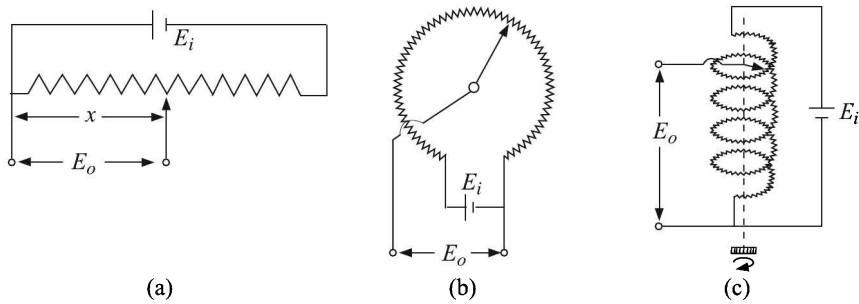


Fig. 6.4 Schematic representations of potentiometers: (a) translational, (b) rotary, and (c) helical.

Construction

Potentiometers are generally constructed in three forms—single slide wire, wire-wound, and cermet.

Single slide wire. The only advantage that the single slide wire offers is the stepless variation of resistance as the wiper travels over it. But since the length of the wire is limited by the desired stroke in a translational device and by the diameter in a rotational one, this type of potentiometer is limited to rather small values of resistance. Although the resistance per unit length can be increased by decreasing the area of cross-section of the wire, it is done only at the expense of its strength and resistance to wear.

Wire-wound. In this case the resistance wire is wound on a straight or circular card or a mandrel (Fig. 6.5), depending on the type of the device—translational or rotational—used.



Fig. 6.5 Wire-wound resistance elements on: (a) straight insulating card and (b) circular insulating card.

The wire-wound construction produces a stepwise increase in resistance (Fig. 6.6) as the wiper moves from one turn of the wire to another, thus imposing a restriction on the resolution of the transducer. For example, if there are 400 turns on a 20 mm long card, the resolution is $20 \div 400 = 50 \mu\text{m}$. In fact, the practical limit of winding is 20 to 40 turns/mm which restricts the resolution to 25 to 50 μm . For rotational devices the resolution R can be figured out from the relation

$$R = \frac{360 \times 10^{-3}}{\pi n D}$$

where D is the diameter of the potentiometer in metres and n is the number of turns/mm.

It may be noted in this context that a higher resolution demands thinner wires which, in turn, means a higher total resistance. Thus resolution and resistance are interdependent.

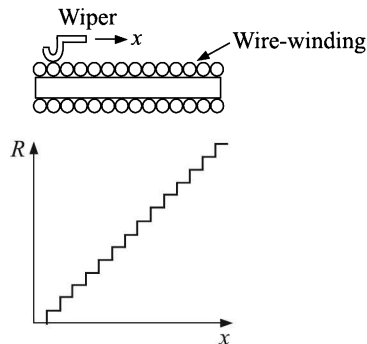


Fig. 6.6 Stepwise increase of resistance of wire-wound potentiometers.

Wires of nickel-chromium (nichrome), nickel-copper (constantan), silver-palladium or some other precious metals are used as resistive elements, their diameters varying between 25 and 50 μm . To avoid surface oxidation, they are annealed in a reducing atmosphere.

On the other hand, hard alloys like phosphor-bronze, beryllium-copper or other precious metal alloys are used to construct wipers and are shaped in such a way that they slide with minimum friction and at the same time maintain a firm contact with the winding.

Cermet. Precious metal particles fused into a ceramic base constitute cermet. This has many advantages such as:

1. Stepless variation of resistance offering a very high resolution
2. Large power ratings because it is not easily fusible
3. Low cost
4. Moderate temperature coefficients
5. Utility in ac applications

Apart from these, hot moulded carbon, carbon films, thin metal films are also used to construct potentiometers.

Characteristics

While choosing a potentiometer for an application, it is necessary to consider the following characteristics:

Loading effects. The resistance element is excited with dc or ac voltage and the input-output relation is ideally linear.

But in practice the voltage measuring arrangement loads the output and as a result the relation is far from linear as will be evident from the following analysis.

From Fig. 6.7(a) it is clear that resistance of the length $x_i = (x_i/x_t)R_p \equiv KR_p$, say. Here, x_t is the total length of the potentiometer wire and R_p is the total resistance of the potentiometer. The circuit in Fig. 6.7(a) can be redrawn to the form shown in Fig. 6.7(b) so that its Thevenin equivalent looks like Fig. 6.7(c). It is clear from Fig. 6.7(c) that the Thevenin equivalent resistance R_o of the potentiometer circuit is $K(1-K)R_p$ and the Thevenin equivalent input voltage E'_i is KE_i , where E_i is the actual input voltage.

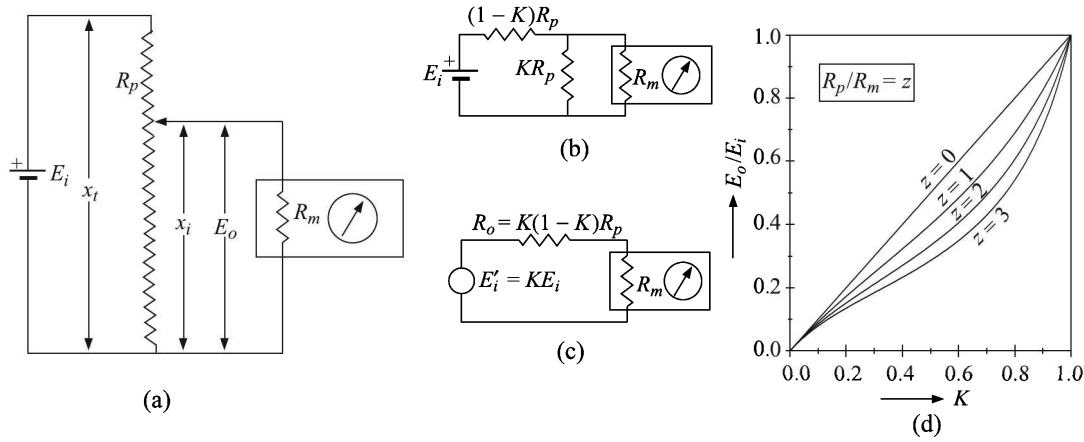


Fig. 6.7 Potentiometer loading effect: (a) circuit arrangement, (b) re-drawn circuit, (c) Thevenin equivalent circuit, and (d) characteristic curves.

Thus, if E_o is the output voltage, we have

$$\frac{E_o}{E_i} = \frac{1}{1 + \frac{R_o}{R_m}} = \frac{K}{1 + K(1 - K) \frac{R_p}{R_m}} \tag{6.5}$$

In actual practice, $R_m \neq \infty$ and, therefore, the characteristics curve is nonlinear. Table 6.1 as well as Fig. 6.7(c) will give an idea about the error caused by the loading effect.

Table 6.1 Error caused by loading of the potentiometer

| R_p/R_m | 1.0 | 0.1 | < 0.1 |
|-------------------|-----|-----|---------------|
| Maximum error (%) | 12 | 1.5 | $15(R_p/R_m)$ |

Power rating. The typical available power rating is 5 W at room temperature. The maximum excitation voltage can be calculated from the relation

$$(E_i)_{\max} = \sqrt{PR_p} \text{ volt}$$

where P is the rated power in watts.

Linearity and sensitivity. For high sensitivity, the output voltage E_o and for that matter input voltage E_i should be high. But the maximum value of E_i is determined by the resistance of the potentiometer R_p and its power rating. The value of R_p has to be kept low in comparison to the resistance of the measuring instrument R_m to achieve linearity. This requirement, thus, is in conflict with the desire for high sensitivity. Typical values of sensitivity are 200 mV/mm for translational or 200 mV/deg for rotational devices.

The advantages and disadvantages of potentiometric displacement transducers are given in Table 6.2.

Table 6.2 Advantages and disadvantages of potentiometers

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|---|
| 1. Inexpensive and simple to set up. | 1. Mechanical loading owing to wiper friction. |
| 2. Rather large displacements can be measured. | 2. Electrical noise from the sliding contact. |
| 3. Sufficient output to drive control circuits. | 3. Wear and misalignment owing to friction. |
| 4. Frequency response and resolution limited for the wire-wound, but unlimited for others. | 4. Quick manipulation generates heat and associated problems. |

Example 6.3

The output of a potentiometer is to be read by a 10 k Ω voltmeter, holding non-linearity to 1%. A family of potentiometers having a thermal rating of 5W and resistances ranging from 100 Ω to 10 k Ω in 100 Ω steps are available. Choose from this family the pot that has the greatest possible sensitivity and meets other requirements. What is the sensitivity if pots are single-turn (360°) units?

Solution

To hold linearity to 1%, $R_p = R_m/15 = 666.7 \Omega$. Pots available in this range are 600 Ω and 700 Ω . To ensure a high sensitivity we should choose 700 Ω , but then the nonlinearity goes above 1%. So, we have no alternative but to choose the 600 Ω pot. With this pot, the maximum excitation voltage is $\sqrt{5 \times 600} \cong 54.8$ V, and, therefore, the required sensitivity is $54.8/360 \cong 152$ mV/degree.

Example 6.4

In a potentiometer transducer, the potentiometer has a total resistance of 24 k Ω for a total wiper travel of 120 mm. During a measurement the wiper moves between 20 mm and 60 mm over the potentiometer.

- If the voltmeter of 15 k Ω is used to read the output voltage of the transducer, find out the error due to the loading effect at the two measuring points.
- If the error due to the loading effect in the above instrumentation is to be kept within $\pm 3\%$, what should be the resistance of the voltmeter?

Solution

(a) The wiper travels between 20 mm and 60 mm. The resistance R between these points is

$$R = \frac{40}{120} \times 24 = 8 \text{ k}\Omega$$

Let the excitation voltage be E . Therefore, the voltage V across R is

$$V = \frac{8}{24}E = \frac{E}{3} \cong 0.3333E \text{ V}$$

Now, the 15 k Ω resistance of the voltmeter lies parallel to R . Their combined resistance R_c is

$$R_c = \frac{15 \times 8}{15 + 8} = 5.217 \text{ k}\Omega$$

Therefore, the voltage V_c developed across R_c is

$$V_c = \frac{5.217E}{(24 - 8) + 5.217} = 0.2459E \text{ V}$$

Note: This result can be obtained by putting $K = x_i/x_l = 40/120 = 1/3$, $R_p = 24 \text{ k}\Omega$ and $R_m = 15 \text{ k}\Omega$ in Eq. (6.5).

Therefore, the error ε in the measurement is

$$\varepsilon = \frac{(0.3333 - 0.2459)E}{0.3333E} \times 100 = 26.22\%$$

(b) To keep the error within 3%, if V_{cx} is the voltage to be developed across the combined resistance of R and the unknown resistance R_{mx} of the measuring voltmeter, we have

$$3 = \frac{(0.3333 - V_{cx})E}{0.3333E} \times 100$$

This gives
$$V_{cx} = 0.3333E - \frac{3 \times 0.3333}{100}E = 0.3233E \text{ V}$$

So
$$0.3233E = \frac{R_{cx}}{16 + R_{cx}}E$$

or
$$R_{cx} = \frac{16 \times 0.3233}{1 - 0.3233} = 7.6442 \text{ k}\Omega$$

Now

$$R_{cx} = \frac{8R_{mx}}{8 + R_{mx}}$$

This gives
$$R_{mx} = \frac{8R_{cx}}{8 - r_{cx}} = \frac{8 \times 7.6442}{8 - 7.6442} \cong 172 \text{ k}\Omega$$

Example 6.5

A potentiometer is used to measure the displacement of a hydraulic ram. The potentiometer is 25 cm long, has a total resistance of 2500 ohms and is operating at 4 W with a voltage source. It has linear resistance-displacement characteristics. Determine

- Sensitivity of the potentiometer in volts/cm (without loading effect)
- Loading error in the measurement of displacement at actual input displacement of 15 cm, when the potentiometer is connected to a recorder having a resistance of 5000 ohms.

Solution

Given, $L = 25 \text{ cm}$, $R_p = 2500 \Omega$, $P = 4 \text{ W}$. Therefore, current in the circuit is

$$I = \sqrt{\frac{P}{R_p}} = \sqrt{\frac{4}{2500}} = 0.04 \text{ A}$$

and excitation voltage is

$$V = 2500 \times 0.04 = 100 \text{ V}$$

(a) Sensitivity = $\frac{100}{25} = 4 \text{ V/cm}$.

(b) Actual input displacement $x = 15 \text{ cm}$. Therefore, resistance across x is

$$R_x = \frac{15}{25} \times 2500 = 1500 \Omega$$

and actual voltage across R_x is

$$V_x = 15 \times 4 = 60 \text{ V}$$

The recorder has been connected parallel to R_x . Their combined resistance is

$$\frac{1500 \times 5000}{1500 + 5000} = 1153.85 \Omega$$

Hence the total resistance of the circuit is now

$$(1153.85 + 1000) = 2153.85 \Omega$$

Therefore, Voltage across $R_x = \frac{60}{2153.85} \times 1153.85 = 32.14 \text{ V}$

$$\text{Loading error} = \frac{60 - 32.14}{60} \times 100 = 46.4\%$$

Inductive Transducers

Inductive transducers can be of various types. We will consider only three, namely

1. Linear variable differential transformer
2. Rotary variable differential transformer
3. Synchros

Linear variable differential transformer (LVDT)

The linear variable differential transformer (LVDT) is the most commonly used variable inductance transducer in industry. It is an electromechanical device designed to produce an ac voltage output proportional to the relative displacement of a transformer and an iron core, as illustrated in Fig. 6.8.

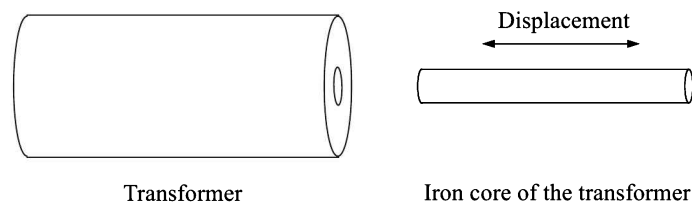


Fig. 6.8 Basics of LVDT.

The basic construction of the LVDT is shown in Fig. 6.9. It consists of one primary winding and two secondary windings. Secondary windings are identical in respect of their number of turns and their placement on both sides of the primary winding. A sinusoidal voltage e_i of amplitude 1 to 15 V and frequency 50 Hz to 20 kHz can be used to excite the primary though 1 V at 2 kHz to 10 kHz is common. A movable core of high μ produces signals proportional to its displacement by changing the mutual inductance between the coils. Nickel-iron alloy, slotted longitudinally to minimise eddy current loss, is normally used to construct the core.

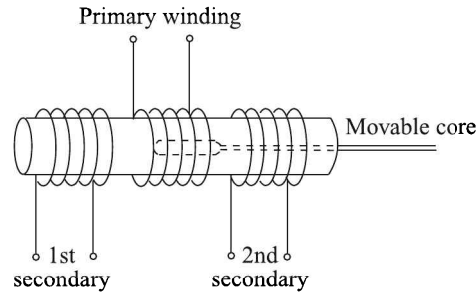


Fig. 6.9 Construction of LVDT.

When the core is in the middle position, a sinusoidal voltage of equal amplitude appears across the two secondaries (Fig. 6.10).

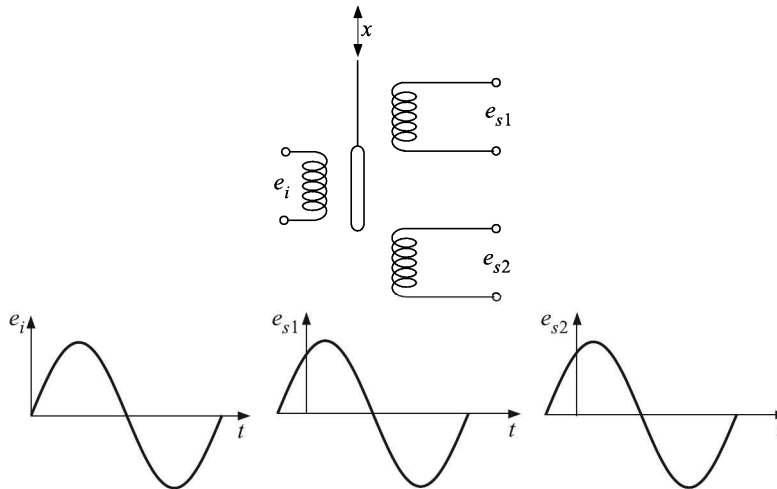


Fig. 6.10 Core in null position and the corresponding voltages.

And if the secondaries are connected in series opposition, as is normally the case, these voltages cancel each other to produce a null voltage (Fig. 6.11).

With the displacement of the core on either side of the null, the combined voltage of the secondaries increases linearly, undergoing a 180° phase-shift while passing through the null (Fig. 6.12). The output loses its linear relationship with displacement beyond some limits and this property restricts the range of the LVDT. The normal range is from $\pm 10 \mu\text{m}$ to $\pm 10 \text{mm}$.

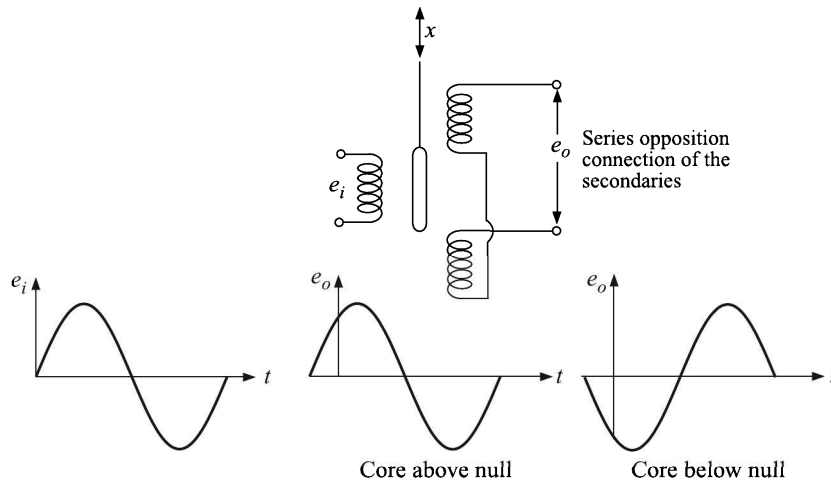


Fig. 6.11 Series opposing connection of secondaries and voltages for different core positions.

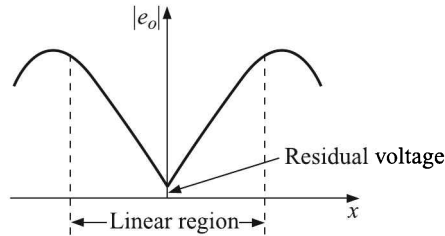


Fig. 6.12 Magnitude of output voltage for core displacement of an LVDT.

Circuit analysis. A simplified analysis of the circuit also reveals that the amplitude linearly varies with the difference in mutual inductances. When the secondary is open circuited, the equation for the primary can be written as

$$i_p R_p + L_p \frac{di_p}{dt} = e_i \tag{6.6}$$

where symbols have their usual meaning. On Laplace transformation Eq. (6.6) yields

$$(sL_p + R_p)I_p = E_i$$

or

$$I_p = \frac{E_i}{sL_p + R_p} \equiv \frac{E_i/R_p}{\tau_p s + 1}$$

where $\tau_p = L_p/R_p$. Now, if e_{s1} and e_{s2} are voltages generated in the secondary coils owing to their mutual inductances of coefficients M_1 and M_2 , equations for the secondaries and their Laplace transforms are

$$\begin{aligned} e_{s1} &= M_1 \frac{di_p}{dt} & e_{s2} &= M_2 \frac{di_p}{dt} \\ E_{s1} &= sM_1 I_p & E_{s2} &= sM_2 I_p \end{aligned}$$

Therefore

$$E_o \equiv E_{s1} - E_{s2} = (M_1 - M_2)sI_p = \frac{(M_1 - M_2)s/R_p}{\tau_p s + 1} E_i$$

Thus

$$\frac{E_o(s)}{E_i(s)} = \frac{s(M_1 - M_2)/R_p}{\tau_p s + 1}$$

whence

$$\frac{E_o(j\omega)}{E_i(j\omega)} = \frac{j\omega(M_1 - M_2)/R_p}{j\omega\tau_p + 1} \equiv \frac{\omega(M_1 - M_2)/R_p}{\sqrt{(\omega\tau_p)^2 + 1}} \angle\phi \quad (6.7)$$

where

$$\phi = \frac{\pi}{2} - \tan^{-1} \omega\tau_p$$

In Eq. (6.7) the expression preceding $\angle\phi$ represents the amplitude ratio of the output and the input. Since ω , R_p , τ_p and the input amplitude are constants for a given set-up, the amplitude of the output A_o can be written as

$$A_o = K(M_1 - M_2) \equiv K'x$$

where K and K' are constants and x is the displacement.

The value of $(M_1 - M_2)$ keeps on increasing with the displacement of the core up to a certain point and then it starts falling as the core moves past one of the secondaries.

Excitation frequency. For a good dynamic response of an LVDT, the excitation frequency must be much higher than the core-movement frequencies. This is necessary for distinguishing them in the amplitude modulated output signal. The rule of the thumb is to set

$$\frac{\text{Maximum core-movement frequency}}{\text{Excitation frequency}} = \frac{1}{10}$$

Residual voltage. Whereas the output voltage at null position is ideally zero, harmonics in the excitation voltage and stray capacitance coupling between primary and secondary usually result in a small but non-zero voltage which is called the *residual voltage*. Under usual conditions this is $< 1\%$ of the full-scale output and may be quite acceptable. Methods of reducing this null when it is objectionable are available.

Wiring variation. Most LVDTs are wired as shown in Fig. 6.11. This wiring arrangement is known as *open wiring*. Since the number of coil windings is uniformly distributed along the transformer, the voltage output is proportional to the iron core displacement when the core slides through the transformer. The corresponding equation is:

$$x = Se_o$$

where x is the displacement of the iron core with respect to the transformer, and S is the sensitivity of the transformer.

Another commonly used LVDT wiring is known as *ratio metric wiring*. It is like the wiring shown in Fig. 6.10 where secondary voltages are measured separately. The displacement for ratio metric LVDTs is given by

$$x = S \frac{e_{s1} - e_{s2}}{e_{s1} + e_{s2}}$$

Typical specifications. Typical specifications of an LVDT are given in Table 6.3.

Table 6.3 Typical specifications of an LVDT

| <i>Normal excitation voltage</i> | <i>Normal range</i> | <i>Operating temperature</i> |
|----------------------------------|---|-------------------------------|
| 1.0 V at 2 to 10 kHz | $\pm 10 \mu\text{m}$ to $\pm 10 \text{ mm}$ | -40 to $+100^\circ\text{C}$ |

Signal conditioning. The LVDT output is basically an amplitude modulated signal. Therefore, there is a possibility of a mix-up between the modulation frequency and the carrier frequency when the displacement varies sinusoidally. We have already told that the rule of thumb is to limit the modulation frequency within 10% of the carrier frequency in order to have a faithful dynamic response. This possibility can be avoided by using a high frequency excitation. Also, the direction of displacement of the core cannot be ascertained by measuring the output ac voltage (Fig. 6.13).

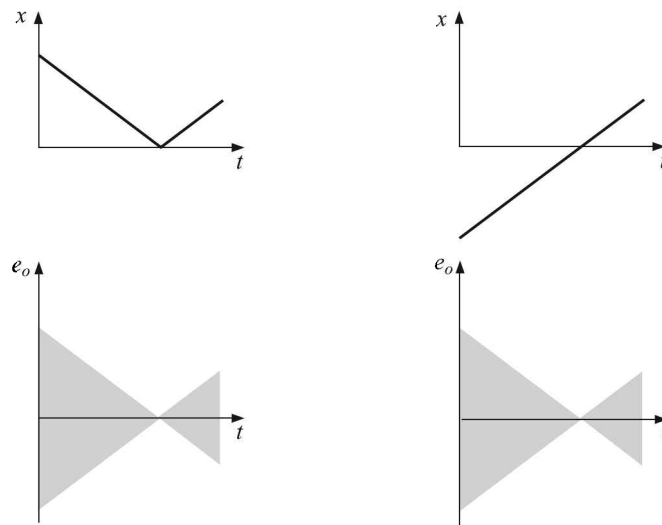


Fig. 6.13 Output ac voltage is virtually the same for two different patterns of core displacements.

Clearly, a phase-sensitive detection³ along with a low-pass filter is of help. The instrumentation of the LVDT will be apparent from Fig. 6.14.

³See Section 16.2 at page 771.

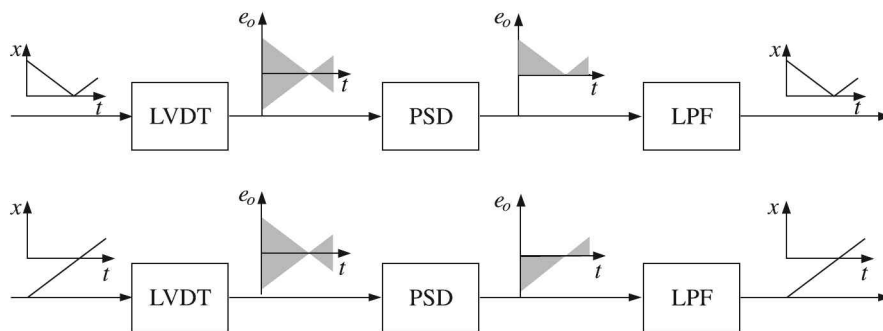


Fig. 6.14 Phase-sensitive detector in combination with a low-pass filter can distinguish between the two different patterns of core movements.

Table 6.4 will give an idea about the advantages and disadvantages of LVDT.

Table 6.4 Advantages and disadvantages of LVDT

| <i>Advantages</i> | <i>Disadvantages</i> |
|---|--|
| 1. Linearity is good up to 5 mm. | 1. Rather large threshold. |
| 2. Output voltage is stepless and hence the resolution is good ($\sim 1\mu\text{m}$). | 2. Affected by stray electromagnetic fields. Hence, proper shielding of the device is necessary. |
| 3. Output is rather high. Therefore, intermediate amplification is not necessary. | 3. AC input generates noises. |
| 4. Sensitivity is high ($\sim 40\text{ V/mm}$). | 4. Sensitivity is lower at higher temperatures. |
| 5. Low power and low hysteresis device. | |
| 6. Short response time, only limited by the inertia of the iron core and the rise time of the amplifiers. | |
| 7. Does not load the measurand mechanically. | |
| 8. Solid and robust, capable of working in a wide variety of environments. No permanent damage to the LVDT if measurements exceed the designed range. | |
| 9. Relatively low cost owing to its popularity. | |

Example 6.6

The output of an LVDT is connected to a 5 V voltmeter through an amplifier of amplification factor 250. The voltmeter scale has 100 divisions and the scale can be read to 1/5th of a division. An output of 2 mV appears across the terminals of the LVDT when the core is displaced through a distance of 0.5 mm. Calculate

- (a) The sensitivity of the LVDT
- (b) That of the whole set-up
- (c) The resolution of the instrument in mm

Solution

- (a) For a displacement of 0.5 mm, the output is 2 mV. Hence the sensitivity of the LVDT is $2 \div 0.5 \text{ mV/mm} = 4 \text{ mV/mm}$.
- (b) This sensitivity is amplified 250 times in the set-up. Hence the sensitivity of the set-up is $4 \times 250 \text{ mV/mm} = 1 \text{ V/mm}$.
- (c) The output of the voltmeter is 5 V with 100 divisions, which means each division equals 0.05 V. And since $(1/5)$ th of a division can be read, the minimum voltage that can be read is 0.01 V, which corresponds to 0.01 mm. Hence the resolution of the instrument is 0.01 mm.

Rotary variable differential transformer

The rotary variable differential transformer (RVDT) is used to measure rotational angles. It works on the LVDT principles. While the LVDT uses a cylindrical iron core, the RVDT uses a rotary ferromagnetic core. A schematic diagram is shown in Fig. 6.15.

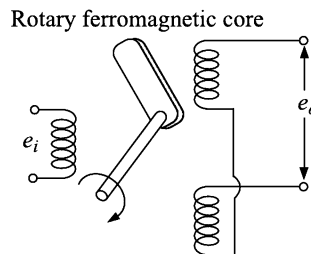


Fig. 6.15 Schematic diagram of an RVDT.

Synchros

Basically electromechanical devices based on variation of reluctance, synchros measure angles or perform functions related to angle measurement such as remote control of angle or computation of rectangular component of vectors.

A synchro consists of a wound rotor and a wound stator which are arranged concentrically so that the motion of the rotor produces a variable mutual inductance between the two windings (Fig. 6.16). The rotor is laminated, has a single-phase winding and is connected to supply lines through precision slip rings, while the stator winding is three-phase, the phases being displaced by 120° in space.

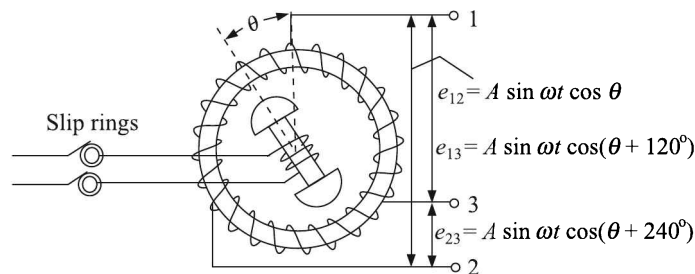


Fig. 6.16 Schematic presentation of a synchro.

Synchros are mostly used to compare the angular position of a load with its commanded position and to generate an appropriate feedback such that deviation is corrected automatically. Such automatic motion control feedback system is called *servomechanism*. For this purpose two synchros are used in conjunction. One is called the transmitter and the other, the receiver (Fig. 6.17).

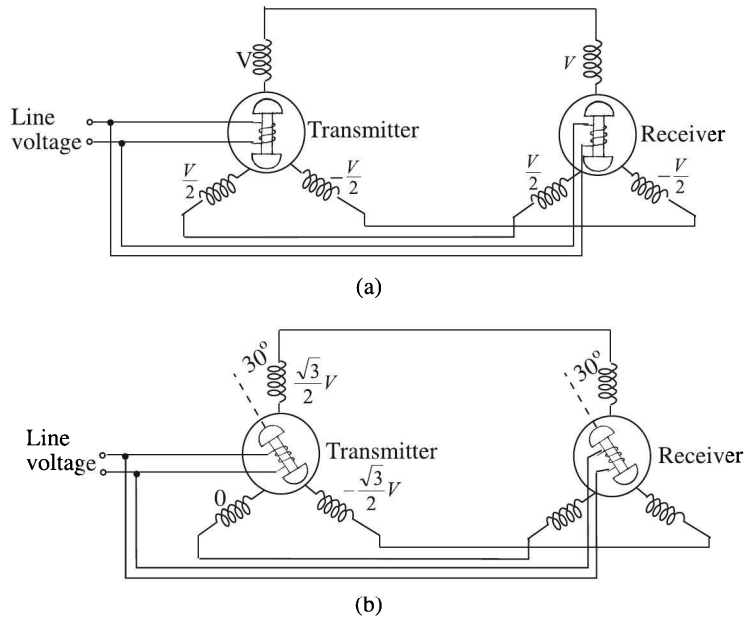


Fig. 6.17 Servomechanism application of synchros.

Suppose, initially the rotors and stators of the transmitter and receiver are in positions as shown in Fig. 6.17(a). Induced voltages on the three phases of the two stators will be $V/2$, V and $-V/2$ respectively, as indicated, because of the $\cos\theta$ coupling between the stator and rotor windings.

Now, suppose the transmitter-rotor rotates by 30° . The corresponding stator voltages will be 0 , $\frac{\sqrt{3}}{2}V$, and $-\frac{\sqrt{3}}{2}V$ as indicated in Fig. 6.17(b). As a consequence, there will be voltage unbalance between corresponding windings of receiver-stator. This will produce a current-flow with a consequent rotation of the receiver-rotor through an identical angle. This is how a synchro is used to produce a feedback of angular displacement.

Typical specifications of a synchro are given in Table 6.5.

Table 6.5 Typical specifications of a synchro

| <i>Normal excitation voltage</i> | <i>Sensitivity</i> | <i>Nonlinearity</i> |
|----------------------------------|--------------------|---------------------|
| 1.0 V at 50 Hz to 400 kHz | 1 V/deg | about 0.25 % |

Capacitive Transducers

Capacitive transducers can directly sense a variety of things—motion, chemical composition, electric field—and, indirectly, pressure, acceleration, fluid level, and fluid composition. The technology is low cost and stable and uses simple conditioning circuits. Capacitive displacement detectors can detect 10^{-8} m displacements with good stability and high speed under wide environmental variations.

Generally, parallel-plate capacitors are used as transducers. According to the theory, the capacitance C of such a capacitor is given by

$$C = \frac{\epsilon A}{x} \text{ farad} \quad (6.8)$$

where $\epsilon = \epsilon_0 \epsilon_r$ is the permittivity of the intervening medium (farad/metre)

x is the distance between the plates (metre)

A is the overlapping area of the plates (metre²)

Therefore, a variable capacitance device can be constructed by effecting variation in either of

1. Distance x between the plates [Fig. 6.18(a)]
2. Effective overlapping area A between the plates [Fig. 6.18(b)]
3. Relative permittivity ϵ_r of the intervening medium between the plates [Fig. 6.18(c)]

We will consider them in that order.

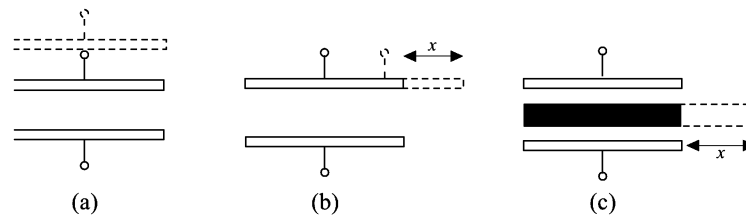


Fig. 6.18 Three kinds of variation in capacitive transducers: (a) change in the gap, (b) change in the area and (c) change in the permittivity.

Change in the gap x between the plates

Since capacitance varies inversely as x , the plot of C vs. x is a rectangular hyperbola. That means

$$\text{Sensitivity} = S = \frac{dC}{dx} = -\frac{\text{constant}}{x^2}$$

is not constant. This is rather inconvenient for measurements. In fact S decreases as x increases.

Linearisation. The linearisation of the input-output relationship is usually achieved by resorting to three different techniques:

1. By measuring the per cent change in capacitance
2. Using a charge amplifier
3. Measuring impedance
4. Differential arrangement

By measuring per cent change in capacitance the input output relation can be linearised. We observe that

$$\frac{dC}{dx} = -\frac{\epsilon A}{x^2} = -\frac{C}{x}$$

$$\Rightarrow \frac{dC}{C} = -\frac{dx}{x} \quad (6.9)$$

Equation (6.9) indicates that the per cent changes of C and x are linearly related provided, of course, the changes are small.

By using a charge amplifier the input-output relation can be linearised as shown in Fig. 6.19. With currents i and i_x , capacitors C and C_x , and voltages e_i and e_o as indicated in Fig. 6.19, we have

$$e_i = \frac{\int i dt}{C} \quad (6.10)$$

$$e_o = \frac{\int i_x dt}{C_x}$$

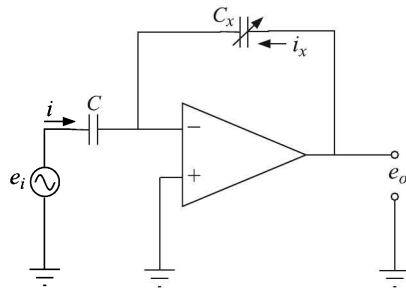


Fig. 6.19 Use of the op-amp to linearise the input-output relation.

Now, because the op-amp has a virtual ground at the input, $i = -i_x$. Hence,

$$e_o = \frac{\int i_x dt}{C_x} = -\frac{\int i dt}{C_x} \quad [\text{because } i_x = -i] \quad (6.11)$$

$$= -\frac{C}{C_x} e_i \quad [\text{using Eq. (6.10)}] \quad (6.12)$$

$$= -\frac{C e_i}{\epsilon A} x \quad [\text{using Eq. (6.8)}] \quad (6.13)$$

In Eq. (6.13), since all other factors are constant, the output voltage varies linearly with the displacement.

Measuring impedance rather than the capacitance is another way of linearisation. Because the impedance is given by

$$X_C = \frac{1}{2\pi f C} = \frac{x}{2\pi f \epsilon_r \epsilon_0 A} \quad (6.14)$$

where f is the frequency of the exciting voltage. It is obvious from Eq. (6.14) that the X_C vs. x curve is linear.

In a typical commercially available transducer, e_i is a 50 kHz sine wave. The output is rectified and fed to a dc voltmeter calibrated directly in distance units.

Having a differential arrangement of capacitors, as shown in Fig. 6.20 is another technique of producing a linear transfer characteristic.

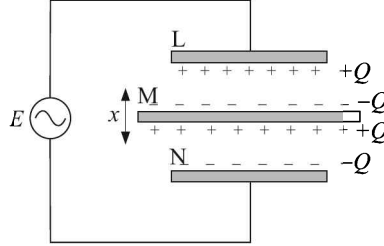


Fig. 6.20 Differential arrangement of capacitors to linearise the output.

Here, a three-plate capacitor is used, keeping the end plates (L and N) fixed and allowing the middle plate M to move. It can be shown that the voltage differential ΔE has a linear relation with the displacement $|x|$ of the movable plate M. At any instant,

$$E_{LM} = \frac{Q}{C_{LM}} = \frac{EC_{LN}}{C_{LM}} \quad (6.15)$$

$$E_{MN} = \frac{Q}{C_{MN}} = \frac{EC_{LN}}{C_{MN}} \quad (6.16)$$

$$C_{LN} = \frac{\varepsilon A}{2d}$$

where Q is the amount of charge on any plate and d is the distance between two adjacent plates.

When M is right at the midway, $C_{LM} = C_{MN}$ and therefore, the voltage differential is zero. If M is displaced upwards by a distance x , then

$$C_{LM} = \frac{\varepsilon A}{d - x} \quad (6.17)$$

$$C_{MN} = \frac{\varepsilon A}{d + x} \quad (6.18)$$

where A is the area of the plates and d is the distance between L and M (or M and N) when M is at the midway. Plugging values from Eqs. (6.17) and (6.18) in Eqs. (6.15) and (6.16), we get

$$E_{LM} = E \cdot \frac{d - x}{2d}$$

$$E_{MN} = E \cdot \frac{d + x}{2d}$$

which give

$$\Delta E = E_{LM} - E_{MN} = \frac{E}{d}x$$

This arrangement, with appropriate instrumentation, can measure displacements between 10^{-8} mm and 10 mm with an accuracy of about 0.1%.

Effect of fringing flux. The variation in spacing x of parallel plates is often used for displacement detection if $x < l$ or w where l and w are the length and width of the electrodes respectively. As long as the l and w of the plates are close compared to the plate spacing, the equations given above will produce more or less accurate results. But as the plate spacing increases relative to the l and w of the plates, more flux lines connect from the edges and backs of the plates. This is called *fringing*. With fringing, the measured capacitance can be much larger than calculated.

Fringing incorporates nonlinearity in the C vs. x^{-1} relation. This effect may very largely be eliminated by introducing what is called a *guard ring* as shown in Fig. 6.21.

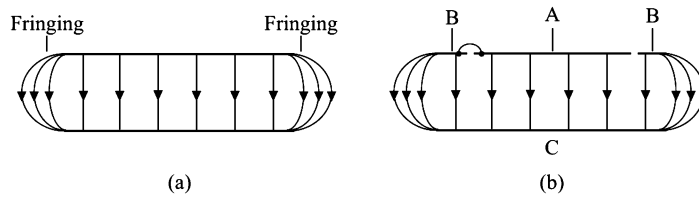


Fig. 6.21 (a) Fringing of flux at the ends of plates of a parallel-plate capacitor. (b) Guard ring (B) to eliminate the effect of fringing in a capacitor. B is at the same electrical potential with A.

One of the circular plates A of the capacitor is surrounded by a concentric annular plate B in the same plane. The inner radius of B is slightly larger than the radius of A and a metallic connection between A and B is made so that they will have the same potential. The other plate C of the capacitor is placed parallel to A and B. It is obvious that the flux between the plates will have edge effects on B, but that between A and C will be practically parallel. However, some correction will be necessary to calculate the capacity of the arrangement because the area of the plates should be A' and not equal to that of A because of its connection to B.

Example 6.7

Figure 6.22 shows a circuit with a variable air gap parallel plate capacitor as the sensing element. Show that the circuit acts as a velocity sensor for very small displacements, and find the proportionality constant between the voltage e_o and the input velocity v . Nominal (zero displacement) capacitance C is 50 pF and the nominal (zero displacement) distance between the capacitor plates x_0 is 5 mm.

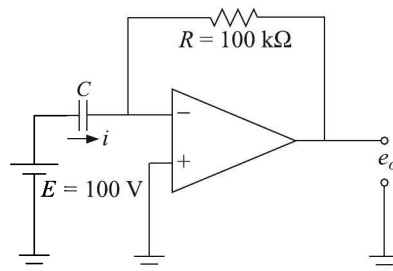


Fig. 6.22 Variable air gap parallel plate capacitor (Example 6.7).

Solution

We know, $C = \frac{\epsilon A}{x}$, where terms have their usual meaning. Therefore,

$$\begin{aligned} \frac{dx}{dt} &= \frac{d}{dt} \left(\frac{\epsilon A}{C} \right) = -\frac{\epsilon A}{C^2} \frac{dC}{dt} = -\frac{Cx}{C^2} \frac{dC}{dt} \\ &= -\frac{x}{C} \frac{dC}{dt} \end{aligned} \tag{i}$$

We also know, $C = \frac{Q}{E}$. Therefore,

$$\frac{dC}{dt} = \frac{1}{E} \frac{dQ}{dt} = \frac{i}{E}$$

where i denotes current. Substituting the value of $\frac{dC}{dt}$ in Eq. (i), we get on rearranging

$$i = -\frac{EC}{x} \frac{dx}{dt}$$

Therefore,

$$e_o = -iR = \frac{ECR}{x} \frac{dx}{dt} = \frac{(100)(50 \times 10^{-12})(100 \times 10^3)}{0.5} \frac{dx}{dt} = 1.0 \times 10^{-3} v$$

Example 6.8

Figure 6.23(a) shows the variable displacement type capacitor sensor with push-pull arrangement.

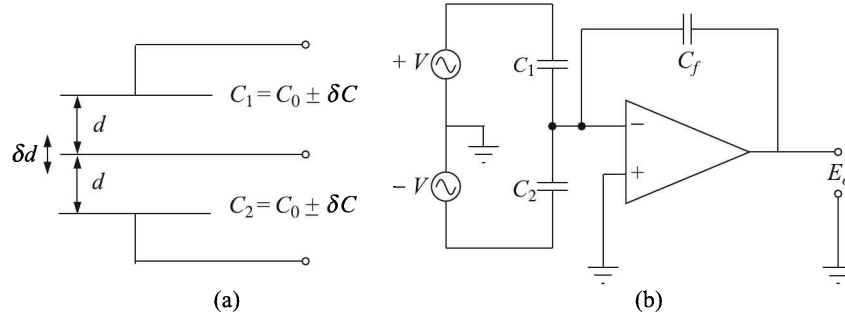


Fig. 6.23 Variable displacement type capacitor sensor (Example 6.8).

If A is the plate area, d the normal plate displacement, ϵ the permittivity and δd the input displacement,

- (a) Calculate $\delta C/C_0$ where C_0 is the nominal capacitance and δC the change in capacitance corresponding to δd .
- (b) Show that the output voltage is

$$E_o = V \frac{C_0}{C_f} \left[\frac{2(\delta d/d)}{1 - (\delta d/d)^2} \right]$$

when the push-pull configuration is connected as shown in Fig. 6.23(b).

Solution

We know, the capacitance C is given by

$$C_0 = \frac{\epsilon A}{d}$$

where terms have their usual significance.

(a) Owing to the displacement δd of the middle plate, C_1 becomes

$$C_1 = \frac{\epsilon A}{d - \delta d} = \frac{\epsilon A}{d} \cdot \frac{1}{1 - (\delta d/d)} = \frac{C_0}{1 - (\delta d/d)}$$

Thus,

$$\frac{\delta C_1}{C_0} = \frac{1}{1 - (\delta d/d)} \qquad \frac{\delta C_2}{C_0} = \frac{1}{1 + (\delta d/d)}$$

Note: Since C_1 and C_2 are connected in series, it is easy to see that the total capacitance between the end plates remains the same, irrespective of the movement of the central plate.

(b) The given circuit to which the push-pull capacitors are connected in Fig. 6.23(b) is a charge amplifier. Because the voltage V is connected in the opposite way to the two capacitors, the charge at the input to the op-amp is

$$VC = -VC_1 + VC_2 = -VC_0 \left[\frac{1}{1 - (\delta d/d)} - \frac{1}{1 + (\delta d/d)} \right] = -VC_0 \left[\frac{2(\delta d/d)}{1 - (\delta d/d)^2} \right]$$

Now from Eq. (6.12), we get

$$e_o = -\frac{VC}{C_f} = \frac{VC_0}{C_f} \cdot \frac{2(\delta d/d)}{1 - (\delta d/d)^2}$$

Change in the effective overlapping area A between the plates

In the x -variation motion detectors as discussed above, when the displacement is as large as the dimension of the electrodes, the accuracy of measurement suffers from the vanishing signal level. The area variation measurement is then preferred.

Parallel-plate capacitor. For a parallel-plate capacitor, having two exactly equal rectangular parallel plates placed one on top of another, if the width of the plate is given by w and the overlap length by l then the area of overlap is lw . Hence,

$$C = \frac{\epsilon lw}{x}$$

Now, if one of the plates is displaced along l , the area of overlap changes. But here, sensitivity = $dC/dl = \epsilon w/x = \text{constant}$. That means, the C vs. l plot is a straight line and, therefore, no extra circuitry is needed to make the calibration linear.

But strictly speaking, some nonlinearity is introduced owing to the fringing effect and this feature restricts the precision of the measurement. This type of transducer is suitable for measurement of linear displacements between 1 cm and 10 cm, the maximum attainable precision being about 0.005% which is quite satisfactory.

Co-axial parallel cylinders. Co-axial parallel cylinders are also conveniently used [Fig. 6.24(a)]. Here if d_1 is the outer diameter of the inner cylinder and d_2 the inner diameter of the outer cylinder, the capacitance is given by

$$C = \frac{2\pi\epsilon l}{\ln(d_2/d_1)}$$

All other factors remaining constant, here $C \propto l$ as well.

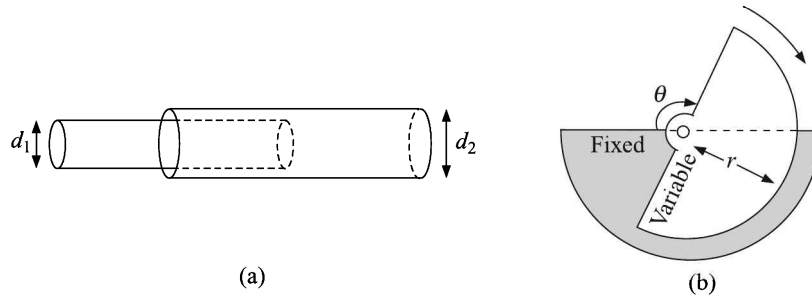


Fig. 6.24 Variable area capacitor transducers—(a) coaxial cylinder type, (b) semi-circular plate type.

Semicircular parallel plates. By making parallel plates semi-circular [Fig. 6.24(b)] angular displacements can also be measured by this method. In such an arrangement, the capacitance is the maximum when plates completely overlap, i.e. when $\theta = 180^\circ$. Then,

$$C_{\max} = \frac{\epsilon A}{x} = \frac{\pi\epsilon r^2}{2x}$$

while at any angle θ degree, the capacitance is

$$C = \frac{\pi\epsilon r^2}{360x}\theta$$

Here also $C \propto \theta$ because all other factors remain constant during an angular displacement.

Serrated parallel electrodes. Serrated parallel electrodes may be used to measure small angular variations. A pair of flat parallel serrated electrodes as shown in Fig. 6.25 are used.

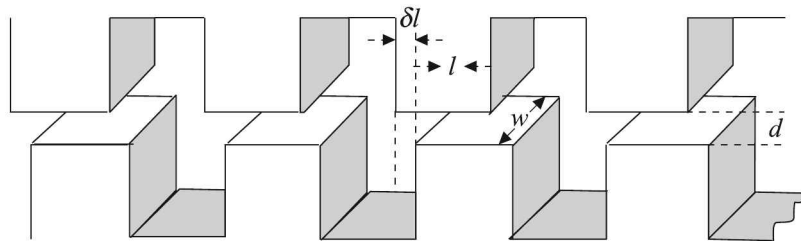


Fig. 6.25 Serrated electrode capacitor having variable effective tooth length.

If l is the active length of a tooth pair
 w is the width of a tooth
 d is the distance between the pair of nearby teeth
 n is the number of pairs of teeth

then the capacitance of the serrated electrode air capacitor is given by

$$C = \frac{\varepsilon_0 n l w}{d}$$

assuming no fringing at the corners of teeth. Now, if there is a small displacement δl , the change in capacitance is

$$\delta C = \frac{\varepsilon_0 n (l + \delta l) w}{d} - \frac{\varepsilon_0 n l w}{d} = \frac{\varepsilon_0 n l w}{d} \left(\frac{\delta l}{l} \right) = C \left(\frac{\delta l}{l} \right)$$

$$\Rightarrow \frac{\delta C}{C} = \frac{\delta l}{l} \quad (6.19)$$

The linear relation given Eq. (6.19) is approximate because substantial fringing will take place. The corrected equation, taking fringing into consideration, is given by

$$\frac{\delta C}{C} = \frac{\delta l}{l} \cdot \frac{1}{1 + k(d/l)} \equiv S \cdot \frac{\delta l}{l} \quad (6.20)$$

where k is constant for a set-up

$$S \text{ is often called the } \textit{sensitivity factor} = \frac{1}{1 + k(d/l)}$$

The variation of the sensitivity factor with the normalised active tooth length d/l for $k = 4/3$ is displayed in Fig. 6.26. It may be seen that S is nonlinear and it decreases as d/l increases.

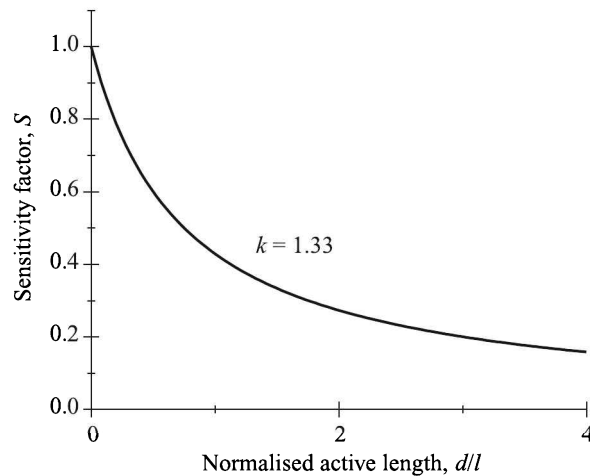


Fig. 6.26 Variation of sensitivity with normalised active tooth length of a serrated electrode capacitor.

Example 6.9

If the air gap between the teeth of two electrodes of a serrated type capacitive transducer is 0.1 cm and the active tooth length is 1 cm, what is the sensitivity factor of the sensor? Assume the constant term as 4.

Solution

We know from Eq. (6.20) that the sensitivity factor is given by

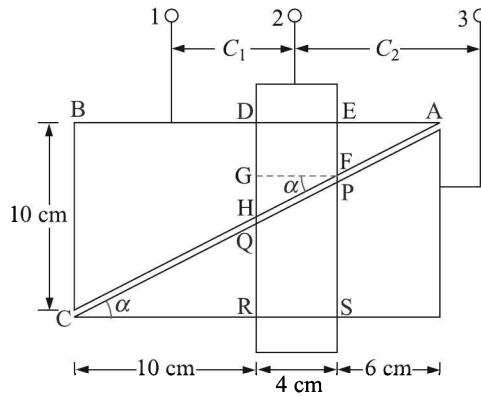
$$S = \frac{1}{1 + (kd/l)}$$

Given: $l = 1$ cm, $d = 0.1$ cm and $k = 4$. Therefore, the required sensitivity factor is

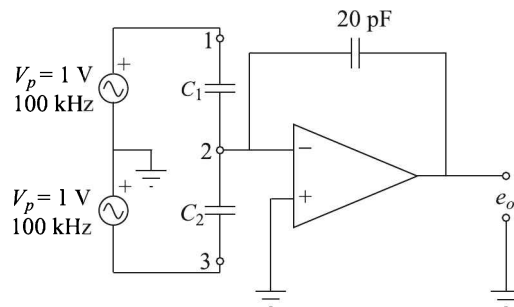
$$S = \frac{1}{1 + [(4)(0.1)]/(1)} = 0.714$$

Example 6.10

A capacitive type displacement transducer consists of two triangular plates, placed side by side, with a negligible gap in between them and a rectangular plate moving laterally with an air gap of 1 mm between the fixed plates and the moving plate. The schematic diagram, indicating appropriate dimensions, is shown below.



- (a) With the position of the moving plate shown in the figure above, what are the values of the capacitances C_1 and C_2 thus formed?
- (b) The above sensor is incorporated in a capacitance measuring circuit as shown in the following figure.



Assuming an ideal op-amp, what is the output voltage under the conditions mentioned above?

Solution

Given: BC = 10 cm, AB = 20 cm, and the air gap between the fixed and moving plates $d = 1 \text{ mm} = 10^{-3} \text{ m}$.

(a) Now

$$\tan \alpha = \frac{BC}{AB} = \frac{10}{20} = \frac{1}{2} = \frac{GH}{FG}$$

$$\therefore GH = \frac{FG}{2} = \frac{4}{2} = 2 \text{ cm} \quad DH = \frac{BC}{2} = \frac{10}{2} = 5 \text{ cm}$$

$$EF = DG = DH - GH = 5 - 2 = 3 \text{ cm}$$

Area of DEFH

$$A_1 = \square DEFC + \triangle FGH = \left(4 \times 3 + \frac{4 \times 2}{2}\right) = 16 \text{ cm}^2 = 16 \times 10^{-4} \text{ m}^2$$

Area of PQRS

$$A_2 = \square DESR - \square DEFH = 4 \times 10 - 16 = 24 \text{ cm}^2 = 24 \times 10^{-4} \text{ m}^2$$

Thus,

$$C_1 = \frac{\epsilon_0 A_1}{d} = \frac{(8.85 \times 10^{-12})(16 \times 10^{-4})}{10^{-3}} = 14.17 \text{ pF}$$

$$C_2 = \frac{\epsilon_0 A_2}{d} = \frac{(8.85 \times 10^{-12})(24 \times 10^{-4})}{10^{-3}} = 21.25 \text{ pF}$$

(b) This being a charge amplifier circuit, the input charge

$$Q = VC_2 - VC_1 = V(C_2 - C_1) = 7.08 \text{ pC}$$

Therefore, from Eq. (6.11) we have

$$e_o = \frac{Q}{C} = \frac{7.08}{20} = 0.354 \text{ V}$$

Change in the relative permittivity ϵ_r

The change in the relative permittivity can be effected in the following two ways.

Lateral shift of the dielectric. Let us consider a rectangular parallel-plate capacitor with each plate of length L and width w , and d is the distance between the plates (Fig. 6.27). We consider an initial situation when the intervening dielectric slab of length L , width w and thickness d is displaced in such a way that only the length l of it is within the plates. Then the capacitance of the arrangement is

$$C = \epsilon_0 \frac{w(L-l)}{d} + \epsilon_0 \epsilon_r \frac{wl}{d}$$

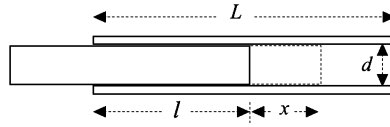


Fig. 6.27 Variable permittivity type capacitive transducer with a lateral shift of the dielectric.

Now, suppose we give the dielectric a displacement of x as indicated in Fig. 6.27. Then,

$$C + \Delta C = \epsilon_0 \frac{w(L - l - x)}{d} + \epsilon_0 \epsilon_r \frac{w(l + x)}{d}$$

or

$$\Delta C = \frac{\epsilon_0 w (\epsilon_r - 1)}{d} x$$

Thus

$$\Delta C \propto x$$

The relative permittivity of a few common dielectrics are given in Table 6.6.

Table 6.6 Relative permittivity of common dielectrics

| Material | Vacuum | Air | Polythene | Silica | Quartz | Glass | Mica | Water | BaTiO ₃ |
|--------------|--------|--------|-----------|--------|--------|---------|------|-------|--------------------|
| ϵ_r | 1.00 | 1.0005 | 2.3 | 3.8 | 4.5 | 5.3–7.5 | 7 | 80 | 10^3 – 10^5 |

Vertical shift of the movable plate. Now, let us consider the situation when the permittivity of the intervening medium changes owing to the vertical shift of one of the plates of the capacitor. Our capacitor consists of a solid dielectric of relative permittivity ϵ_r and of thickness d_1 plus a variable air gap of initial thickness d_2 (Fig. 6.28).

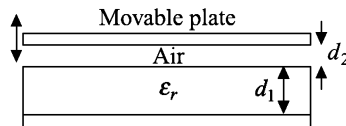


Fig. 6.28 Vertical shift of plate causing a variable permittivity.

Then, the capacitance of the arrangement is given by

$$C = \frac{\epsilon_0 A}{d_2 + (d_1/\epsilon_r)} \tag{6.21}$$

where A indicates the area of each of the plates and the dielectric. Next, we displace the movable top plate so that the air gap becomes $(d_2 - \delta d_2)$. This increases the capacitance by δC and we have

$$C + \delta C = \frac{\epsilon_0 A}{d_2 - \delta d_2 + (d_1/\epsilon_r)} \tag{6.22}$$

Subtracting Eq. (6.21) from Eq. (6.22), we get on rearrangement

$$\begin{aligned} \delta C &= \frac{\varepsilon_0 A}{d_2 + \frac{d_1}{r}} \cdot \frac{\delta d_2}{d_2 - \delta d_2 + \frac{d_1}{\varepsilon_r}} = C \cdot \frac{\delta d_2}{d_2 - \delta d_2 + \frac{d_1}{\varepsilon_r}} && \text{[Using Eq. (6.21)]} \\ \Rightarrow \frac{\delta C}{C} &= \frac{\delta d_2}{d_2 - \delta d_2 + \frac{d_1}{\varepsilon_r}} = \frac{\delta d_2}{d_1 + d_2} \cdot \frac{d_1 + d_2}{d_2 - \delta d_2 + \frac{d_1}{\varepsilon_r}} && (6.23) \end{aligned}$$

Now,

$$\frac{d_1 + d_2}{d_2 - \delta d_2 + \frac{d_1}{\varepsilon_r}} = \frac{S}{1 - S \cdot \frac{\delta d_2}{d_1 + d_2}}$$

where

$$S \equiv \frac{d_1 + d_2}{d_2 + \frac{d_1}{\varepsilon_r}} = \frac{1 + \frac{d_1}{d_2}}{1 + \frac{d_1}{d_2 \varepsilon_r}} \quad (6.24)$$

Substituting Eq. (6.24) in Eq. (6.23), we get

$$\frac{\delta C}{C} = \frac{\delta d_2}{d_1 + d_2} \cdot \frac{S}{1 - S \cdot \frac{\delta d_2}{d_1 + d_2}} \quad (6.25)$$

S is called the *sensitivity factor*. From Eq. (6.24) we find that its value depends on

1. The ratio (d_1/d_2)
2. The relative permittivity ε_r of the dielectric layer

Owing to the presence of S in the denominator of Eq. (6.25), it is obvious that the per cent change of capacitance is nonlinear. But for small δd_2 , the nonlinear term can be neglected and the relation becomes linear.

Frequency response of capacitive transducers

Capacitance devices are excited by ac and therefore it is necessary to study their behaviour with respect to frequency of the ac supply. We study here the frequency response for a simple circuit [Fig. 6.29(a)]. As long as the capacitor plates are stationary, no current flows through the circuit and therefore $e_o = E_b$. The moment the distance between the capacitor plates starts changing, the capacitor behaves as a voltage generator. Therefore, in the frequency domain we may write for the circuit

$$e_o(j\omega) = \frac{R}{R + (1/j\omega C)} e_i(j\omega) \quad (6.26)$$

where $e_i(j\omega)$ is the input voltage generated by the capacitor. Now, at any instant

$$e_i \propto \frac{1}{C} = \frac{x}{\varepsilon A}$$

or

$$e_i = Kx \quad (6.27)$$

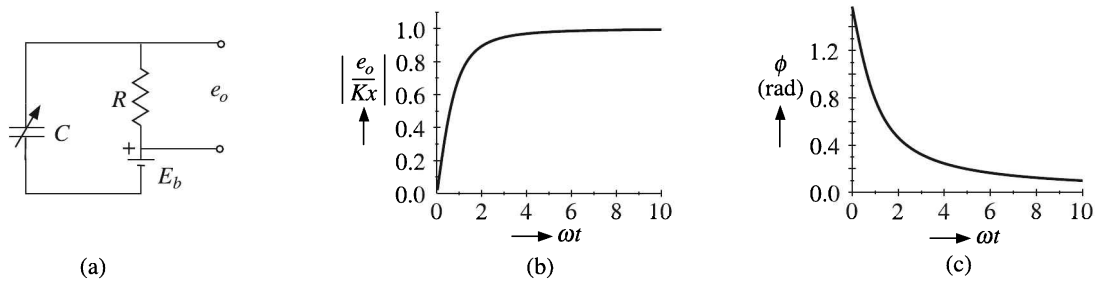


Fig. 6.29 Capacitive transducer: (a) a simple circuit, (b) the magnitude, and (c) phase of its frequency response.

where x is the displacement of the movable plate of the capacitor (assumed to be parallel plate type) and K is a constant which represents the sensitivity of the capacitive transducer. Substituting the value of e_i from Eq. (6.27) in Eq. (6.26), and putting $\tau = RC$, we get

$$\frac{e_o}{Kx}(j\omega) = \frac{j\omega\tau}{j\omega\tau + 1} \equiv \frac{1}{\sqrt{1 + (1/\omega\tau)^2}} \angle \phi \quad (6.28)$$

where

$$\phi = \frac{\pi}{2} - \tan^{-1} \omega\tau$$

The frequency response curves are shown in Fig. 6.29(b). From Eq. (6.28) we find, for $\omega\tau \gg 1$, $e_o/Kx \approx 1$, and, $\phi \rightarrow 0$. Thus e_o faithfully follows x under this condition.

For low frequencies, τ has to be large to make $\omega\tau \gg 1$. Now, $\tau = RC$, and for a certain x , C remaining fixed τ can be increased by increasing R only. Typically, R is about 1 M Ω or more. This necessitates that the voltage measurement device must have a very high (10 M Ω or more) input impedance.

In general, the impedance of a capacitor is very high. The following example will show that.

Example 6.11

Calculate the capacitance of an air capacitor of 3 cm \times 3 cm plates, separated by 0.3 mm and find its impedance at 10 kHz.

Solution

The required capacitance is

$$C \cong \frac{8.9 \times 10^{-12} \times 3 \times 10^{-2} \times 3 \times 10^{-2}}{0.3 \times 10^{-3}} \text{ F} \cong 26.7 \text{ pF}$$

At 10 kHz, its impedance is

$$X_c = \frac{1}{2\pi\nu C} \cong \frac{1}{2\pi \times 10^4 \times 26.7 \times 10^{-12}} \Omega \cong 0.6 \text{ M}\Omega$$

Such high impedance causes many problems such as development of noise voltages and making the transducer sensitive to the length and position of connecting cables. These call for a careful design of the output circuitry.

The merits and demerits of capacitive transducers are given in Table 6.7.

Table 6.7 Merits and demerits of capacitive transducers

| <i>Merits</i> | <i>Demerits</i> |
|--|--|
| 1. High sensitivity. | 1. Noise generation from stray capacitance arising from the transmission line. |
| 2. Good frequency response. | 2. Temperature sensitivity. |
| 3. High input impedance. So, not much loading. | 3. Complex instrumentation. |
| 4. Minimum mechanical loading. | |

6.3 Optical Transducers

Optical devices, such as various interferometers, can help us measure very small displacements to a very high degree of precision. So far those had been typical stand-alone laboratory instruments which could hardly be integrated to an instrumentation system. Of late laser-based transducers for measuring displacements have been made available commercially.

Laser Transducers

The laser displacement transducers measure the displacement by two methods — triangulation and time-of-flight (TOF).

Triangulation measurement

For distances of a few inches with high accuracy requirements, laser triangulation transducers measure the location of the spot within the field of view of the detecting element. They are so named because the transducer enclosure, the emitted laser and the reflected laser light form a triangle. A schematic diagram is shown in Fig. 6.30.

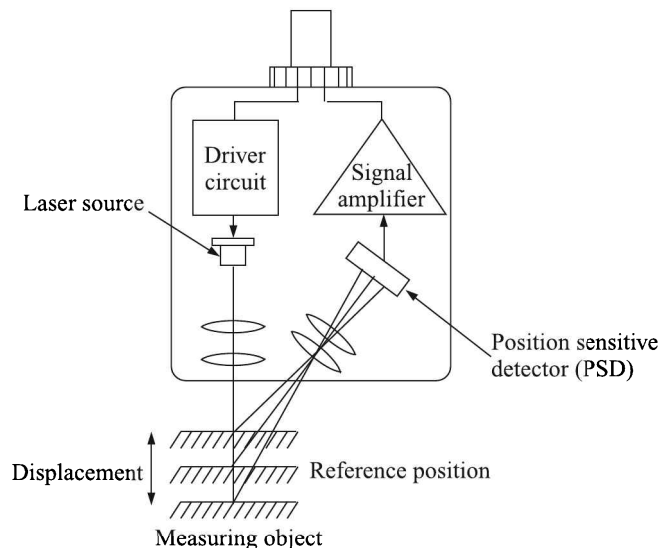


Fig. 6.30 Laser displacement transducer.

These transducers consist of a semiconductor laser as light source. The laser beam is projected from the instrument and is reflected from the target surface and focussed via an optical lens system onto a light sensitive receiving element. The receiving element, located at an angle that varies from 45 to 65 degrees at the centre of the measurement range, may be a position sensitive device (PSD) or a charge coupled device (CCD). If the target changes its position from the reference point, the position of the projected spot on the detector changes as well. The signal conditioning electronics of the laser detects the current generated by the receiving element. When this current is maximum, the lens system has focussed on the target. The corresponding distance is the location of the target.

Receiving element. The most critical component in the optical triangulation system is the light receiving element—the PSD or the CCD.

PSD. The PSDs are analogue detectors that rely on a current generated in a photodiode divided into one or two resistive layers. The amount of current from each output is proportional to the reflected light spot position on the detector.

The PSD receiver finds the position of the *centre of the average light intensity distribution*. There is no other information that the PSD element can provide. Now, the maximum of the averaged light intensity may not correspond to the peak light intensity. Depending on surface conditions, target texture or tilt, the intensity distribution of the reflected light spot also changes. This change in the intensity distribution will then change the centre of the average light distribution, even though the true peak position of the reflected light from the target has not changed [Fig. 6.31(a)].

Another aspect that affects the PSD performance is that they are very sensitive to light intensity. As a result, if the ambient light intensity changes while the spot position remains the same, it will result in a change in the output.

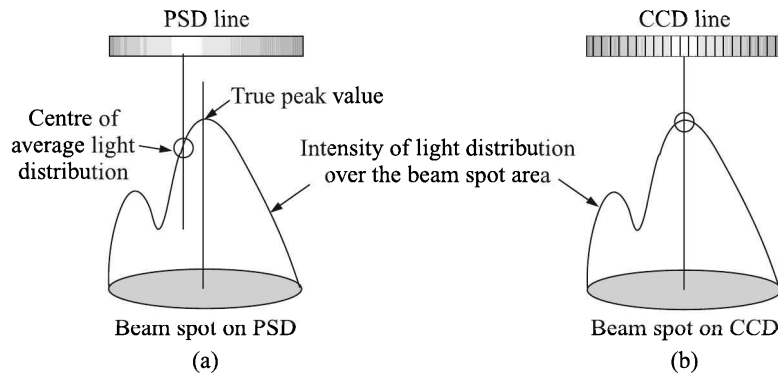


Fig. 6.31 Light intensity distribution of beam spot on light-receiving elements: (a) PSD, and (b) CCD.

CCD. CCD⁴ detectors are essentially a form of the image producing screen of the television camera and come in one- or two-dimension forms. In most simple triangulation sensors, a single-dimension CCD is used.

⁴See, for example *Solid State Electronic Devices*, B G Streetman, Prentice-Hall of India, New Delhi (1993), pp 357–62.

The CCD element is a digital pixelised array detector. It has 1024 discrete voltages representing the amount of light on each pixel of the detector. A CCD element detector can carry 1024×1024 pieces of light intensity information. With the help of a powerful DSP, the intensity distribution of the imaged spot [Fig. 6.31(b)] can be completely “viewed”. As a result, it can always detect the true peak position of the reflected light distribution.

Smart displacement sensors with CCD technology have revolutionised the scope of applications for triangulation sensors. The CCD sensors overcome almost all of the application problems that PSD sensors are unable to solve. But they are costlier.

Surface effects. The reflected light from a target is a mixture of diffused reflection and regular reflection. The proportion of diffused light and regular light depends on the material and surface conditions of the target. A regular reflection is dominant for a specular body while a diffused reflection predominates for objects with a normal surface. Since specular or nearly specular bodies are usually measured with a high accuracy laser displacement meter, the transducer structure is suitably modified such that it accepts regular reflections for this type of displacement meter (Fig. 6.32).

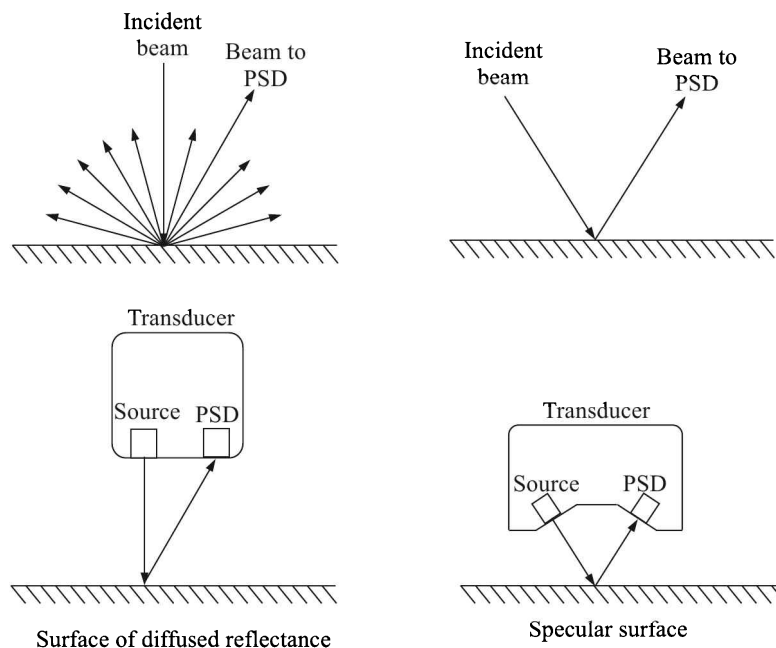


Fig. 6.32 Transducer structures for diffused and regular reflection displacement meters.

On the other hand, for the displacement meters that measure normal objects, the transducer detects a diffused reflection component in order to expand the measuring distance.

Triangulation devices that are ideal for measuring distances of a few inches with high accuracy, may be built on any scale, but the accuracy falls off rapidly with increasing range. The depth of field (minimum to maximum measurable distance) is typically limited.

Time-of-flight method

The time-of-flight (TOF) sensors derive range from the time light takes to travel from the sensor to the target and return. For very long range distance measurements (up to many kms) time-of-flight laser range finders using pulsed laser beams are used.

Phase measurement. A variant of the TOF method is to use modulated beams. These systems also use the time light takes to travel to the target and back, but the time for a single round-trip is not measured directly. Instead, the strength of the laser beam is rapidly varied to produce a signal that changes over time. The time delay is indirectly measured by comparing the signal from the laser with the delayed signal returning from the target. One common example of this approach is the *phase measurement* in which the output of the laser is typically sinusoidal and the phase of the outgoing signal is compared with that of the reflected light.

Phase measurement is limited in accuracy by the frequency of modulation and the ability to resolve the phase difference between the signals.

Range-to-frequency conversion. Some modulated beam range finders work on a range-to-frequency conversion principle, which offers several advantages over phase measurement. In these cases, laser light reflected from a target is collected by a lens and focussed onto a photodiode inside the instrument (Fig. 6.33).

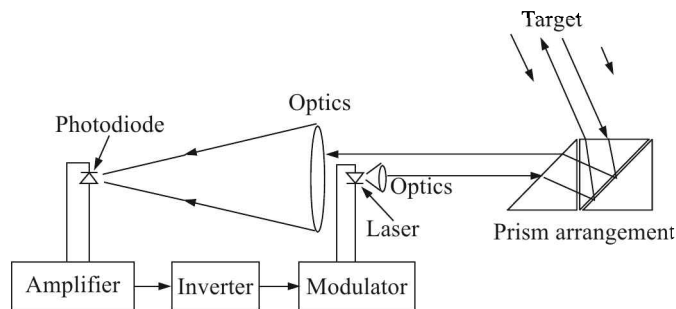


Fig. 6.33 Modulated beam TOF laser displacement transducer.

The resulting signal is amplified up to a limited level, inverted, and used directly to modulate a laser diode. The light from the laser is collimated and emitted from the centre of the front face of the sensor. This configuration forms an oscillator, with the laser switching itself ON and OFF using its own signal. The time that the light takes to travel to the target and return plus the time needed to amplify the signal determines the period of oscillation, or the rate at which the laser is switched ON and OFF. This signal is then divided and timed by an internal clock to obtain a range measurement. The measurement is somewhat nonlinear and dependent on signal strength and temperature. So, the sensor can be calibrated to remove these effects.

Modulated beam sensors are typically used in intermediate range applications, for distances from a few centimetres to several tens of metres on targets of diffuse reflectance. With specular targets, the range can be extended to several hundreds or thousands of metres.

Laser Confocal Microscope

This is a new non-contact measurement method which combines the confocal principle and the tuning fork, unlike the triangulation method which is commonly used for laser displacement meters. A confocal imaging system can reject out-of-focus objects by two strategies:

1. By illuminating a single point of the specimen at any one time with a focussed beam, so that illumination intensity drops off rapidly above and below the plane of focus, and
2. By the use of a blocking pinhole aperture in a conjugate focal plane to the specimen so that light emitted away from the point in the specimen being illuminated is blocked from reaching the detector. The principle is shown through diagrams in Fig. 6.34.

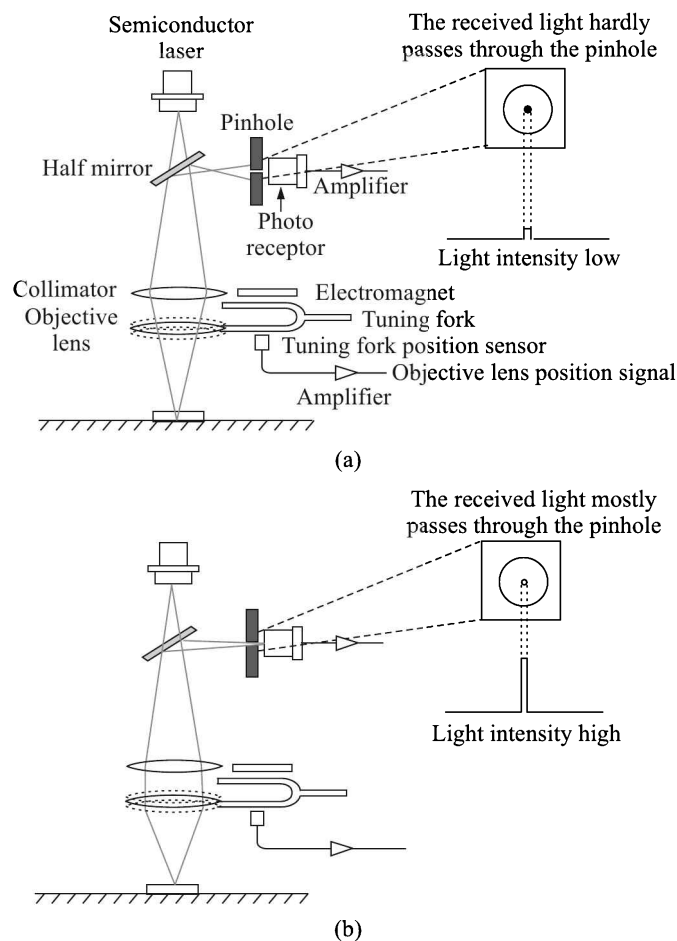


Fig. 6.34 The principle of distance measurement using the confocal microscope: (a) when focus is not on the object, and (b) when focus is on the object.

The laser beam passes through the objective lens, which moves rapidly up and down, based on the movement of the tuning fork, and focuses on the object. The reflected light from the target object passes through a half mirror and a pin hole, and reaches the photoreceptor.

According to the confocal principle, when the laser beam focuses on the object, the reflected light is then concentrated at the pinhole, through which it enters the photoreceptor. By measuring the position of the tuning fork at that time with the sensor, the distance to the object can be accurately measured.

The advantages of laser confocal displacement sensors are

1. Measuring the distance at the focal position means it is unaffected by any changes in the surface reflectance ratio.
2. The fact that the transmitted and received beams are located on the same axis eliminates any possible errors relating to the gradient or entry direction of the object.
3. If necessary, it can be used to measure thickness of thin films of transparent objects because the point where the refractive index changes can be identified.

The scanning confocal microscopy was invented by Mervin Minsky⁵ in 1958. But, it was not until 1987 when intense light sources such as lasers and image reconstruction techniques with the help of computers were available that attention of other workers was drawn to this novel method. Commercial instruments, mainly for studying structures of biological specimens using raster scan, were available in 1990s. Now, these sensors for measuring small displacements are commercially available.

Micro Laser Interferometer

Interferometers can measure very small displacements with a high degree of accuracy. More than a century ago, AA Michelson developed an interferometer that could be used to measure lengths or changes in length by applying the principle of interference of coherent monochromatic radiation. His arrangement is shown in Fig. 6.35.

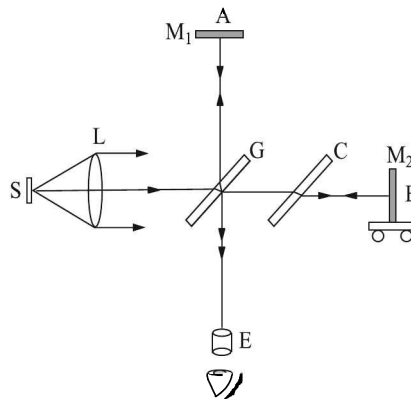


Fig. 6.35 Michelson's interferometer.

In Fig. 6.35, G is a half-silvered glass plate. A parallel beam of light, produced by an extended source (S) and lens (L) combination, is split into two components by G which is inclined at 45° to the beam direction. One component travels towards the plane mirror M_1 , gets reflected there and then travels through G towards the eyepiece E. The other component

⁵The interested reader may see his memoir at web.media.mit.edu/minsky/papers/ConfocalMemoir.html

of almost equal intensity travels towards the second plane mirror M_2 , gets reflected first at M_2 and then at G to proceed towards E . To make the optical paths of both the beams nearly equal, a compensator (C) plate of glass, of the same material and thickness of G , is placed in the path of the axial beam. The setup produces interference fringes that can be observed through the eyepiece. The interferometer was used to measure changes in length by counting the number of fringes that passed the field of view as the mirror M_2 was moved.

Although this looks like a simple arrangement, in practice this kind of measurement could be carried out in laboratories rather than in industrial environments until laser-based instrumentation came into being.

Now, a monochromatic and coherent light emitted from a small semiconductor laser is split into two beams by a tiny beam splitter, one beam being incident on a tiny reference mirror located in the sensor head and the other on the object to be measured. The sensor measures the interference pattern produced by the two differing optical paths. The replacement of a bright fringe by a dark fringe at the photoreceptor is equivalent to movement of the target by $\lambda/2$ where λ is the wavelength of the incident radiation. The displacement is calculated based on conversion of the interference with a special digital signal processing (DSP) board, making a high-speed, high-resolution measurement possible.

The whole interferometer along with the DSP board package measures a mere $4\text{ cm} \times 5\text{ cm} \times 2\text{ cm}$ and weighs about 50 g. This dramatic reduction in size enables the interferometer to be used in

1. Piezoelectric measurement equipment
2. Wafer-stage position control for electron beam drawing systems
3. Surface measurement of silicon wafers.

Commercially available microlaser interferometer based on the Michelson interferometer method can achieve a resolution of 0.08 nm.

While oscillations of ultra-low-frequency is naturally measurable with an interferometer, measurement of static displacement is also possible. The spot diameter of the beam incident on the object to be measured is an extremely small $5\text{ }\mu\text{m} \times 2.5\text{ }\mu\text{m}$. This allows measurement even if the surface being measured does not appear reflective, as long as there is sufficient flatness in the tiny area irradiated with the laser spot.

Fibre-optic Transducer

The fibre-optic displacement transducer⁶ contains two sets of optical fibres. One set, connected to a light source, is termed the transmitting fibres, and the other set, connected to a photodetector (photodiode), is known as the receiving fibres. These two sets of fibres are bundled into a common probe (Fig. 6.36).

The light from the source, channelled through the transmitting fibres to the probe tip, travels to the target surface and part of it is reflected back to the probe. A portion of the reflected light is caught by the receiving fibres and transmitted to the photodetector where its intensity is measured. The intensity of the reflected light is a function of distance (gap) between the probe tip and the target surface.

⁶aka the *fotonic* sensor after the name of a popular optical fibre.

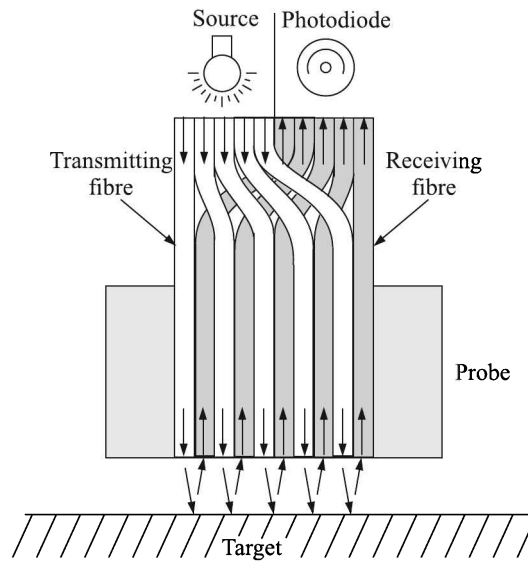


Fig. 6.36 Fibre-optic displacement transducer.

An optical fibre is a flexible strand of glass or plastic capable of transmitting light along its length by maintaining near total internal reflection of the light accepted at its input end. The most commonly used fibres are called 'step index' type. They consist of an inner core to carry the light flux and an outer cladding (Fig. 6.37).

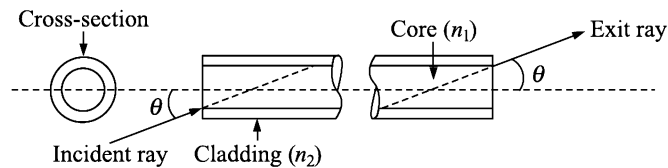


Fig. 6.37 Schematics of an optical fibre.

Numerical aperture

For the total internal reflection to occur, the refractive index of the transparent material in the core n_1 must be greater than that of the cladding n_2 . The numerical aperture (N_A) is defined as the sine of the half angle θ of the light which will be accepted into the core.

Multimode⁷ optical fibre will only propagate light that enters the fibre within a certain cone, known as the 'acceptance cone' of the fibre. The half-angle of this cone is called the 'angle of acceptance' θ_{\max} . For a step-index multimode fibre, the acceptance angle is determined only by the indices of refraction given by the relation

⁷Multimode fibre has a larger core-size than single-mode fibre. It supports more than one propagation mode, hence it is limited by modal dispersion, while single mode is not. Also, because of their larger core size, multimode fibres have higher numerical apertures which means they are better at collecting light than single-mode fibres. They are mostly used for communication over shorter distances, such as within a building or a campus.

$$n \sin \theta_{\max} = \sqrt{n_1^2 - n_2^2}$$

For a light ray incident from a medium of refractive index n to the core of index n_1 , we get from Snell's law

$$n \sin \theta_i = n_1 \sin \theta_r$$

which gives

$$\sin \theta_r = \frac{n}{n_1} \sin \theta_i \quad (6.29)$$

From Fig. 6.38, we get

$$\sin \theta_r = \sin(90^\circ - \theta_c) = \cos \theta_c \quad (6.30)$$

We know,

$$\sin \theta_c = \frac{n_2}{n_1} \quad (6.31)$$

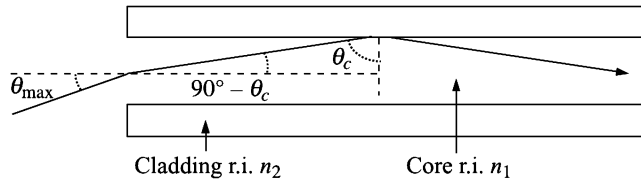


Fig. 6.38 Geometrical configurations of the incident and refracted rays in an optical fibre.

Substituting the value of θ_r from Eq. (6.29) in Eq. (6.30), we get

$$\frac{n}{n_1} \sin \theta_i = \cos \theta_c \quad (6.32)$$

By squaring both sides of Eq. (6.32) and utilising the result of Eq. (6.31), we get

$$\frac{n^2}{n_1^2} \sin^2 \theta_i = \cos^2 \theta_c = 1 - \sin^2 \theta_c = 1 - \frac{n_2^2}{n_1^2} \quad (6.33)$$

On simplifying Eq. (6.33), we arrive at the result

$$n \sin \theta = \sqrt{n_1^2 - n_2^2}$$

This form, which resembles the numerical aperture used for other optical systems, is defined as the N_A of any type of fibre so that

$$N_A = \sqrt{n_1^2 - n_2^2} = n \sin \theta_{\max} \quad (6.34)$$

The angle of acceptance is the maximum angle at which a light ray incident on the fibre can be trapped within the core and reflected along its length. The exiting light rays at the other end of the fibre are also limited to the same angle.

The diameters of a single fibre usually fall in the range of about 0.025 to 0.25 mm. Recent advances in the manufacturing technology however have extended the size up to about 1.5 mm. The efficiency of transmission depends on the composition and purity of the glass used in the core and cladding as well as on the quality of the optical finish of the end surfaces of the fibre.

Figure 6.39 shows how the transmitting fibre illuminates the target and an adjacent receiving fibre collects the reflected light. It is easy to see that when the distance between

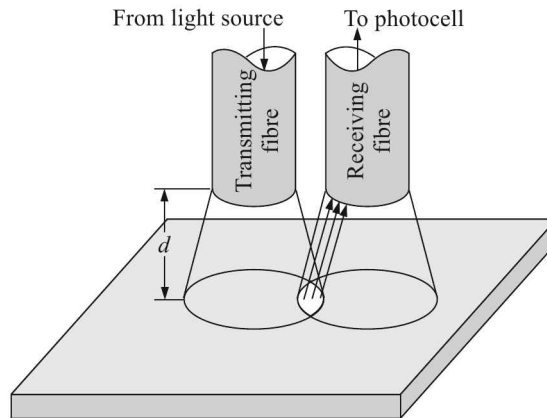


Fig. 6.39 Transmission and reception of light by adjacent fibres.

the probe and the target d is zero, the light in the transmitting fibre would be reflected directly back into itself and little light would be collected by the receiving fibre. As d increases, the numerical apertures of two fibres overlap over an area and so, some of the reflected light is captured by the receiving fibre and carried to the photodetector. In other words, the receiving fibres become more and more illuminated. As a result, the measured intensity of the reflected light increases almost linearly with gap distance in this *front slope region* (Fig. 6.40).

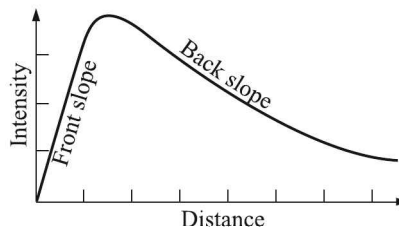


Fig. 6.40 Probe distance vs. photoreceptor characteristic.

The measured intensity keeps increasing until the gap distance is about the same order as the probe diameter. In this transition region, the receiving fibres are now fully illuminated and the maximum measured reflection is reached. With further increases in d past the transition region, the measured intensity drops off following roughly an inverse-square law. This results from the fact that, even though the receiving fibres are fully illuminated, the actual intensity of the reflected light as seen by the probe diminishes. Consequently, the intensity of light monitored at the photoreceptor drops.

The displacement range over which the initial rise in signal takes place and at which the maximum occurs is primarily determined by the diameter and the N_A of the fibres and the intensity distribution within the operating field of the fibres. Most commercial devices of this type use multiple transmitting and receiving fibres in order to obtain the higher levels of intensity at the photodetectors. This is needed to ensure acceptable levels of performance. The transmitting and receiving fibres are arranged in hemispherical, concentric or random patterns as shown in Fig. 6.41.

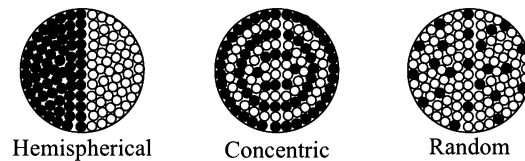


Fig. 6.41 Different patterns of arrangement of transmitting and receiving fibres in the probe—○ transmitting fibre, ● receiving fibre.

Going back to Fig. 6.40, the front slope region is more or less linear and therefore, the most suitable for displacement measurement. The back slope region is also useful, but is less linear and sensitive. The gap at which the maximum, or zero slope occurs, provides a convenient and readily usable calibration reference position at which the output signal can be normalised in order to obtain a consistent sensitivity factor relatively independent of the colour or finish of the surface of the target.

An interesting and useful variation on this device is obtained when a focussing lens system is placed near the sensing end of the fibre-optic probe. The results of one such combination of lenses and fibre-optics is shown in Fig. 6.42.

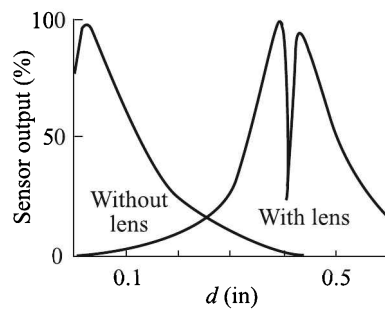


Fig. 6.42 Sensor output vs. displacement curves for system with lens and without lens near the sensing end of the fibre-optic probe.

As can be seen, two optical peaks are now present with a sharp null occurring midway between the peaks. Either of the steep response areas on each side of the null point can be used to measure small displacements at high resolutions with the help of self-focussing continuous tracking optical measurement systems.

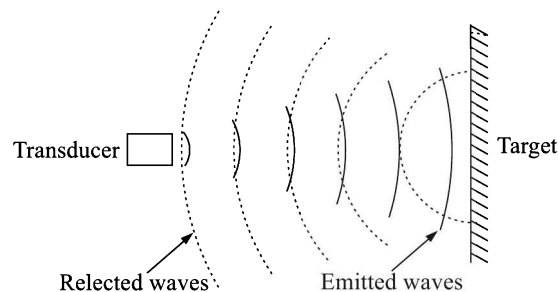
The advantages and disadvantages of fibre-optic displacement transducers are listed in Table 6.8.

Table 6.8 Advantages and disadvantages of fibre-optic displacement transducers

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|--|
| 1. Can operate directly with a large variety of surfaces, from specular to diffuse. | 1. Works well on highly reflective surfaces less effective on duller surfaces. |
| 2. Works well with materials from conductors to insulators. | 2. Re-calibration is generally required often, since the reflectance of target surfaces may vary. |
| 3. Non-contacting measurement. | 3. Can measure the roughness of the target surface only up to the order of the spacing of the transmitting and receiving fibres. |
| 4. Contains neither moving parts nor electrical circuitry. Therefore, completely immune to all forms of electrical interference. | 4. A misalignment, or dust accumulation on the cable tip degrades sensor performance. |
| 5. No possibility of a spark. Therefore, safe even in the most hazardous environments. Also, no danger of electrical shock to personnel repairing broken fibres. | |

6.4 Ultrasonic Transducer

The ultrasonic transducer measures displacement by applying the TOF principle. Typically, the transmitter of the transducer sends a 'ping' and its receiver waits to hear an echo. Sound waves propagate from the transmitter and bounce off the target, returning an echo to the receiver (Fig. 6.43). If the velocity of sound propagation is known, the distance to an object can be calculated from the time delay between the emitted and reflected sounds.

**Fig. 6.43** Ultrasonic displacement measurement.

This technique is frequently referred to as *echo ranging*. A block diagram given in Fig. 6.44 illustrates the basic design concept and functional elements in a typical ranging system.

The oscillator output is gated to the ultrasonic transmitter for a brief period that will result in the transmission of a few cycles of ultrasonic energy. The gate signal also starts a counter which is stopped by the detected returning echo. The count is thus directly proportional to the propagation time of the ultrasonic sound. The frequency of the clock that is driving the counter is selected to produce a count which represents the distance to the object in the desired engineering units.

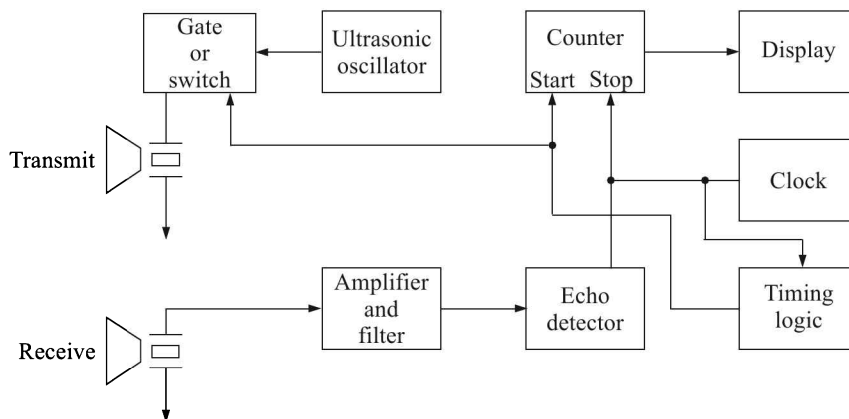


Fig. 6.44 Block diagram of a typical ultrasonic displacement measurement system.

The returning ultrasonic echo is usually very weak and the key to designing a good ranging system is to utilise a high Q -tuned frequency amplifier stage that will significantly amplify any signal at the frequency of the ultrasonic echo while rejecting all other higher or lower frequencies.

Another useful technique is to make the gain of the echo amplifier increase with time such that the amplifier gain compensates for the proportional decrease in the signal strength with distance or time. The most common approach is to utilise the counter state outputs to drive a digital programmable amplifier such that the gain is automatically related to the distance the sound travels.

Factors to consider

While the ultrasonic method of finding the distance of an object is simple, there are a few limiting factors to consider.

Sound generation. The sound generator features a piezo ceramic disc that resonates at a nominal frequency of 20 to 60 kHz and radiates or receives ultrasonic energy.

The 'open' type transducer exposes the piezo element bonded with a metal conical cone behind a protective screen. The 'enclosed' type has the piezo element mounted directly on the underside of the top of a casing which is then machined to resonate at the desired frequency. The enclosed type is necessary for dusty or outdoor applications. The face of the transducer must be kept clean and free of damage to prevent losses.

The transmitter should have a low impedance at the resonant frequency to obtain high mechanical efficiency. The receiver should have maximum impedance at the specified anti-resonant frequency to provide high electrical efficiency. The open type receiver will develop more electrical output at a given sound pressure level (high sensitivity) and exhibit less reduction in output as the operating frequency deviates from normal resonant frequency (greater bandwidth).

Surface-to-beam angle. Sound diverges very rapidly, so transducers are to be carefully designed to produce as small a beam angle as possible. Though a wide angle beam is suitable for some applications, a narrow beam improves the range and reduces background interference.

Since sound propagates as a wave, its velocity of propagation and directivity are related to its wavelength λ . A typical radiation power pattern for either a generator or receiver of waves is shown in Fig. 6.45. Owing to the reciprocity of transmission and reception, the graphical pattern portrays both radiated power and the reception sensitivity along a given direction.

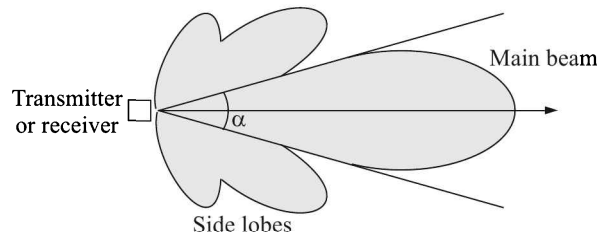


Fig. 6.45 Intensity distribution pattern of ultrasonic beam.

The angular half-width ($\alpha/2$) of the main beam is given by

$$\sin \frac{\alpha}{2} \approx \frac{\lambda}{d} = \frac{v}{df} \quad (6.35)$$

where d is the effective diameter of the piezo ceramic disc

v is the velocity of sound propagation ($344 \text{ m}\cdot\text{s}^{-1}$ in air at 20°C)

f is the operating frequency.

Equation (6.35) holds good if $\lambda < d$. For $\lambda > d$, the power distribution pattern tends to become spherical in form. Thus, narrow beams and high directivity are achieved by selecting $d \gg \lambda$.

For open type transducers, the beam divergence is decided by the shape of the conical guide attached to the piezo ceramic disc inside the housing. Therefore, it cannot be simply calculated by the diameter of the disc.

Even if the generated beam angle is small, but the object surface is severely tilted (generally $> 12^\circ$) away from the perpendicular of the beam axis, the echo is deflected away from the sensor, preventing the object from being detected (Fig. 6.46).

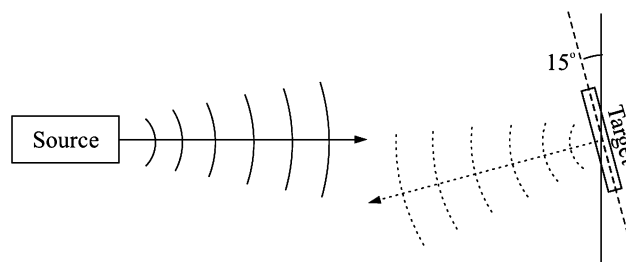


Fig. 6.46 Considerable tilting away from the perpendicular prevents the target from being detected.

Example 6.12

Calculate the diameter of the piezo ceramic disc that generates ultrasonic waves of frequency 60 kHz and propagates it with full width of 20° .

Solution

Here, $\alpha = 20^\circ$, $f = 60$ kHz. Therefore, from Eq. (6.35)

$$d = \frac{v}{f \sin(\alpha/2)} = \frac{344}{(60 \times 10^3)(\sin 10^\circ)} = 0.033 \text{ m} = 3.3 \text{ cm}$$

Object surface area vs. distance. The amount of ultrasonic energy reflected back to the sensor from the object depends largely on the object surface area and its distance from the sensor. The intensity of sound waves decrease with the distance from the sound source due to two effects:

- (a) *The spherical divergence which gives rise to the inverse square law:* This implies that an intensity drop of 6 dB occurs if the distance is doubled. This is common to all wave phenomena regardless of their frequencies.
- (b) *The absorption of the wave by the air:* It varies with humidity and dust content of the air and most importantly, with the frequency of the wave. Absorption at 20 kHz is about 0.02 dB per 30 cm. However, though a lower frequency is better suited for a long range propagation, it will result in less directivity for a given diameter of source or receiver.

If an object is positioned, for example, 200 mm from the sensor, the received echo is approximately 4 times stronger than if the object is at 400 mm. Thus, it is possible that the echo strength from an object with a small surface area, placed at the maximum sensing distance, may be too weak to detect.

Surface reflection properties. Almost all materials reflect ultrasonic energy and can be detected. However, a flat, hard, smooth surface, located perpendicular to the transmitted sonic beam, is the ideal condition for reliable detection. Conversely, materials with coarse and/or textured surfaces diffuse or absorb much of the transmitted energy. For example, granular products, foam rubber materials and certain textiles and papers may need a stronger transmitted signal for reliable detection.

Temperature. The velocity of propagation of sound varies with temperature. As air gets warmer, sound travels faster. Hence, ultrasonic systems must incorporate a thermometer to estimate the current speed of sound. While the ambient air temperature can be measured, other disturbing effects, such as convection and turbulence, can cause errors in the calculated distance.

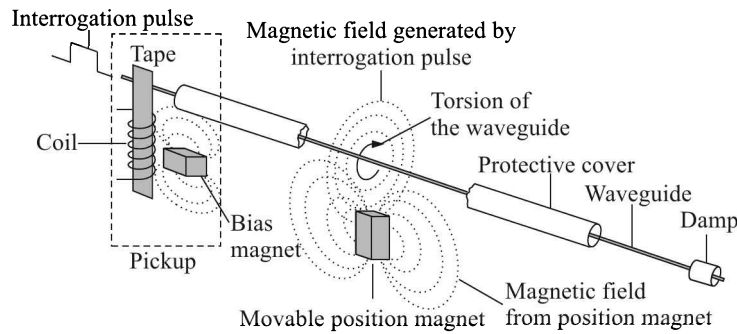
The advantages and disadvantages of ultrasonic method of measurement of displacement are given in Table 6.9.

Table 6.9 Advantages and disadvantages of echo ranging

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|---|
| 1. Non-contact measurement. | 1. Moderate accuracy: 0.1 to 2% of the range. |
| 2. Works with almost any surface type. | 2. Requires near-perpendicular incidence on the target. |
| 3. Resistant to vibration, radiation, background light, and noise. | 3. Affected by dust, dirt, high humidity or air turbulence. |
| 4. Low cost. | 4. Limited speed of response. |

6.5 Magnetostrictive Transducer

Suppose, a current pulse, called *interrogation pulse*, is sent through a waveguide which has a magnet in its path at a place (Fig. 6.47). Then, a torsional force is induced in the waveguide at the location of the magnet⁸. The force owes its origin to the interaction between the magnetic field produced by the current pulse and the permanent magnet. The torsional pulse generates a strain wave that travels at the acoustic speed (~ 2850 m/s) in the waveguide changing its magnetic permeability instantaneously⁹ at the points through which the wave passes.

**Fig. 6.47** Magnetostrictive displacement transducer action.

Initially, an interrogation pulse is sent through the ferromagnetic waveguide (normally a wire or tube) and simultaneously a timer circuit is switched on from one end of the waveguide which we call *source*. The pulse is of very short duration, about 1 to 2 μs . Its minimum current density is along the centre of the wire and the maximum at the wire surface. This is due to the skin effect. As the pulse reaches the position of the target (a permanent magnet), it generates the torsional sonic wave that travels along the waveguide in both the directions. A pick-up at the source detects a signal as the wave reaches there because at that moment the permeability of the waveguide material changes instantaneously. This signal stops the timer. The elapsed time, Δt , multiplied by the speed of acoustic wave in the waveguide material gives the location of the target.

⁸Wiedemann effect, see Section 5.2 at page 119.

⁹Villari effect.

The sonic wave also travels in the direction away from the pickup. In order to avoid an interfering signal from waves travelling in this direction, its energy is absorbed by a damping device (called the *damp*).

The pickup consists of a small piece of magnetostrictive material, called the *tape*. The tape is welded to the waveguide near one end of the waveguide. It passes through a coil and is magnetised by a small permanent magnet called the *bias magnet*. When the sonic wave propagates down the waveguide and then down the tape, the stress induced by the wave causes a wave of changed permeability¹⁰ in the tape. This, in turn, causes a change in the magnetic flux density in the tape, and thus a voltage output pulse is produced from the coil.¹¹ The voltage pulse is detected by the electronic circuitry and conditioned into the desired output.

Magnetostrictive absolute linear displacement sensors in the range of 25 mm to 25 cm having an accuracy of 0.001% are commercially available. They are intrinsically safe and therefore, can be used in hazardous areas. Magnetostrictive level gauges¹² of much higher ranges are also available.

6.6 Digital Displacement Transducers

So far we have talked about transducers which generate analogue signals proportional to displacements. Of course, these signals can be converted to digital ones by analogue-to-digital converters (ADCs), and suitable digital counters can be used to produce digital readouts.

But a transducer that presents information as discrete samples and that does not introduce a quantisation error¹³ when the reading is represented in the digital form may be classified as a digital transducer. These transducers are called *encoders* because they generate coded messages of a measurement. The encoders employ binary coded strips or discs depending on whether linear or angular displacement is being measured. The binary values of 0's and 1's may be generated by

1. Either using electrically conducting and non-conducting segments with sliding brush contacts and applying a voltage at conducting ones
2. Or, magnetically storing values which can be sensed by a read-head much similar to the one utilised in the floppy-disc drive of a computer
3. Or, optically with opaque and transparent windows through which a tiny light beam can be sent to be sensed by a light sensor.

Encoders can be divided into three categories:

1. Tachometer type
2. Incremental type
3. Absolute type.

¹⁰Villari effect.

¹¹Faraday effect.

¹²See Section 12.1 at page 491.

¹³See Section 16.4 at page 800.

Tachometer Type

The coding in such a transducer is schematically shown in Fig. 6.48. Because of this kind of coding, it has a single output signal which consists of a pulse for each increment of the displacement. These pulses can be counted by a digital counter which, in turn, can be calibrated in terms of displacement in suitable units.

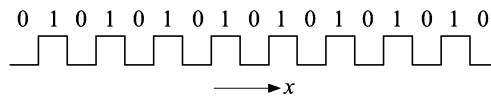


Fig. 6.48 Coding of a tachometer encoder.

If motion is always in one direction, these encoders are very simple aids to measure the corresponding displacements. However, any reversed motion will produce identical pulses causing errors.

Tachometer encoders, as their name suggests, are suitable for measuring speed rather than displacement because the quicker the arrival of pulses, the faster is the motion.

Incremental Type

In this type at least two, sometimes three, tracks of coding are employed to solve the reverse-motion problem (Fig. 6.49).

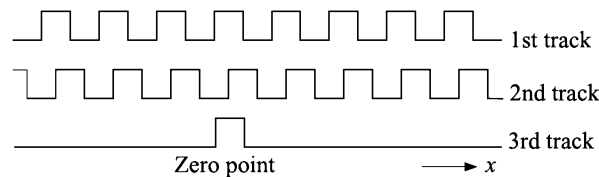


Fig. 6.49 Coding of an incremental encoder.

Track 1 and track 2 are coded in such a way that there is a phase-shift of one-fourth of a cycle relative to the other. So the direction of motion can be detected by noting the signal from which track rises first. When a third track is incorporated, it is used as a reference track which produces a single pulse per revolution at a distinct point, called the *zero point*.

Absolute Type

Such encoders employ four or more bit binary coded strips or discs, as shown in Fig. 6.50. Here, data from multiple tracks are read out in parallel to produce a binary representation of the displacement or angular position.

From the measurement point of view, binary code is rather inconvenient because in this code sometimes a number of bits change in two consecutive decimal positions. For example, the representations of 7_{10} and 8_{10} are 0111_2 and 1000_2 , where all the four bits change between two consecutive decimal numbers. If the readout device is slightly misaligned to the left, the count may go from 0111 (7_{10}) to 0011 (3_{10}), producing a substantial error. To take care of

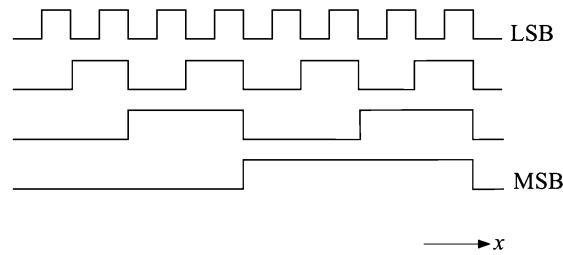


Fig. 6.50 Coding of an absolute encoder.

such situations, commercial encoders use Gray¹⁴ code where only one bit changes at each transition and as a result the error arising out of a small misalignment is kept to a minimum. The following comparison between the binary and Gray codes for decimal numbers 0 to 9, as shown in Table 6.10 will make the point clear.

Table 6.10 Comparison between binary and Gray codes

| <i>Decimal</i> | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|--------------------|------|------|------|------|------|------|------|------|------|------|
| <i>Binary code</i> | 0000 | 0001 | 0010 | 0011 | 0100 | 0101 | 0110 | 0111 | 1000 | 1001 |
| <i>Gray code</i> | 0000 | 0001 | 0011 | 0010 | 0110 | 0111 | 0101 | 0100 | 1100 | 1101 |

Note: To convert a standard binary number to its Gray equivalent, we first examine the rightmost digit and then consider each digit to the left in turn. If the digit to the left is 0 then let the original digit stand. If the next digit to the left is 1, then change the original digit. The digit at the extreme left is assumed to have a 0 to its left and therefore remains unchanged.

For example, if the standard binary number is 110111, we show below how to obtain its Gray representation

| | | | | | | |
|---|---|---|---|---|---|-------------|
| <i>Given binary (from right to left):</i> | 1 | 1 | 1 | 0 | 1 | 1 |
| <i>Conversion to Gray (from right to left):</i> | 0 | 0 | 1 | 1 | 0 | 1 |
| <i>Because the left digit is:</i> | 1 | 1 | 0 | 1 | 1 | 0 (assumed) |

Writing the number from left to right, the Gray representation is 101100. Sometimes, the Gray code is called the *reflected binary code*.

The binary is a positional code. That means, bits in different positions have their own weights. For example, a 1 bit in the fourth position from the right has a weight of $2^{4-1} = 8$. This is why it is often referred to as 8421 binary code. So, the Gray code is better converted to binary code for display or further instrumentation. This conversion can easily be done with the help of XOR gates as shown in Fig. 6.51.

¹⁴Named after Frank Gray (1887–1969), an American physicist and researcher at Bell Labs who patented it in 1953. However, the codes had been used in telegraphy by the French engineer Émile Baudot in 1878.

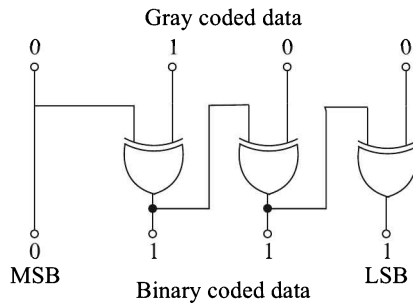


Fig. 6.51 Conversion of Gray code to 8421 binary code.

Binary and Gray coded 4-bit strips are shown in Figs. 6.52(a) and (b).

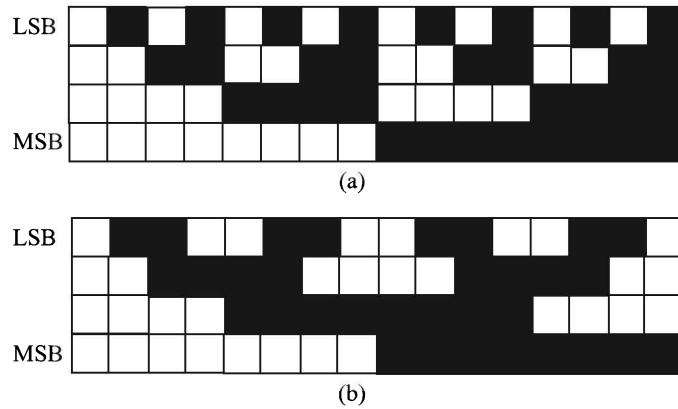


Fig. 6.52 (a) 8421 binary coded strip, and (b) Gray-coded strip.

As already pointed out, encoders of all three types can be constructed as non-contacting devices using magnetic or optical principles, or as contact devices with stationary brush contacts (Fig. 6.53). Contact devices have low resolution and they suffer from mechanical wear and tear. For the finest resolution, optical encoders are the most suitable ones although one may have to use a monochromatic light source to avoid diffraction effects.

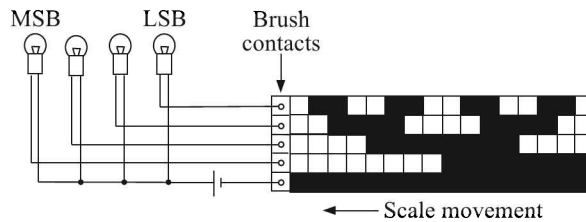


Fig. 6.53 Digital readout device through brush contact on a translational encoder based on Gray code.

Figure 6.54 shows a rotary encoder which is used for the measurement of angular displacement. The resolution of commercially available translational encoders is around $1 \mu\text{m}$ and that of angular ones is of the order of fraction of a second arc.

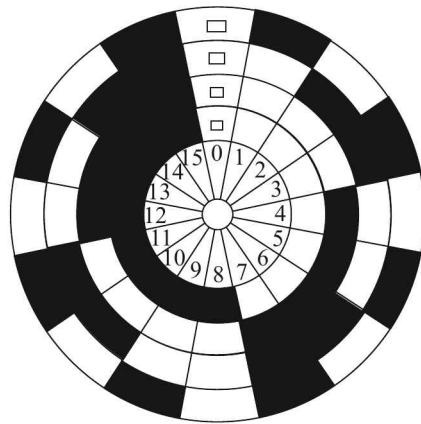


Fig. 6.54 Rotary encoder based on binary code.

A comparison of the accuracy and range of commercially available transducers is given in Fig. 6.55. The graph is more suggestive than exhaustive.

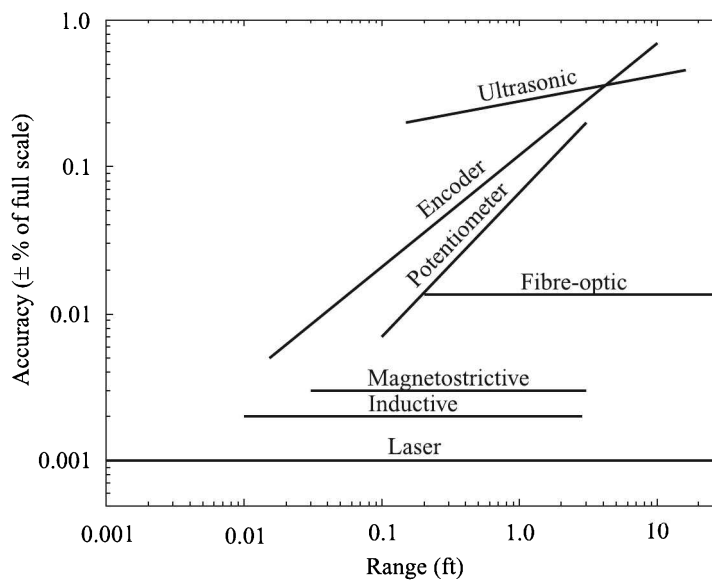


Fig. 6.55 Comparison of accuracy and range of commercially available displacement transducers.

There are, of course, other kinds of transducers for measurement of displacement. For example, piezoelectric transducers can measure very small displacements, and so on.

6.7 Proximity Sensors

The device which helps detect and measure distance of nearby objects without physical contact is called a *proximity sensor*.

The maximum distance that a proximity sensor can detect is called its *nominal range*. Some sensors have adjustments of the nominal range or means to report a graduated detection distance.

Apart from measuring roundness, straightness or surface roughness of flat objects, proximity sensors are also used in monitoring the machine vibration or to measure the variation in distance between a shaft and its support bearing. This is common in large steam turbines, compressors, and motors that use sleeve-type bearings. They are also widely used in general industrial automation such as conveyor lines (counting, jam detection, etc.), machine tools (safety interlock, sequencing).

Proximity sensors can have a high reliability and long functional life because of the absence of mechanical parts and lack of physical contact between the sensor and the target.

Proximity detectors are generally of the following types:

1. Inductive
2. Capacitive
3. Hall effect
4. Optical
5. Ultrasonic

Inductive Proximity Sensors

An inductive proximity sensor detects only metallic objects. It generates an *electromagnetic field* and measures the change in the field owing to the presence of the object. The detected object is referred to as the proximity sensor's *target*.

They are usually based on the following principles:

1. Variation of reluctance
2. Eddy-current generation

Variation of reluctance

The self-inductance L of a coil can be written as

$$L = \frac{N^2}{R} \quad \text{where,} \quad R = \frac{l}{\mu A} \quad (6.36)$$

Here, N is the number of turns in the coil

R is the reluctance of the magnetic circuit

l is the length of the magnetic path

A is the area of cross-section of the magnetic path

μ is the effective permeability of the medium in and around the coil.

So, R can be written as

$$R = \frac{1}{\mu G} \quad \text{where,} \quad G = \frac{A}{l} = \text{geometric form factor}$$

Thus, variation in either μ or G varies reluctance which, in turn, varies the inductance.

μ or G can vary due to displacement of various parts of the magnetic circuit. The resulting change in inductance can be measured to find the displacement.

Consider a magnetic circuit consisting of (i) a semicircular ring with an electrical winding, working as an electromagnet, (ii) an air gap, and (iii) a plate of magnetic material [Fig. 6.56(a)]. The magnetic circuit is shown by the dashed line.

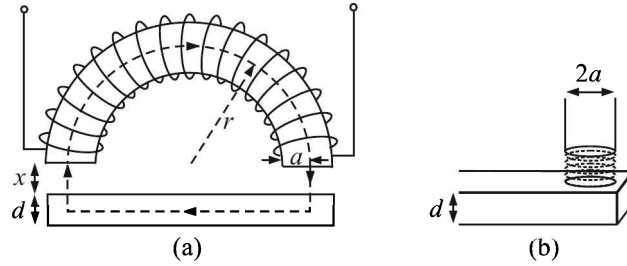


Fig. 6.56 (a) Magnetic circuit, and (b) path of the magnetic flux through the plate.

Let μ_m be the permeability of the magnet material

μ_p be the permeability of the material of the plate

μ_0 be the permeability of the air-gap between the magnet and the plate.

Since the magnetic circuit consists of three components, the total reluctance of the circuit is the sum of reluctances of individual components. That is,

$$R_{\text{total}} = R_{\text{magnet}} + R_{\text{gap}} + R_{\text{plate}}$$

Now,

$$R_{\text{magnet}} = \frac{l_{\text{magnet}}}{\mu_m A} = \frac{\pi r}{\mu_m (\pi a^2)} = \frac{r}{\mu_m a^2} \quad (6.37)$$

where r is the radius of curvature of the semicircular ring
 a is the radius of cross-section of the ring.

$$R_{\text{gap}} = \frac{2x}{\mu_0 (\pi a^2)} \quad (6.38)$$

because the magnetic flux area in the air gap is nearly the same as that of the cross-section of the semicircular ring and

$$R_{\text{plate}} = \frac{2r}{\mu_p (2ad)} = \frac{r}{\mu_p ad} \quad (6.39)$$

assuming that the magnetic flux originating from the magnet passes through an area of $(2a)d$ of the plate [see Fig. 6.56(b)].

Thus, from Eqs. (6.37) – (6.39)

$$\begin{aligned} R_{\text{total}} &= \frac{r}{\mu_m a^2} + \frac{2x}{\mu_0 \pi a^2} + \frac{r}{\mu_p ad} \\ &= \frac{r}{a} \left[\frac{1}{\mu_m a} + \frac{1}{\mu_p d} \right] + \frac{2x}{\mu_0 \pi a^2} \end{aligned}$$

or $R_{\text{total}} = R_0 + kx$

where $R_0 = \frac{r}{a} \left[\frac{1}{\mu_m a} + \frac{1}{\mu_p d} \right] = \text{reluctance for zero air-gap}$

$$k = \frac{2}{\mu_0 \pi a^2}$$

The corresponding change in self-inductance of the coil is, from Eq. (6.36)

$$L = \frac{N^2}{R} = \frac{N^2}{R_0 + kx} = \frac{L_0}{1 + \alpha x} \quad (6.40)$$

where $L_0 = \frac{N^2}{R_0} = \text{inductance from zero air-gap}$

$$\alpha = \frac{k}{R_0}$$

It may be seen from Eq. (6.40) that the self-inductance has a nonlinear relationship with displacement of the plate (i.e. air-gap) x . This can be converted to a linear relation by push-pull variable reluctance transducers in ac bridge as shown in Section 16.1 at page 746.

Their applications can be to detect open/close functions or to count the number of rotations. They have a small range, typically $3 \text{ mm} \pm 10\%$.

Eddy-current proximity sensor

Eddy-currents are generated in a conducting plate if it is placed near a coil carrying alternating current (Fig. 6.57). Eddy-currents flowing in the target produce a magnetic field of their own which opposes that produced by the coil. As a result the flux around the coil is reduced (i.e. μ is lowered) with a consequent reduction of the inductance of the coil. This effect depends on the distance between the target and the coil, because, the nearer the target to the coil, the higher the eddy currents and the lower the inductance and vice versa.

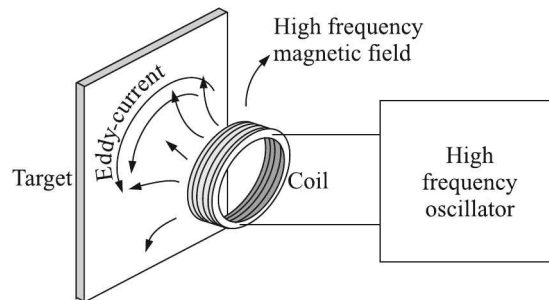


Fig. 6.57 Eddy-current generation.

Two arrangements are common:

1. Target placed perpendicularly to the axis of the coil [Fig. 6.58(a)]
2. Target in the form of a coaxial cylinder running over the coil [Fig. 6.58(b)]

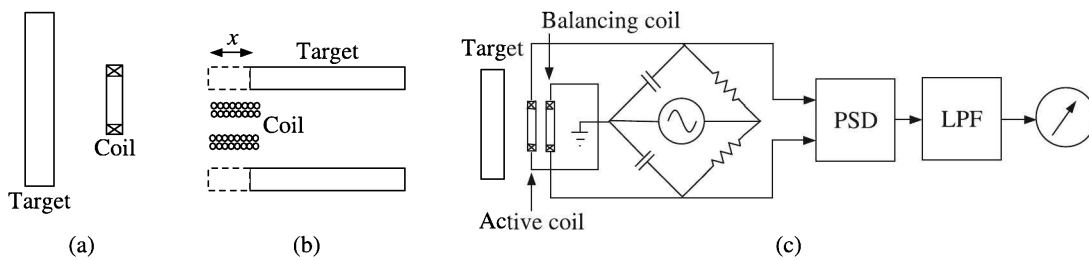


Fig. 6.58 Schematic diagram of placement of eddy-current generation type transducers: (a) target and coil at right angles, (b) coaxial target and coil, and (c) measuring arrangement.

Set-up. In a practical set-up [Fig. 6.58(c)], the probe consists of two coils—one (active) for the measurement and the other (balancing) for temperature compensation. The two coils are connected in parallel to two capacitors forming two arms of a bridge.

The distance between target and the probe determines the inductance of the active coil. So, when the bridge is balanced, any displacement of the target causes a bridge unbalance producing a voltage proportional to the target displacement. This unbalanced voltage is demodulated and low-pass filtered to produce a dc output which can be calibrated to indicate the extent of target displacement.

Alternative method. There is an alternative method of measurement of displacement with the help of an eddy-current transducer. It is through measurement of the phase difference and amplitude of the incident ac and that generated by the eddy current in the target. As the target comes closer to the sensor head, the oscillation amplitude becomes smaller and the phase difference from the reference waveform becomes larger (Fig. 6.59). By detecting changes in the amplitude and phase, the sensor can obtain a value approximately proportional to the change in the distance between the sensor head and the target.

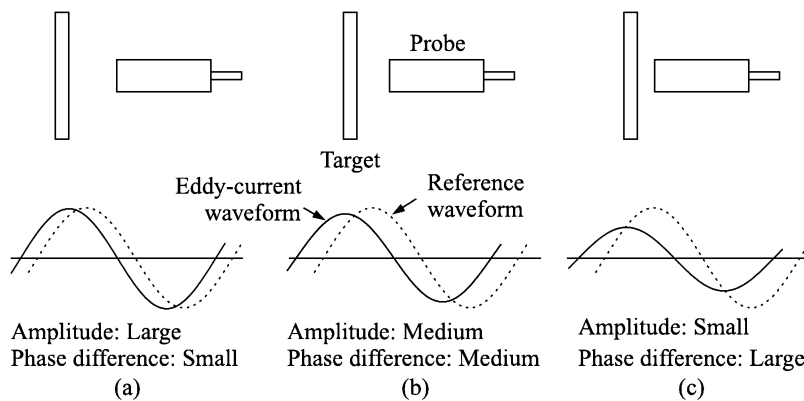


Fig. 6.59 Reference and eddy-current waveforms when the distance between the target and probe is (a) large, (b) medium, and (c) small.

Based on the target material, values are processed and corrected using a lineariser circuit. The linear output is proportional to the distance between the sensor head and the target.

Target. The eddy currents are confined to shallow depths, called *skin depth*, near the conductive target surface. The skin depth δ is given by

$$\delta = 50.3 \sqrt{\frac{\rho}{f\mu}} \text{ mm}$$

where ρ is the resistivity in $\mu\Omega\text{-cm}$
 f is the excitation frequency in Hz
 μ is the permeability of the target material

The target material must be at least three times thicker than the skin depth of the eddy currents to make the transducer successful. This is because a lower thickness will make the target hot and will change its thickness owing to thermal expansion.

The flat surface area of the target should not be smaller than the probe tip diameter. If the target surface is smaller than 50% of the probe diameter, output signals decrease substantially.

Offset. In practice, the measuring range of an eddy-current transducer is at about 20% offset of the range given by the vendor. For example, a 0 to 2.5 mm range eddy-current transducer is generally considered effective from 0.5 to 3 mm from the target surface.

Table 6.11 lists typical specifications of commercially available probes.

Table 6.11 Typical specifications of eddy-current probes

| | |
|-----------------------------|--|
| <i>Size</i> | About 2 to 75 mm in diameter, 20 to 40 mm long |
| <i>Range</i> | 0.25 to 30 mm |
| <i>Nonlinearity</i> | 0.5% |
| <i>Maximum resolution</i> | 0.1 μm |
| <i>Excitation frequency</i> | 50 kHz to 10 MHz |

Advantages and disadvantages. The advantages and disadvantages of an eddy-current transducer are given in Table 6.12.

Table 6.12 Advantages and disadvantages of eddy-current proximity sensors

| <i>Advantages</i> | <i>Disadvantages</i> |
|--------------------------------|---|
| 1. Non-contacting measurement. | 1. The displacement vs. the impedance characteristic is nonlinear and temperature dependent. Though a balance coil can compensate for the temperature effect, the nonlinearity has to be appropriately taken care of. |
| 2. High resolution. | 2. Works only on conductive materials with sufficient thickness. Cannot be used for detecting the displacement of nonconductive materials or thin metal foils. |
| 3. High frequency response. | |

Capacitive Proximity Sensor

The capacitive proximity sensor does the job by detecting the alteration in the *electrostatic field* set-up by the sensor. Therefore, the capacitive proximity sensor is capable of detecting any dielectric target like plastic or paper over and above metallic targets.

Parallel-plate capacitors are deployed for this purpose.¹⁵ The sizes of the sensor and the target, and the intervening material are assumed to be constant. Therefore, any change in capacitance is a result of a change in the distance between the probe and the target.

Working principle. The sensing surface of a capacitive sensor is formed by two concentrically shaped metal electrodes of an unwound capacitor (Figure 6.60).

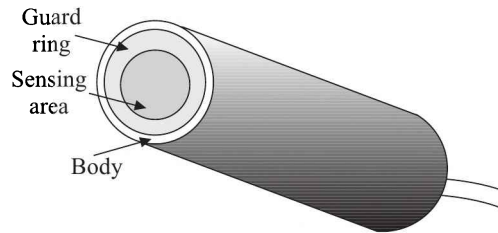


Fig. 6.60 Components of the capacitive proximity sensor probe.

When an object nears the sensing surface area, it enters the electrostatic field of the electrodes and changes the capacitance in an oscillator circuit where the capacitive sensor is attached. As a result, the oscillator begins oscillating. The trigger circuit reads the amplitude of oscillation and when it reaches a specific level the output state of the sensor changes. The block diagram of the electronic circuit and the schematic of the oscillation are shown in Figures 6.61(a) and (b) respectively.

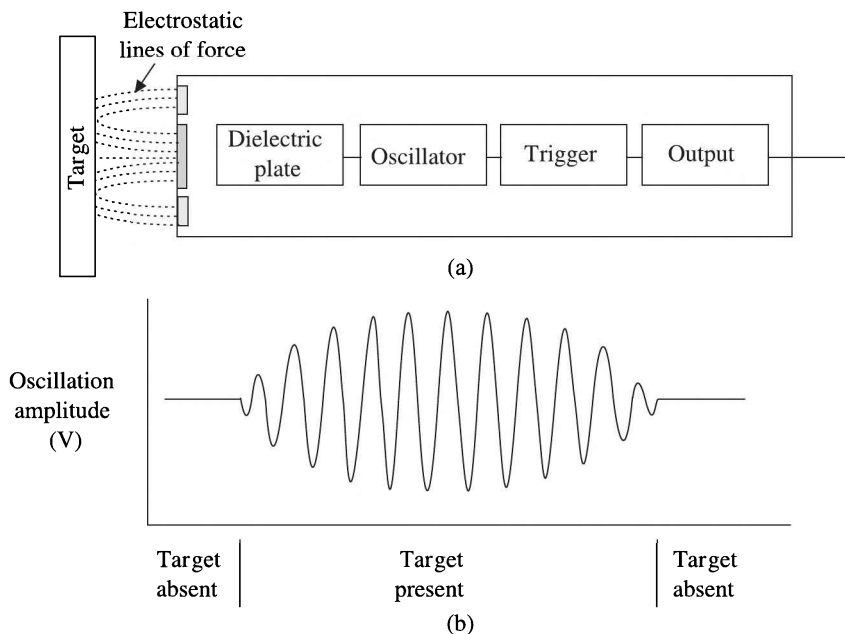


Fig. 6.61 (a) Block diagram of the electronic circuit, and (b) the oscillation generation.

¹⁵See *Change in the gap x between the plates* under Section 6.2 at page 187.

The electronic circuit is calibrated to generate specific voltage changes for corresponding changes in capacitance. These voltages are scaled to represent specific changes in distance. A common sensitivity setting is 1.0 V/100 μm . That means that for every 100 μm change in distance, the output voltage changes exactly 1.0 V. With this calibration, a +2.5 V change in the output means that the target has moved 250 μm closer to the probe.

Focussing the field. When a voltage is applied to a conductor, the electrostatic lines of force emanate from every surface. In a capacitive sensor, the sensing voltage is applied to the sensing area of the probe. For accurate measurements, the electrostatic field from the sensing area needs to be restricted within the space between the probe and the target. If the field is allowed to spread to other items or other areas on the target, then a change in the position of the other item will be measured as a change in the position of the target. This is known as *fringing*. A technique called *guarding* is used to prevent this from happening.¹⁶

To create a guard, the back and sides of the sensing area are surrounded by another conductor that is kept at the same voltage as the sensing area itself (Figure 6.62). When the

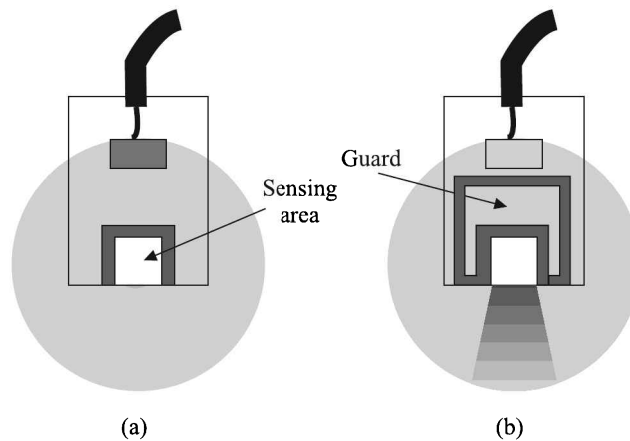


Fig. 6.62 Schematic diagrams showing electrostatic fields, through shading, of (a) unguarded, and (b) guarded capacitive proximity sensors.

voltage is applied to the sensing area, a separate circuit applies exactly the same voltage to the guard. Because there is no difference in voltage between the sensing area and the guard, there exists no field between them. Other conductors beside or behind the probe form a field with the guard instead of the sensing area. Only the unguarded front of the sensing area is allowed to form an electrostatic field with the target.

Target size. The target size is important when selecting a probe for a specific application. A slightly conical field that is a projection of the sensing area, is produced when the sensing electrostatic field is focussed by guarding. The minimum target diameter for standard calibration is

$$\text{Target diameter} = 1.3 \times (\text{Sensing area diameter})$$

¹⁶We have already discussed it in Section 6.2 under the heading *Effect of fringing flux* at page 190.

If the target size is too small, the electrostatic field will wrap around the sides of the target which means the field will extend further than it did while it was calibrated. As a result, it will measure the target further away.

Range. The range in which a probe is useful is a function of the size of the sensing area. The greater the area, the larger the range. Therefore, a smaller probe should be closer to the target. Although it is adjustable, but there is a limit to the range of adjustment. In general,

$$\text{Maximum gap} = 0.4 \times (\text{Sensing area diameter})$$

Accuracy. Capacitive proximity sensors are generally used for short ranges only. A wide range of accuracies is possible depending on the needs and economics of the application. Typical accuracies and cost at the present market rate for different applications are:

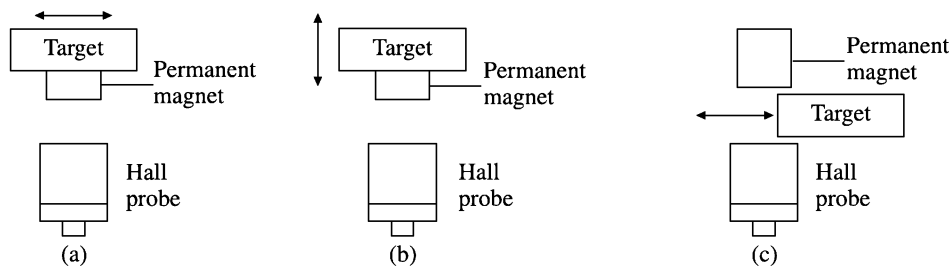
| <i>Application</i> | <i>Range (mm)</i> | <i>Accuracy</i> |
|--------------------|-------------------|--|
| Low cost | 10 | 0.1 % of FSD e.g. ± 0.01 mm per mm |
| Medium cost | 10 | 0.01% of FSD e.g. ± 0.001 mm per mm |
| High-accuracy | 100 | 0.001% of FSD e.g. ± 0.001 mm per mm |

They are regarded as robust, reliable and accurate sensors.

Hall Effect Proximity Sensor

A Hall effect¹⁷ proximity sensor is a non-contact electronic sensor, which consists of a permanent magnet or ferromagnetic part as trigger intermediary and a Hall effect sensor IC. The Hall sensor IC detects the change of the magnetic field when the permanent magnet comes in the close proximity to it and generates an electric signal. This signal is amplified and rectified to control the output signal of the proximity sensor.

The Hall proximity sensor can be used to measure horizontal, vertical and interception proximity modes as illustrated in Fig. 6.63.



South pole of the permanent magnet is oriented towards the sensing side of the Hall probe

Fig. 6.63 Hall proximity measurement: (a) horizontal, (b) vertical, and (c) interception modes. The double-edged arrow indicates the direction of movement of the target.

Figure 6.64 depicts the dimensions of a typical Hall proximity probe.

¹⁷See Section 5.2 at page 121 for details on Hall effect.

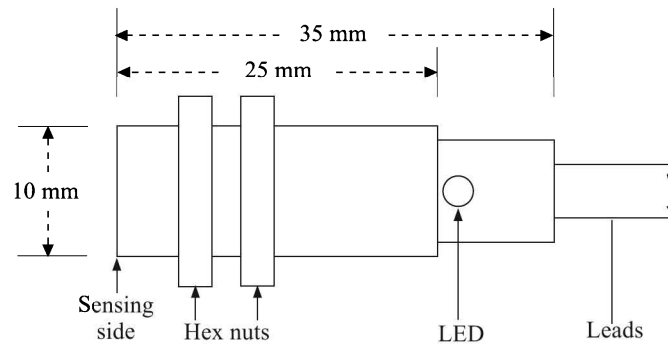


Fig. 6.64 Typical dimensions of a Hall proximity probe.

Advantages and disadvantages. The advantages and disadvantages of Hall proximity sensors are given in Table 6.13.

Table 6.13 Advantages and disadvantages of Hall proximity sensors

| <i>Advantages</i> | <i>Disadvantages</i> |
|---|---|
| 1. Good output wave shape. | 1. Maximum sensing distance is about 10 cm with strong magnets. |
| 2. High stability | 2. Affected by temperature variation because mobilities of carriers are temperature dependant. |
| 3. Low cost. | 3. Presence of an offset voltage. This occurs even with well centred electrodes. It can be as high as 100 mV for a 12 V source ^a . |
| 4. Unaffected by oil, dirt and vibration. | |
| 5. High suitability for integrating to PC systems and various kinds of industrial control equipment | |
| 6. Adaptability as optimal switches for position control, speed measurement, counting, direction detection and automatic protection | |

^a An offset voltage occurs when there are physical inaccuracies and material non-uniformities. To solve this problem an additional control electrode needs to be added and through it necessary current can be injected to obtain a null output when no magnetic field is present.

Optical Proximity Sensors

The laser transducers (see Section 6.3 at page 200) and fibre-optic transducers (see Section 6.3 at page 206) can be used as optical proximity sensors.

Optical proximity sensors can also be of the simple light beam type as shown in Fig. 6.65.

A simple optical proximity sensor includes a light source and a sensor that detects the light reflected from the target. The light source, usually an LED, generates light of such a wavelength that the sensor can detect best and that which is not likely to be present in the

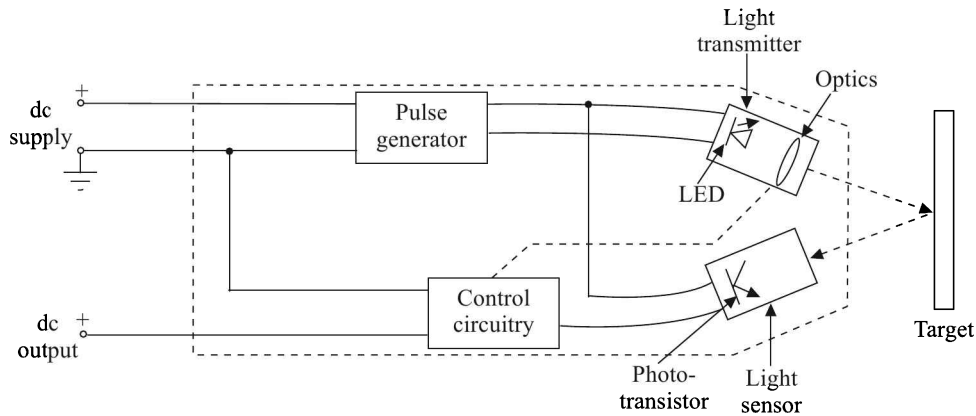


Fig. 6.65 Simple optical proximity sensor.

ambience. Usually, the infra-red light is used in most proximity sensors. Also, to make the sensing system foolproof, most optical proximity sensors use pulsed light sources.

The sensor can be a photodiode, which generates a small current when light energy strikes it, or more commonly a phototransistor or a photo-Darlington that allows current to flow if light strikes it.

The control circuitry may have to match the pulsing frequency of the transmitter with the light sensor. Control circuitry is also often used to switch the output circuit at a certain light level and if the distance measurement is necessary, to focus the optics of the transmitter on the target to track a maximum power point at the sensor. Light beam sensors that output voltage or current proportional to the received light level are also available.

Optical proximity sensors can be of three types, namely:

1. Through beam
2. Diffuse scan
3. Retro-reflective

Through beam. This type of sensors consist of separate transmitter and receiver. The target is sensed when the beam is interrupted.

Diffuse scan. This type of sensors have both transmitter and receiver in the same enclosure. The closely focussed beam is reflected back by the target.

Retro-reflective. This type of sensors use a special reflector to reflect the beam back. When the target interrupts the beam, an output signal is generated. The sensing distance of retro-reflective types is greater than the diffuse scan types. The receiver rejects reflections from all other sources or objects other than the special reflector.

The three types of optical proximity sensing are shown in Fig. 6.66.

Advantages and disadvantages. The advantages and disadvantages of the optical proximity sensor are given in Table 6.14.

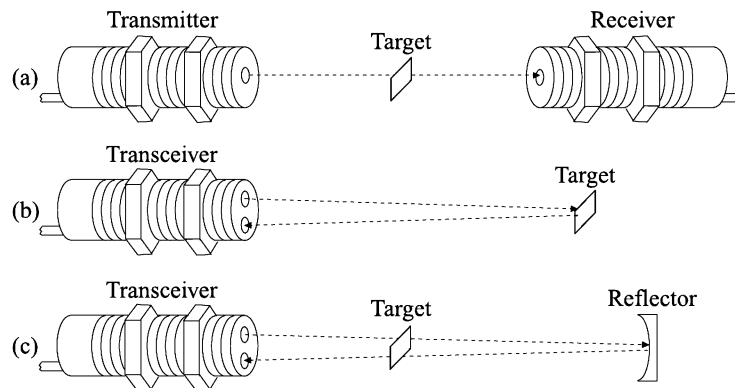


Fig. 6.66 Different types of optical proximity sensing: (a) through beam, (b) diffuse scan and (c) retro-reflection.

Table 6.14 Advantages and disadvantages of optical proximity sensor

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|---|
| 1. Small in size. | 1. Alignment, i.e. proper incidence of the light is necessary. |
| 2. Fast switching, no switch bounce. | 2. Can be disturbed by ambient light conditions such as arc welding in the proximity. |
| 3. Insensitive to vibration and shock. | 3. Dust-free, low moisture and clean environment is necessary. |
| 4. Many configurations are commercially available. | |

Applications. In pharmaceutical, food, paper, plastic and automobile industries, some of the uses of optical proximity sensors are:

1. Stack height control or box, bottle, container counting
2. Fluid level control such as bottle filling; level sensing applications for solid, grains, sand, ice and other bulk material
3. Glass sheet position sensing
4. Security and safety, e.g. collision prevention, in machine tools, presses
5. Colour sensing applications
6. Breakage and jam detection in conveyors.

Ultrasonic Proximity Sensors

The ultrasonic transducer discussed in Section 6.4 actually serves as a proximity sensor.

Ultrasonic proximity sensors are usually used to monitor the liquid and solid levels or as an approach warning system against collisions.

We now move on to the next type of measurement in Chapter 7 which deals with the measurement of strain.

Review Questions

- 6.1 What is an LVDT? What are the parameters that can be measured by this? Describe with a diagram and output characteristics the principle of its construction and operation giving typical design data. How are the readings affected by variations in ambient temperature and transverse displacements of its cylindrical core? What are the remedial measures taken to counter this effect?
- 6.2 (a) Describe, in brief, a variable resistance transducer used for measurement of small displacements.
 (b) Show how such a device of resistance R_p loses its linearity when the output is measured with the help of measuring instrument of resistance R_m .
- 6.3 The response of a variable-gap parallel-plate (two plates) capacitor transducer is nonlinear. Show, with analysis, how the response of such a device can be made linear by appropriate instrumentation.
- 6.4 What are the three major classes of digital displacement transducers? Draw the track diagrams of each of them. Discuss their relative merits and demerits. Why is the Gray code preferred to binary code in commercial encoders? How is Gray code converted to binary code?
- 6.5 What is process loading? If a potentiometer is used as displacement sensor, then prove that the relationship between load voltage and fractional displacement of wiper is nonlinear. If this nonlinearity is due to loading effect, then establish the conditions for minimum nonlinear error.
- 6.6 Draw the schematic diagram of LVDT and explain its electromechanical transfer characteristics. Show an arrangement to extract the amplitude as well as the phase information contained in the ac output of an LVDT.
- 6.7 (a) Derive the expression for error of a resistance potentiometer (pot) when connected across a load of finite resistance. Draw typical curves to show the variation of errors with input displacement for different values of load resistance.
 (b) Explain why the sensitivity and linearity are two conflicting requirements of a resistance potential divider.
- 6.8 A slide wire potentiometer AB is shown in Fig. 6.67. It has a resistance of $1 \Omega/\text{cm}$. A voltage of $E = 1 \text{ V}$ is balanced against 100 cm.

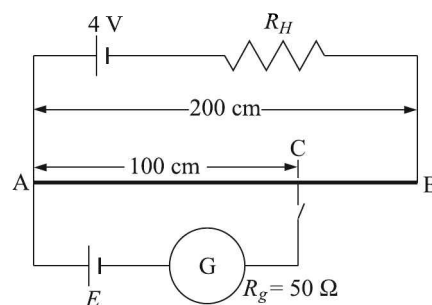


Fig. 6.67

- (a) Calculate the potentiometer current and resistance of the rheostat R_H .
- (b) Calculate the current that would flow through the galvanometer if the terminals of the battery get reversed accidentally.

6.9 Indicate the correct choice(s):

- (a) An optical fibre is characterised by
 - (i) total internal reflection
 - (ii) a core material of refractive index lower than that of the cladding
 - (iii) scattering loss
 - (iv) diffraction
- (b) A step index optical fibre, whose refractive indices of the core and cladding are 1.44 and 1.40 respectively, is surrounded by air. Its numerical aperture is
 - (i) 0.12
 - (ii) 0.75
 - (iii) 0.06
 - (iv) 0.34
- (c) The two secondary coils of an LVDT have wrongly been connected as shown in Fig. 6.68. Then the input-output relationship would be

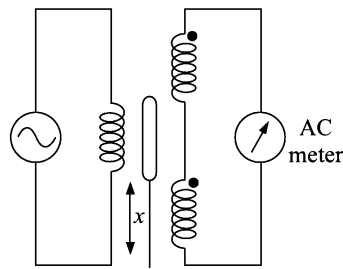
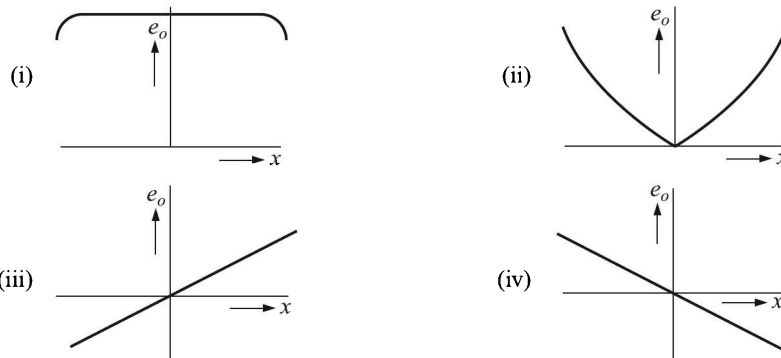
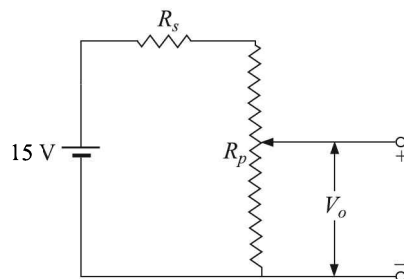


Fig. 6.68



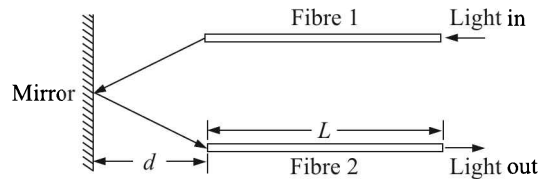
- (d) For a shaft encoder, the most appropriate 2-bit code is
 - (i) 11, 10, 01, 00
 - (ii) 11, 10, 00, 01

- (iii) 01, 10, 11, 00
 (iv) 01, 00, 11, 10
- (e) A resistance potentiometer has a total resistance of $10000\ \Omega$ and is rated $4\ \text{W}$. If the range of the potentiometer is 0 to $100\ \text{mm}$, then its sensitivity in V/mm is
 (i) 1.0
 (ii) 2.0
 (iii) 2.5
 (iv) 25
- (f) Linear variable differential transformer has
 (i) two primary coils connected in phase and a secondary coil
 (ii) two primary coils connected in opposition and a secondary coil
 (iii) one primary coil and two secondary coils connected in phase
 (iv) one primary coil and two secondary coils connected in opposition
- (g) A shaft encoder attached to a dc motor has a sensitivity of 500 pulses per revolution. A frequency meter connected to the output of the encoder indicates the frequency to be $5500\ \text{Hz}$. The speed of the motor in rpm is
 (i) 110
 (ii) 220
 (iii) 550
 (iv) 660
- (h) In a synchro pair, the control transmitter excites the three stator windings of the control transformer. The stator winding voltages will have
 (i) equal magnitudes but different phases
 (ii) different magnitudes and different phases
 (iii) equal magnitudes and phases
 (iv) different magnitudes but equal phases
- (i) The excitation frequency of an LVDT is $2\ \text{kHz}$. The maximum frequency of displacement should be limited to
 (i) $200\ \text{Hz}$
 (ii) $1.5\ \text{kHz}$
 (iii) $2\ \text{kHz}$
 (iv) $2.5\ \text{kHz}$
- (j) A $4\ \text{k}\Omega$, $0.02\ \text{W}$ potentiometer is used in the circuit shown below.

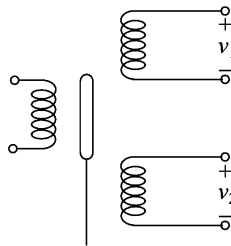


The minimum value of the resistance R_s in order to protect the potentiometer is

- (i) 2.23 k Ω
 - (ii) 2.71 k Ω
 - (iii) 3.82 k Ω
 - (iv) 8.92 k Ω
- (k) In a fibre-optic displacement sensor, shown in the following figure, the ratio of the output light intensity *does not* depend on

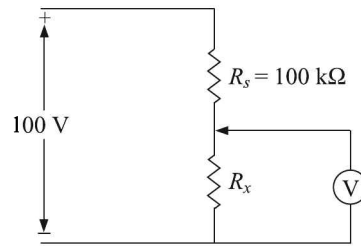


- (i) numerical aperture of the fibres
 - (ii) length L of the fibres
 - (iii) distance d
 - (iv) reflectivity of the mirror
- (l) The secondary induced voltages of an LVDT, shown in the following figure, at null position are $\bar{v}_1 = 1.0 \text{ V} \angle 0^\circ$ and $\bar{v}_2 = 1.0 \text{ V} \angle 10^\circ$.



Then the null voltage of the LVDT is

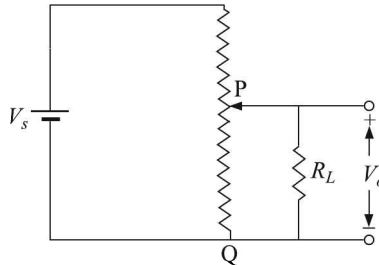
- (i) 0 V
 - (ii) 0.014 V
 - (iii) 0.174 V
 - (iv) 2 V
- (m) A variable air gap type capacitor consists of two parallel plates—a fixed plate and a plate at a distance x . If a potential V is applied across the two plates, then the force of attraction between the plates is related to x as
- (i) $F \propto x^2$
 - (ii) $F \propto \frac{1}{x^2}$
 - (iii) $F \propto \frac{1}{x}$
 - (iv) $F \propto x$
- (n) The voltmeter shown in the following figure has a sensitivity of 500 Ω/V and a full scale of 100 V. When connected to the circuit as shown, the meter reads 20 V.



The value of R_x is

- (i) 75 kΩ
- (ii) 50 kΩ
- (iii) 25 kΩ
- (iv) 10 kΩ

(o) The figure shows a potentiometer of total resistance R_T with a sliding contact.



The resistance between the points P and Q of the potentiometer at the position of the contact shown is R_C and the voltage ratio $\frac{V_o}{V_s}$ at this point is 0.5. If the ratio

$\frac{R_L}{R_T} = 1$, the ratio $\frac{R_C}{R_T}$ is

- (i) $\frac{-1 + \sqrt{5}}{2}$
- (ii) $\frac{1 + \sqrt{5}}{2}$
- (iii) $-1 + \sqrt{5}$
- (iv) $1 + \sqrt{5}$

(p) The numerical aperture of a step index fibre used in air (refractive index = 1) is 0.39. The diameter of the core is 200 μm . The angle of acceptance when the fibre is used in water (refractive index = 1.33) is closest to

- (i) 15°
- (ii) 16°
- (iii) 17°
- (iv) 18°

-
- (q) A linear variable differential transformer (LVDT) is
- (i) a displacement transducer
 - (ii) an impedance matching transformer
 - (iii) a differential temperature sensor
 - (iv) an autotransformer
- (r) To reduce the effect of fringing in a capacitive type transducer
- (i) the transducer is shielded and the shield is kept at ground potential
 - (ii) a guard ring is provided and it is kept at ground potential
 - (iii) the transducer is shielded and the shield is kept at the same potential as the moving plate
 - (iv) a guard ring is provided and it is kept at the same potential as the moving plate

6.10 Fill in the blanks:

- (a) The moving core in a linear variable differential transformer (LVDT) is made of _____ material.
- (b) In an optical fibre with the core and cladding having refractive indices n_1 and n_2 respectively, ($n_1 > n_2$), the numerical aperture is _____ when the light enters the fibre in air.
- (c) A synchro transmitter is used with a synchro repeater for _____ and with a control transformer for _____
- (i) Amplification
 - (ii) Error detection
 - (ii) Remote sensing
 - (iii) Addition

Strain Measurement

When a force is applied to a solid at rest, it gets mechanically deformed to a certain extent. If it is a tensile force, the length of the solid increases while a decrease in length results if the applied force is compressive. The ratio of the change in length ΔL to its original length L is called the longitudinal (or *axial*) strain. Thus,

$$\varepsilon = \frac{\Delta L}{L} \quad (7.1)$$

We know, atoms in a solid are bound together by what is called the interatomic forces. When no force acts on the solid, the atoms maintain definite distances among themselves at a given temperature. In equilibrium, these distances are called interatomic distances or bond lengths. The bond lengths increase or decrease when a force is applied to the solid. Then, because of the existence of interatomic forces, forces of restitution come into play to restore the atoms to their original positions of equilibrium, which is essentially the minimum energy state of the solid. Collectively, these forces of restitution per unit area constitute the stress of the solid. Following Newton's third law of motion, the stress, which is a *reaction*, is equated to the applied force per unit area which is the *action*. Therefore, the longitudinal stress σ is given by

$$\sigma = \frac{F}{A}$$

where F is the force applied on the area A .

7.1 Stress-Strain Relations

Within the elastic limit, the stress-strain relation is given by Hooke's law¹ as

$$E = \frac{\sigma}{\varepsilon} \text{ kg/m}^2$$

where E is the Young's modulus

σ is the longitudinal stress, in kg/m^2

ε is the longitudinal strain, in m/m .

When a body of length L is elongated by ΔL owing to the application of a force F , its perpendicular dimension D contracts by ΔD . The strain generated in the perpendicular direction is called the lateral strain (Fig. 7.1).

¹Named after Robert Hooke (1635–1703), an English natural philosopher, architect and polymath.

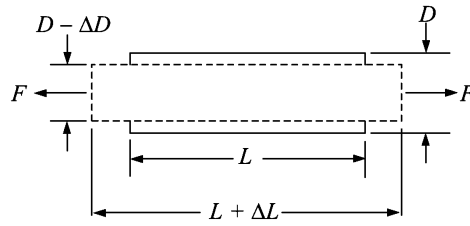


Fig. 7.1 Longitudinal and lateral strains of a solid.

Poisson² showed that the ratio between lateral strain and longitudinal strain is constant for a material. It is called the Poisson's ratio denoted by ν . Thus,

$$\nu = -\frac{\Delta D/D}{\Delta L/L} \quad (7.2)$$

From a theoretical analysis it can be shown that Poisson's ratio for all solids must lie between 0 and 0.5, i.e. $0 < \nu < 0.5$. For most of the materials, $\nu \approx 0.3$. Therefore, for an isotropic 3-D solid, with a generated stress in the x -direction, we can write the following equations:

$$\begin{aligned} \frac{\Delta x}{x} &= \varepsilon_x = \frac{\sigma_x}{E} \\ -\frac{\Delta y}{y} &= \varepsilon_y = -\nu \varepsilon_x = -\nu \frac{\sigma_x}{E} \\ -\frac{\Delta z}{z} &= \varepsilon_z = -\nu \varepsilon_x = -\nu \frac{\sigma_x}{E} \end{aligned}$$

If stresses are generated simultaneously along the other two directions, we can write

$$\left. \begin{aligned} \varepsilon_x &= \frac{\sigma_x}{E} - \nu \frac{\sigma_y}{E} - \nu \frac{\sigma_z}{E} \\ \varepsilon_y &= \frac{\sigma_y}{E} - \nu \frac{\sigma_x}{E} - \nu \frac{\sigma_z}{E} \\ \varepsilon_z &= \frac{\sigma_z}{E} - \nu \frac{\sigma_x}{E} - \nu \frac{\sigma_y}{E} \end{aligned} \right\} \quad (7.3)$$

On re-arranging Eqs. (7.3), we get

$$\left. \begin{aligned} \sigma_x &= E\varepsilon_x + \nu\sigma_y + \nu\sigma_z \\ \sigma_y &= \nu\sigma_x + E\varepsilon_y + \nu\sigma_z \\ \sigma_z &= \nu\sigma_x + \nu\sigma_y + E\varepsilon_z \end{aligned} \right\} \quad (7.4)$$

For a 2-D system, $\sigma_z = 0$. Then Eqs. (7.4) yield

$$\sigma_x = \frac{E}{1 - \nu^2} (\varepsilon_x + \nu\varepsilon_y)$$

²Siméon Denis Poisson (1781–1840) was a French mathematician, geometer and physicist.

$$\sigma_y = \frac{E}{1 - \nu^2} (\varepsilon_y + \nu\varepsilon_x)$$

This exercise shows that although strain is a quantity which is related to the volume of a substance, the measurement of its components along the three co-ordinate axes is sufficient to compute stresses in those directions.

Strain Measurement Considerations

We will confine ourselves to transducers related to measurement of longitudinal or axial strain only because all other strains can be computed from it. For axial strain, the changes in length are so small that it is difficult to measure such displacements directly except in a few isolated cases. Hence the necessity of strain gauges.³

The factors that one has to consider while measuring strain are

1. To measure strain in the concerned solid, one has to measure Δx and x in each direction. Usually, the strain magnitude is of the order of a few microstrains. Microstrain is a unit which is frequently used in strain measurements. It is defined as the change in length in μm per unit metre of length. That is

$$\text{Microstrain} = \frac{\Delta L \text{ (in } \mu\text{m)}}{L \text{ (in m)}} = \text{actual strain magnified } 10^6 \text{ times}$$

2. Strains are likely to vary from point to point. This necessitates that the strain gauge should be as small as possible in size. Usually, the gauge length is around 5 mm.

Strain measurement thus boils down to the measurement of very small (about 1 μm) displacement and gauges can be fabricated in various ways as follows:

1. *Mechanical*: ΔL is measured after magnification with the help of levers and gears. The earliest strain gauges were mechanical devices that measured strain by measuring the change in length and comparing it to the original length of the object. For example, the extension meter (extensometer) uses a series of levers to amplify strain to a readable value. In general, however, mechanical devices tend to provide low resolutions, are bulky and difficult to use.
2. *Electrical*: Changes in resistance (simple or piezo) or inductance or capacitance are measured. Although capacitance- and inductance-based strain gauges have been constructed, their sensitivity to vibration, mounting requirements, and circuit complexity have limited their application.
3. *Optical*: The phenomena of interference, diffraction and scattering of light waves are utilised to measure strain.

Of these, gauges based on the principle of change in resistance are frequently used while optical interferometric gauges are used in situations that demand very accurate measurements. We will consider only these two kinds of strain gauges.

³pronounced as *gage*.

7.2 Resistance Strain Gauges

Principle

If a conducting wire is held under tension, its length increases slightly with a consequent reduction of its area of cross-section. Let us consider a conductor of length L , cross-sectional area A and resistivity ρ . Its resistance R is given by

$$R = \rho \frac{L}{A} \quad (7.5)$$

The relation (7.5) generally holds good for common metals and many nonmetals at room temperature when subjected to direct or low frequency currents.⁴ Now we consider that the conductor is strained, i.e. either stretched or compressed. The consequent change in resistance can be expressed as

$$\Delta R = \rho \frac{L}{A} - (\rho + \Delta\rho) \frac{L + \Delta L}{A + \Delta A} \quad (7.6)$$

where Δ 's indicate changes in the corresponding quantities. For metallic wires subjected to engineering strain levels, $\Delta L \ll L$ and $\Delta A \ll A$. If $\Delta\rho \ll \rho$ as well, we can simplify Eq. (7.6) by approximating Δ with the infinitesimal differential change d as

$$\Delta R \equiv dR = d\left(\rho \frac{L}{A}\right) \quad (7.7)$$

The differential expression on the right-hand side is easier to compute by having recourse to *logarithmic differentiation* technique. We take the logarithm of Eq. (7.5) yielding

$$\ln R = \ln \rho + \ln L - \ln A \quad (7.8)$$

Now taking differential of Eq. (7.8), we get

$$\frac{dR}{R} = \frac{d\rho}{\rho} + \frac{dL}{L} - \frac{dA}{A} \quad (7.9)$$

In general, A can be written as

$$A = CD^2 \quad (7.10)$$

where D denotes a cross-section dimension and C is a constant⁵. By logarithmic differentiation of Eq. (7.10), we get

$$\frac{dA}{A} = 2 \frac{dD}{D} \quad (7.11)$$

From Eq. (7.2), we can write

$$\nu = - \frac{dD/D}{\varepsilon} \quad (7.12)$$

Therefore, from Eqs. (7.11) and (7.12), we get

$$\frac{dA}{A} = -2\nu\varepsilon = -2\nu \frac{dL}{L} \quad (7.13)$$

⁴Resistivity can also be affected by electromagnetic fields, nuclear or optical radiations, pressure and surface effects.

⁵For a wire of circular cross-section, $D = r$ and $C = \pi$.

Combining Eqs. (7.9) and (7.13), we finally get

$$\frac{dR}{R} = \underbrace{\frac{d\rho}{\rho}}_{\text{resistivity change}} + \underbrace{\frac{dL}{L}}_{\text{length change}} + \underbrace{2\nu \frac{dL}{L}}_{\text{cross-section change}} \quad (7.14)$$

The three terms which appear on the right-hand side of Eq. (7.14) indicate contributions from three factors changing the resistance of the gauge, namely,

1. Change in resistivity of the material
2. Change in length
3. Change in cross-sectional area of the gauge as indicated.

The first factor is often called the *piezoresistive change*⁶.

The gauge factor G_f , which is basically the sensitivity factor of the gauge, is defined as the change in resistance of the gauge per unit strain. From Eq (7.14), we get

$$G_f \equiv \frac{dR/R}{\varepsilon} = \frac{dR/R}{dL/L} = 1 + 2\nu + \frac{d\rho/\rho}{\varepsilon} \quad (7.15)$$

In the absence of a direct resistivity change, i.e. for $d\rho/\rho = 0$, the value of the gauge factor should be

$$1 \leq G_f \leq 2 \quad (7.16)$$

corresponding to the theoretical limits of Poisson's ratio as $0 \leq \nu \leq 1/2$. But in practice, the G_f does not behave so well for all materials. This is evident from Fig. 7.2 where we have plotted the fractional resistance change against the fractional strain, both expressed in per cents. The gauge factors are simply the slopes of curves.

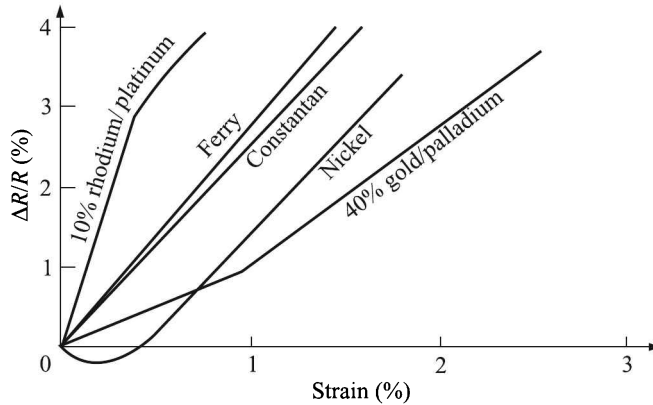


Fig. 7.2 Plot of fractional resistance change vs. fractional strain for a few materials.

⁶See Section 5.2 at page 150 for piezoresistivity.

It can be seen from the plot as well as from Table 7.1 that

1. Constantan (Ni-Cu alloy) and ferry (Ni-Cu alloy)⁷ possess a gauge factor of nearly 2. Nichrome (Ni-Cr alloy), karma (Ni-Cr-Al-Fe alloy), advance (Cu-Ni-Mn alloy)⁸—not shown in the plot—have similar curves. That means, for these materials the piezoresistive factor is negligibly small and the Poisson ratio ν is nearly 0.5.
2. Pure nickel exhibits a negative G_f (-12) for small strain. This is why it is never used alone but is almost always alloyed with other metals to construct strain gauges.
3. The 10% rhodium/platinum alloy exhibits a high G_f which is a desirable feature. But its behaviour changes abruptly at about 0.4% strain which is not desirable.
4. Semiconductors, such as p - and n -doped silicon and germanium, exhibit high G_f values, making them suitable for the construction of strain gauges.

Table 7.1 Gauge factors for strain gauge materials

| <i>Material</i> | <i>Low-strain G_f</i> | <i>High-strain G_f</i> | <i>Elongation (%)</i> |
|--------------------|------------------------------------|-------------------------------------|-----------------------|
| Copper | 2.6 | 2.2 | 0.5 |
| Constantan/ferry | 2.1 | 1.9 | 1.0 |
| 40% gold/palladium | 0.9 | 1.9 | 0.8 |
| Nickel | -12 | 2.7 | – |
| Platinum | 6.1 | 2.4 | 0.4 |
| Silver | 2.9 | 2.4 | 0.8 |
| Semiconductor | ~ -100 | ~ -600 | – |

Types

Construction and material-wise resistive strain gauges can be divided into three types, namely

1. Wire-wound
2. Foil and
3. Semiconductor

Wire-wound gauges

Wire-wound gauges are further classified into two types

1. Bonded
2. Unbonded

Bonded wire-wound gauge. A bonded wire-wound gauge is bonded directly to the surface of the specimen being tested, with a thin layer of adhesive cement. The cement not only serves to transmit the strain from the specimen to the gauge wires but also acts as an insulator. Fig. 7.3 gives a schematic view of a bonded strain gauge.

⁷It has medium-range electrical resistivity and a very low temperature coefficient of resistance (TCR).

⁸A popular material for construction of strain gauges because of its low coefficient of thermal expansion and low TCR.

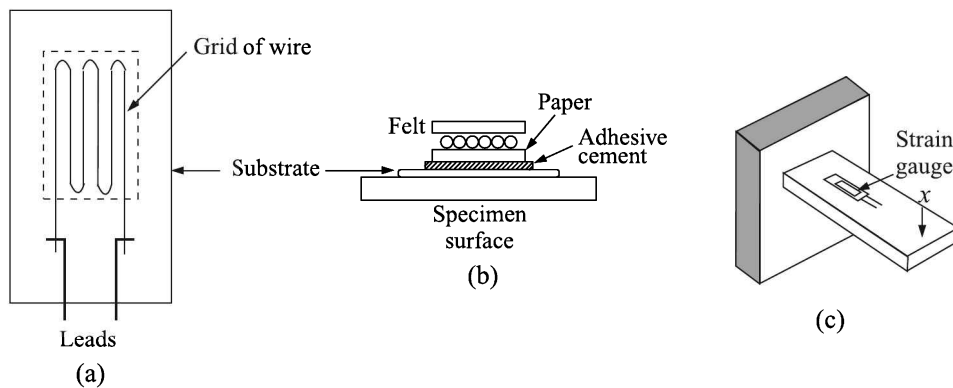


Fig. 7.3 Schematic view of a bonded strain gauge: (a) construction, (b) bonding on the surface, and (c) actual placement.

Depending upon the requirement, bonded strain gauges can have various structures though basically they are grids of fine resistance wire of about $25\ \mu\text{m}$ in diameter or less. Their general features are listed in Table 7.2.

Table 7.2 General specifications of bonded strain gauges

| | |
|-----------------------------------|--|
| <i>Size</i> | Typically $3\ \text{mm} \times 3\ \text{mm}$, but seldom bigger than $2.5\ \text{mm} \times 12.5\ \text{mm}$ |
| <i>Resistance value</i> | 120 – 1000 Ω |
| <i>Maximum excitation voltage</i> | 5 – 10 V |
| <i>Material</i> | Ni-Cu, Ni-Cr or Ni-Fe alloys |

Unbonded metal-wire gauge. An unbonded metal-wire gauge employs a set of preloaded resistance wires connected to a Wheatstone bridge. The wires are equally taut and the bridge is initially balanced at the preload (Fig. 7.4).

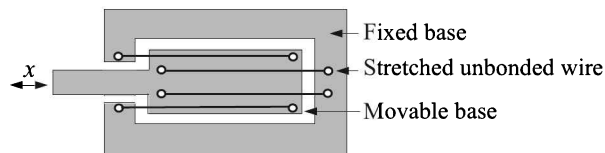


Fig. 7.4 Schematic view of unbonded strain gauge.

A small motion of the movable base increases tension in two wires while decreasing it in two others (the wires are initially stretched enough so that they never go slack) with a consequent bridge unbalance because of resistance changes. The output voltage is proportional to the input displacement which can be calibrated in terms of strain.

Foil type gauges

Basically an extension of wire gauge, a foil type strain gauge consists of sensing elements formed from thin sheets or foils, less than $5\ \mu\text{m}$ thick, by photo-etching or masked vacuum

deposition process which allows one to shape it in a suitable form according to the requirement (Fig. 7.5). The end turns are made fat so as to reduce the contribution from the transverse strain which is a spurious input.

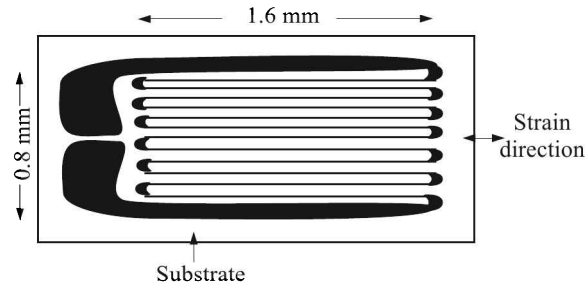


Fig. 7.5 Foil type strain gauge.

Semiconductor type gauges

As discussed in Section 5.2 at page 150, the resistivity of doped silicon and germanium undergoes a change when stressed. This property, called piezoresistivity, is utilised to construct strain gauges with these materials [Fig. 7.6(a)]. Unlike other types of strain gauges, it is interesting to note that the semiconductor strain gauges measure the change in resistance with *stress as opposed to strain*. As already pointed out, here the gauge factor is well around 100.

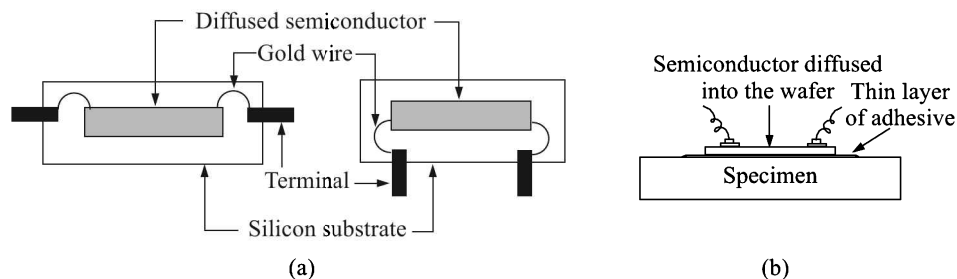


Fig. 7.6 Bonded-type semiconductor strain gauge: (a) construction, and (b) bonding.

The semiconductor bonded strain gauge is a wafer of about $150\ \mu\text{m}$ thickness with the resistance element diffused into a substrate of silicon. Usually, gold wires are used to construct electrodes because the Fermi level of gold matches that of the semiconductors and hence an ohmic contact is ensured. The wafer element is not provided with a backing, and bonding to the strained surface is done by applying a thin layer of epoxy [Fig. 7.6(b)]. Their size is smaller and the cost is lower than that of a metallic foil gauge.

Two more improvements have been made in the construction of semiconductor strain gauges:

1. A thin film of an electrical insulator, typically SiO_2 , is deposited onto the specimen surface. The thin film of the semiconductor material is then deposited on this insulator [Fig. 7.7(a)]. To ensure molecular bonding between them, vacuum deposition or

sputtering techniques are used. In this way, this thin-film gauge eliminates the need for adhesive bonding. Because this thin-film gauge is molecularly bonded to the specimen, the installation is much more stable and the resistance values experience less drift. The other advantage is that the specimen can either be a thin diaphragm or a thick beam.

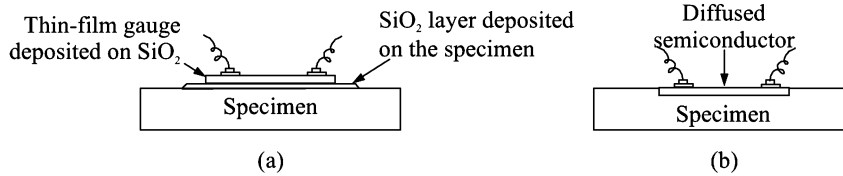


Fig. 7.7 Semiconductor strain gauges requiring no bonding: (a) thin-film having molecular bonding with insulator substrate and the specimen, and (b) directly diffused semiconductor into the specimen.

- The strain gauge material is diffused into the specimen. Such semiconductor strain gauges use photolithography masking techniques and solid state diffusion of boron to molecularly bond the resistance elements. Electrical leads are directly attached to the pattern [Fig. 7.7(b)]. By eliminating bonding agents, errors due to creep and hysteresis are eliminated. But diffused gauges are limited to moderate-temperature applications requiring temperature compensation. They are small, inexpensive, accurate and generate a strong output signal. They are often used as sensing elements in pressure transducers.

Advantages and disadvantages. Table 7.3 lists the advantages and disadvantages of semiconductor strain gauges.

Table 7.3 Advantages and disadvantages of semiconductor strain gauges

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|------------------------------------|
| 1. High unit resistance and high G_f | 1. High temperature sensitivity. |
| 2. Low hysteresis. | 2. Tendency to drift. |
| 3. Good frequency response which makes them amenable to ac measurements. | 3. Nonlinear characteristic curve. |
| 4. Very small size (0.7 to 7 mm). | |

But the disadvantages can easily be tackled by employing a second gauge for temperature compensation, and software compensation in computer-controlled instrumentation to make the calibration curve linear.

Strain Measurement Method

We have already pointed out that the displacements associated with strains are very small and therefore, corresponding changes in resistance are small. In engineering materials, the strain typically varies between 2 and 10000 microstrain or 0.000002 and 0.01. The corresponding maximum change in resistance is about 1% as will be evident from the following example.

Example 7.1

A strain gauge, having $G_f = 2.0$ and $R = 120 \Omega$, is used to measure strains generated by

pressures of 50 psi and 50000 psi in aluminium. The corresponding strains are 5 and 5000 microstrains. Calculate the per cent changes of resistance of the strain gauge.

Solution

For 5 microstrain: $\Delta R = G_f \varepsilon R = 2(5 \times 10^{-6})(120) = 0.0012 \Omega = 0.0012 \Omega$

$$\therefore \frac{\Delta R}{R} = \frac{0.0012}{120} \times 100 = 0.001$$

For 5000 microstrain: $\Delta R = 2(5000 \times 10^{-6})(120) = 1.2 \Omega$

$$\therefore \frac{\Delta R}{R} = 1\%$$

Conventional methods

Measuring such small changes in resistance with sufficient accuracy is indeed a challenging task. Let us consider two conventional methods, namely

1. Current injection
2. Ballast circuit

of measuring resistances and see how they are not adequate for strain measurements.

Current injection. The circuit is shown in Fig. 7.8(a). A constant current of known value is passed through the resistor R from a constant current source and the voltage across R is measured by a voltmeter of high input impedance. Since the current and voltage values are known, the resistance can be calculated from the Ohm's law.

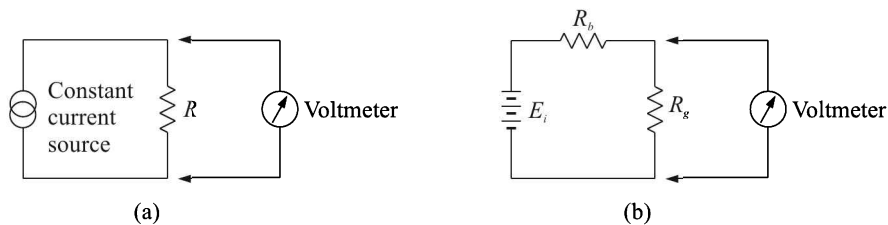


Fig. 7.8 Resistance measurement: (a) by the current injection method, and (b) by the ballast circuit method.

The drawback of using the circuit for strain measurement is that to avoid Joule heating of the resistor, a very low current has to be passed. Since the gauge resistance is not high, the voltage drop across the gauge will be small. Moreover, we have to measure the *change in resistance* which will produce an even smaller value of the voltage. The generated change in voltage is often on the order of the thermal noise.

Ballast circuit. The ballast circuit, shown in Fig. 7.8(b) is similar to that used in the current injection method. Here, in lieu of the constant current source, a voltage source in series with a ballast (high resistance) R_b is used to produce a low current to be passed through the gauge R_g . We note that for $R_b \rightarrow \infty$, the output voltage equals E_i , irrespective of the value of R_g .

The output voltage E_o is given by

$$E_o = \frac{R_g}{R_b + R_g} E_i$$

We have noted that the change in the strain gauge resistance is small. So, it is appropriate to assume that $\Delta R_g \approx dR_g$ so that the change in the output voltage can be written as

$$dE_o = \left[\frac{dR_g}{R_b + R_g} - \frac{R_g dR_g}{(R_b + R_g)^2} \right] E_i = \frac{R_b R_g}{(R_b + R_g)^2} E_i \cdot \frac{dR_g}{R_g} = \frac{R_b R_g}{(R_b + R_g)^2} E_i \cdot G_f \varepsilon \quad (7.17)$$

To write Eq. (7.17), we have utilised Eq. (7.15). Now, we need to determine the value of R_b that will maximise the sensitivity of the circuit. The sensitivity of the circuit can be defined as

$$S = \frac{dE_o}{\varepsilon} = \frac{R_b R_g}{(R_b + R_g)^2} E_i \cdot G_f \quad [\text{From Eq. (7.17)}]$$

So, to maximise S w.r.t. R_b we note

$$0 = \frac{dS}{dR_b} = \frac{(R_g - R_b)}{(R_b + R_g)^3} R_g E_i G_f \quad (7.18)$$

Equation (7.18) yields

$$R_b = R_g$$

Then from Eq. (7.17), we get

$$dE_o = \frac{G_f}{4} \varepsilon E_i \quad (7.19)$$

From Eq. (7.19), we observe that the change in voltage is indeed small. let us consider a typical case of $G_f = 2$, and $\varepsilon = 5$ microstrain. Then Eq. (7.19) yields a voltage change of $0.0000025E_i$ V which requires measurement by a digital voltmeter of 6 decades of precision!

From the above discussion it is amply clear that both the conventional current injection and ballast circuit methods are not suitable for strain measurements.

Bridge circuit method

A bridge circuit, such as a Wheatstone bridge, can be utilised in such situations with advantage. It can be utilised to make both static and dynamic measurements. In a static (or null) measurement, no current flows through the measuring instrument which is equivalent to having a measuring instrument of infinite input impedance. Thus the loading of the measured medium is almost nil here.

A dynamic measurement can be either voltage sensitive or current sensitive. In both cases, very small current flows through measuring instrument thus loading the measured medium minimally. Moreover, the advantage of a bridge measurement, be it static or dynamic, is that it is very easy to eliminate stray inputs, like temperature effects, by incorporating compensatory devices in suitable arms of the bridge.

We will consider static measurements first and then the dynamic measurements.

Static Measurements

Suppose in Fig. 7.9, R_1 corresponds to the strain gauge resistance. Then, if V_B and V_D indicate potentials at B and D respectively, we have from KVL

$$V_B = \frac{R_1}{R_1 + R_2} E_i \qquad V_D = \frac{R_3}{R_3 + R_4} E_i \qquad (7.20)$$

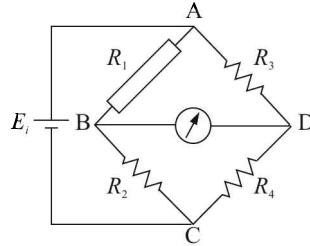


Fig. 7.9 Wheatstone bridge arrangement for strain measurement. The resistance R_1 represents the strain gauge.

When the bridge is balanced, no current flows through the galvanometer. Then

$$V_B = V_D$$

Equating the right-hand sides of Eq. (7.20), we get on simplification

$$\frac{R_1}{R_2} = \frac{R_3}{R_4} \qquad (7.21)$$

Now, if R_1 changes to $R_1 + \Delta R_1$ owing to strain, R_2 has to change to $R_2 + \Delta R_2$ to balance the bridge. It can be seen from Eq. (7.21) that

$$\Delta R_2 = \Delta R_1 \frac{R_4}{R_3}$$

And if $R_1 = R_2 = R_3 = R_4 = R_g$, which normally is, then

$$\Delta R_2 = \Delta R_1 \equiv \Delta R_g = G_f R_g \varepsilon$$

as can be seen from Eq. (7.15). Thus, Eq. (7.2) shows us that the change in resistance R_2 is a direct measure of the strain and it can be calibrated accordingly.

Dynamic Measurements

As already pointed out, dynamic measurements can be voltage-sensitive or current-sensitive. First we will consider voltage-sensitive measurements which are made with the help of a quarter-, half- or full-bridge. That means, strain gauges occupy one-fourth, half or the entire of the bridge.

Voltage-sensitive bridge

Quarter-bridge. Here, the strain-gauge constitutes one arm of the bridge (see Fig. 7.9). And if initial resistance of all the arms are equal, which normally is the case, the output voltage caused by a change in resistance in the strain gauge can be calculated from the difference in potentials at B and D as follows:

$$\begin{aligned}\Delta E_o &= \left(\frac{R + \Delta R}{2R + \Delta R} - \frac{1}{2} \right) E_i = \frac{\Delta R/R}{4 + 2\Delta R/R} E_i \\ &\cong \frac{\Delta R/R}{4} E_i \quad \left[\because 4 \gg \frac{2\Delta R}{R} \right] \\ &= \frac{G_f \varepsilon}{4} E_i\end{aligned}\quad (7.22)$$

From Eq. (7.22) we can see that the sensitivity S of the bridge is given by

$$S = \frac{\Delta E_o}{\varepsilon} = \frac{G_f E_i}{4}$$

Half-bridge. Here, strain gauges are bonded on top and bottom of the stressed member (shown in Fig. 7.10) such that if one gauge is stretched the other is compressed. That means, if the resistance of one gauge changes to $R + \Delta R$, that of the other becomes $R - \Delta R$.

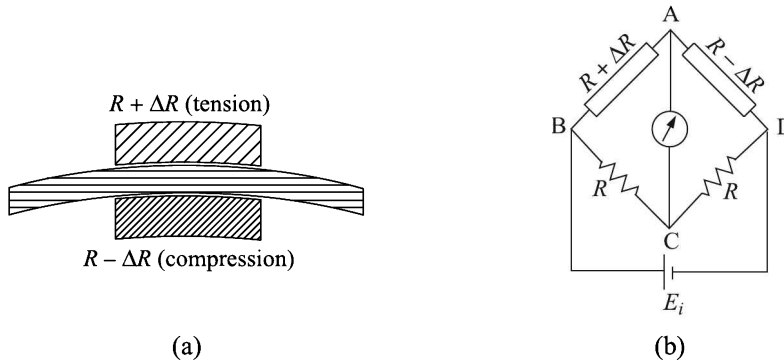


Fig. 7.10 Half-bridge arrangement for measurement of strain: (a) fixing of gauges on a cantilever, and (b) bridge configuration.

Hence, the output voltage is

$$\begin{aligned}\Delta E_o &= \left[\frac{R + \Delta R}{(R + \Delta R) + (R - \Delta R)} - \frac{1}{2} \right] E_i \\ &= \frac{\Delta R}{2R} E_i \\ &= \frac{G_f \varepsilon}{2} E_i\end{aligned}\quad (7.23)$$

and therefore, the sensitivity is

$$S = \frac{G_f E_i}{2}$$

The advantages of a half-bridge over a quarter-bridge are

1. The sensitivity is doubled
2. Unlike quarter-bridge it is not susceptible to errors arising out of change in the ambient temperature

Of course, the full bridge, which we will discuss presently, scores even higher as far as sensitivity is concerned and that it is also immune to temperature effects. Its drawback is that the gauges occupy a considerable space in this arrangement. As a result, the measurement gives an average strain value over a rather large area which may not be desirable sometimes.

Full-bridge. Here, gauges are fixed as shown in Fig. 7.11. We may easily see that the output voltage and the sensitivity of the bridge are given by

$$\Delta E_o = G_f \varepsilon E_i \qquad S = G_f E_i \qquad (7.24)$$

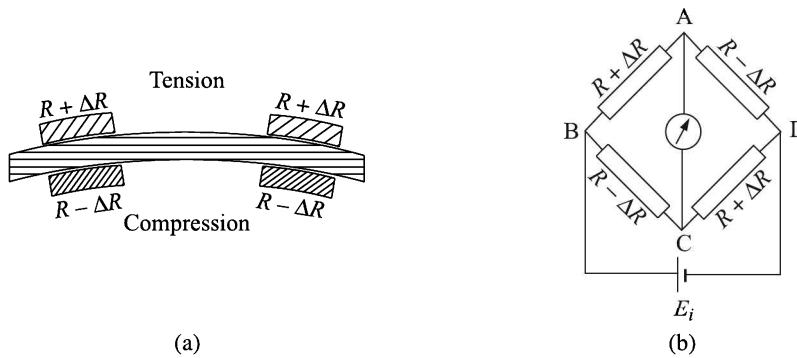


Fig. 7.11 Full-bridge arrangement for measurement of strain: (a) fixing of gauges on the specimen, and (b) bridge configuration.

Current-sensitive bridge

As already stated, here the output current is measured by a suitable ammeter. To calculate the output current, we can find out the Thevenin-equivalent resistance R_o and voltage E_o of the bridge as follows (Fig. 7.12):

$$R_o = \frac{(R + \Delta R)R}{2R + \Delta R} + \frac{R}{2} = \frac{4 + (3\Delta R/R)}{4 + (2\Delta R/R)} R \qquad (7.25)$$

$$\begin{aligned} E_o &= E_A - E_C = \left(\frac{R + \Delta R}{2R + \Delta R} - \frac{R}{2R} \right) E_i \\ &= \frac{\Delta R/R}{4 + (2\Delta R/R)} E_i \end{aligned} \qquad (7.26)$$

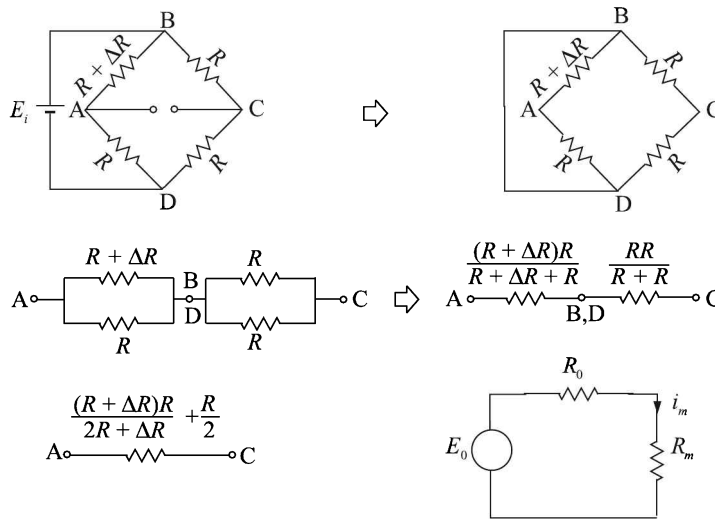


Fig. 7.12 Reduction of the bridge to its Thevenin-equivalent resistance.

The current through the measuring ammeter of resistance R_m is

$$i_m = \frac{E_m}{R_m} = \frac{E_o}{R_o + R_m}$$

which on substituting the values of E_o and R_o from Eqs. (7.26) and (7.25) becomes

$$\begin{aligned} i_m &= \frac{\Delta R}{R} \frac{1}{R_m[4 + (2\Delta R/R)] + R[4 + (3\Delta R/R)]} E_i \\ &\cong \frac{\Delta R}{R} \frac{1}{4(R + R_m)} E_i \\ &= G_f \varepsilon K \quad \left[K \equiv \frac{E_i}{4(R + R_m)} \right] \end{aligned} \quad (7.27)$$

K is a constant for the set-up. Thus, in this arrangement $i_m \propto \varepsilon$.

We now consider a few examples concerning the measurement of strain by different bridge arrangements.

Example 7.2

A 100Ω strain gauge of gauge factor 2 is connected to the first arm of a Wheatstone bridge. Under no strain condition, all the arms have equal resistance. When the gauge is subjected to a strain, the second arm resistance has to be changed to 100.56Ω to obtain a balance. Find the value of the strain.

Solution

Substituting the respective values in Eq. (7.2), we get

$$\varepsilon = \frac{\Delta R}{G_f R_g} = \frac{0.56}{2 \times 100} = 0.0028$$

Example 7.3

A bridge circuit has two fixed resistors and two strain gauges all of which have a value of 120Ω . The gauge factor is 2.04 and the strain applied to twin strain gauges, one in tension and the other in compression, is 0.000165. If the battery current in the initial balanced condition of the bridge is 50 mA, determine

- The voltage output of the bridge, and
- The sensitivity in volt per unit strain.

If the galvanometer connected to output terminals reads $100 \mu\text{V}$ per scale division and if 1/10th of a division can be read, determine the resolution.

Solution

In the initial balanced condition of the bridge, the equivalent bridge resistance is $R = 120 \Omega$. Hence, the battery voltage is

$$E_i = 50 \times 10^{-3} \times 120 = 6 \text{ V}$$

- A strain of 0.000165 produces an output of

$$\Delta E_o = \frac{G_f \varepsilon}{2} E_i = \frac{2.04 \times 0.000165 \times 6}{6} = 1.01 \text{ mV}$$

- Therefore,

$$\text{Sensitivity} = \frac{1.01 \times 10^{-3}}{0.000165} = 6.12 \text{ V/strain} = 6.12 \mu\text{V}/\mu\text{-strain.}$$

The instrument has $100 \mu\text{V}$ graduation and 1/10th of a division can be read. That means, $10 \mu\text{V}$ can be read. This corresponds to $10 \div 6.12 \cong 1.63 \mu\text{-strain}$. Therefore, the resolution is $1.63 \mu\text{-strain}$.

Example 7.4

The resistance of a strain gauge is 120Ω and its gauge factor is 2. It is connected to a current sensitive Wheatstone bridge in which all resistances are 120Ω . If the input voltage is 4 V and the resistance of the galvanometer is 100Ω , calculate the detector current in μA for $1 \mu\text{-strain}$. Also calculate the voltage output if $1 \mu\text{-strain}$ is applied to the gauge and the voltmeter has an infinite input impedance.

Solution

For the first part, we get from Eq. (7.27)

$$i_m = \frac{2 \times 1 \times 10^{-6} \times 4}{4 \times (120 + 100)} \text{ A} = 9.09 \text{ nA}$$

The second part can be calculated by multiplying both sides of Eq. (7.27) by R_m to obtain

$$E_m = i_m R_m = \frac{\Delta R}{R} \frac{E_i R_m}{4(R + R_m)} = \frac{\Delta R}{4R} \frac{E_i}{(R/R_m) + 1}$$

whence on making $R_m \rightarrow \infty$ and substituting the relevant data we get

$$E_m = \frac{G_f \varepsilon E_i}{4} = \frac{2 \times 1 \times 10^{-6} \times 4}{4} \text{ V} = 2 \mu\text{V}$$

Example 7.5

Figure 7.13 shows a Wheatstone bridge circuit for strain measurement. All the arms of the bridge are strain gauges with identical no-strain resistances of value 120 ohms and identical gauge factors of value 2.

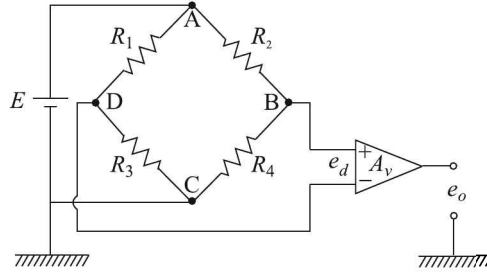


Fig. 7.13 Example 7.5.

- Find an expression for the bridge output e_d in terms of the four strains ε_1 , ε_2 , ε_3 and ε_4 .
- Assume that all the gauges experience the same magnitudes of strain. Gauge R_1 is in tension. Which of the remaining three gauges should be in tension and which in compression for maximum sensitivity of the bridge?
- If the gain of the instrumentation amplifier $A_v = 1000$, excitation voltage $E = 1$ V, strain magnitude $|\varepsilon| = 10^{-4}$ for all four arms, find the output voltage e_o .

Solution

If V_D and V_B indicate voltages at the points D and B respectively,

$$V_D = \frac{R_1}{R_1 + R_3} E \quad V_B = \frac{R_2}{R_2 + R_4} E$$

Therefore,

$$\begin{aligned} e_d = V_D - V_B &= \left(\frac{R_1}{R_1 + R_3} - \frac{R_2}{R_2 + R_4} \right) E \\ &= \left[\frac{R_1(R_2 + R_4) - R_2(R_1 + R_3)}{(R_1 + R_3)(R_2 + R_4)} \right] E \\ &= \left[\frac{R_1 R_4 - R_2 R_3}{(R_1 + R_3)(R_2 + R_4)} \right] E \end{aligned}$$

$$\text{Now, } R_1 = R + \Delta R_1 \quad R_2 = R + \Delta R_2 \quad R_3 = R + \Delta R_3 \quad R_4 = R + \Delta R_4$$

$$\text{and } \Delta R_1 = G_f \varepsilon_1 R \quad \Delta R_2 = G_f \varepsilon_2 R \quad \Delta R_3 = G_f \varepsilon_3 R \quad \Delta R_4 = G_f \varepsilon_4 R$$

So,

$$\begin{aligned} R_1 R_4 - R_2 R_3 &= (R + \Delta R_1)(R + \Delta R_4) - (R + \Delta R_2)(R + \Delta R_3) \\ &= [R^2 + R(\Delta R_1 + \Delta R_4)] - [R^2 + R(\Delta R_2 + \Delta R_3)] \\ &\quad \text{[neglecting product terms]} \\ &= R[(\Delta R_1 + \Delta R_4) - (\Delta R_2 + \Delta R_3)] \end{aligned}$$

$$\begin{aligned}
(R_1 + R_3)(R_2 + R_4) &= (2R + \Delta R_1 + \Delta R_3)(2R + \Delta R_2 + \Delta R_4) \\
&= 4R^2 + 2R(\Delta R_1 + \Delta R_2 + \Delta R_3 + \Delta R_4) \\
&\quad \text{[neglecting product terms]} \\
&\cong 4R^2 \\
[\because 4R^2 &\gg 2R(\Delta R_1 + \Delta R_2 + \Delta R_3 + \Delta R_4)]
\end{aligned}$$

(a) Therefore,

$$\begin{aligned}
e_d &= \frac{(\Delta R_1 + \Delta R_4) - (\Delta R_2 + \Delta R_3)}{4R} E \\
&= \frac{G_f R [(\varepsilon_1 + \varepsilon_4) - (\varepsilon_2 + \varepsilon_3)]}{4R} E \\
&= \frac{G_f}{4} [(\varepsilon_1 + \varepsilon_4) - (\varepsilon_2 + \varepsilon_3)] E
\end{aligned}$$

(b) For maximum sensitivity, i.e. $(e_d)_{\max}$, the numerator of the equation in (a) should be maximum. This means that, ε_1 and ε_4 should be high and ε_2 and ε_3 should be low. In other words, R_1 and R_4 should be in tension, and R_2 and R_3 should be in compression.

(c) Given: $G_f = 2$, $R = 120 \Omega$, $|\varepsilon| = 10^{-4}$, $E = 1 \text{ V}$ and $A_v = 1000$. $\therefore R_1$ and R_4 are in tension, and R_2 and R_3 are in compression

$$\begin{aligned}
e_o &= A e_d = A \cdot \frac{G_f}{4} \cdot 4|\varepsilon| \cdot E \\
&= (1000) \left(\frac{2}{4} \right) (4 \times 10^{-4})(1) \\
&= 0.2 \text{ V}
\end{aligned}$$

Temperature Effects and Compensation

Variations in the ambient temperatures affect the strain measurements by Wheatstone bridges in the following three ways:

1. Change in the gauge factor of the strain gauge
2. Temperature-induced strain in the gauge element
3. Temperature-induced resistance changes in long lead wires.

Change in the gauge factor

The change in the gauge factor with temperature for a few materials is presented in graphical form in Fig. 7.14.

The gauge factors of copper-nickel alloys, such as advance, are relatively insensitive to operating temperature variations, making them the most popular choice for strain gauge materials. To have an idea as to how the temperature variation of G_f affects the strain measurement, let us consider a few examples.

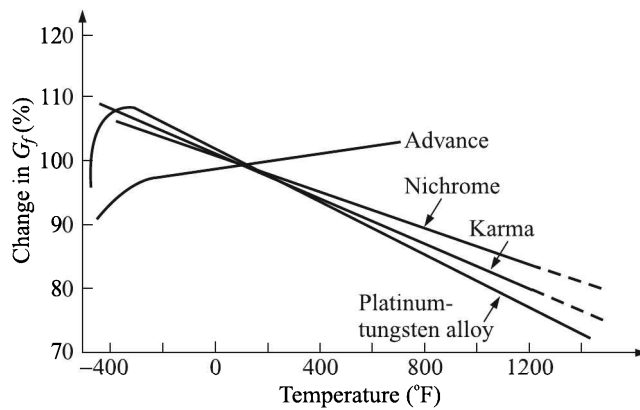


Fig. 7.14 Variation in gauge factors of the various strain gauge materials as a function of operating temperature.

Example 7.6

The TCR for isoelastic⁹ is 260 ppm/°F. Its G_f is 3.5. Calculate the apparent strain produced by a variation of 1°F of temperature.

Solution

The generated apparent strain is

$$\varepsilon = \frac{dR/R}{G_f} = \frac{260 \times 10^{-6}}{3.5} = 74 \text{ microstrain}$$

Example 7.7

A gauge, made of a material having a temperature coefficient of resistance of 12×10^{-4} per °C, has a resistance of 120 Ω and a gauge factor of 2. It is connected to a bridge having resistances of 120 Ω each. The bridge is balanced at ambient temperature. If the temperature changes by 20°C, find

- the output voltage of the bridge if the input voltage is 10 V, and
- the equivalent strain represented by the change in temperature.

Solution

- Change in resistance of the gauge due to the change in temperature is

$$\Delta R = \alpha R \Delta T = (120 \times 12 \times 10^{-4} \times 20) \Omega = 2.88 \Omega$$

Since this is a quarter bridge, from Eq. (7.22), we get

$$\Delta E_o = \frac{\Delta R}{4R} E_i = \frac{2.88}{4 \times 120} \times 10 \text{ V} = 0.06 \text{ V}$$

- The strain corresponding to a resistance change of 2.88 Ω is

$$\frac{\Delta R}{G_f \varepsilon} = \frac{2.88}{120 \times 2} = 0.012 = 12000 \text{ microstrain}$$

⁹An alloy of 36% Ni, 8% Cr, 0.5% Mo and 55.5% Fe.

Temperature-induced strain in the gauge element

A difference in the coefficients of thermal expansion between the gauge and the substrate material may also generate spurious strain readings. Though the total cross-sectional area of the gauge element transverse to its measuring axis is very small compared to the similar area of the substrate material, actually all the substrate thermal strain is transferred to the gauge. The situation will be clear from Example 7.8.

Example 7.8

A constantan strain gauge is fixed on an aluminium substrate. The coefficients of thermal expansion of constantan and aluminium are 8 ppm/°F and 13 ppm/°F respectively. The TCR and G_f of constantan are 6 ppm/°F and 2.0. Calculate the net apparent strain for a temperature change of 10°F.

Solution

The strain due to thermal expansion is

$$\varepsilon = 13 - 8 = 5 \text{ microstrain}/^\circ\text{F}$$

The strain due to the resistive variation of constantan is

$$\varepsilon = \frac{dR}{G_f} = \frac{6}{2} = 3 \text{ microstrain}/^\circ\text{F}$$

Therefore, the net apparent strain due to resistive as well as expansion effects for a variation of temperature of 10°F will be nearly

$$\varepsilon = (3 + 5) \times 10 = 80 \text{ microstrain}$$

It is a common practice for strain gauge manufacturers to treat the gauge element materials so that their thermal expansions equalise those of the substrates over a reasonable temperature range.

Temperature-induced resistance changes in the lead wires

Strain gauges are sometimes mounted at a distance from the measuring equipment. This introduces the possibility of errors creeping in owing to

1. Temperature variations
2. Reduction of the sensitivity of the bridge
3. Lead-wire resistance changes.

Such errors can be substantial if the lead wires are over 3 m long. In a two-wire installation [Fig. 7.16(a)], the two leads are in series with the strain gauge element and therefore, any change in the lead wire resistance (R_l) cannot be distinguished from the changes in the resistance R_g of the strain gauge. Because then the bridge output can be expressed, using Eq. (7.22), as

$$\Delta E_o = \frac{1}{4} \frac{\Delta R_g}{R_g + R_l} E_i = \frac{1}{4} \frac{\Delta R_g}{R_g(1 + \alpha)} E_i \quad (7.28)$$

where

$$\alpha = \frac{R_l}{R_g} = \frac{\text{Total leadwire resistance}}{\text{Gauge resistance}} \quad (7.29)$$

Then using the definition of gauge factor, we get from Eq. (7.28)

$$\Delta E_o = \frac{1}{4} \frac{G_f}{1 + \alpha} \varepsilon E_i \equiv \frac{1}{4} (G_f)_{\text{eff}} \varepsilon E_i \quad (7.30)$$

where
$$(G_f)_{\text{eff}} = \frac{G_f}{1 + \alpha} \approx G_f(1 - \alpha) \quad \text{for } \alpha \ll 1 \quad (7.31)$$

Equation (7.31) demonstrates that the error introduced by a substantial lead-wire resistance becomes significant if the ratio α exceeds 0.1%. It effectively lowers the sensitivity of the bridge circuit by reducing the gauge factor.

Temperature compensation

We have already discussed in Section 7.2 that the half-bridge and full-bridge configurations for measurement of strain are automatically compensated for temperature effects. However, the quarter-bridge arrangement does not enjoy this immunity. The temperature compensation can be made here by incorporating a *dummy* gauge in one of the arms of the bridge as shown in Fig. 7.15.

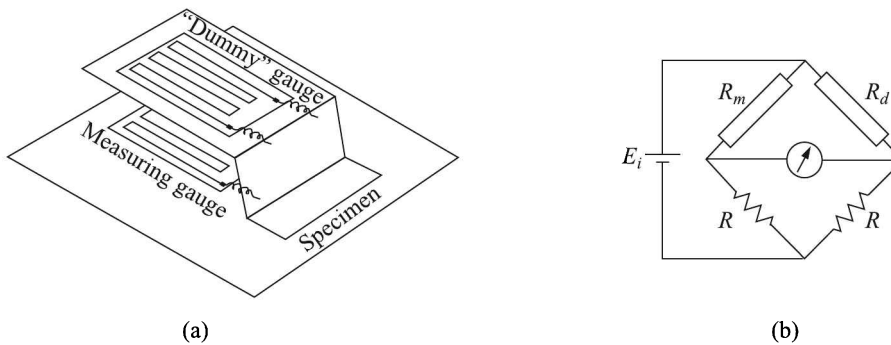


Fig. 7.15 Temperature compensation for quarter-bridge: (a) Placement of the *dummy* gauge on the specimen, and (b) incorporation of the *dummy* (R_d) in the bridge.

To correct for lead-wire effects, an additional third lead can be incorporated at the top arm of the bridge as shown in Fig. 7.16(b). In this configuration, the third wire is not a part of the bridge. It merely acts as a sense lead with almost no current flowing through it. Theoretically, if the lead wires to the strain gauge have the same nominal resistance, the same TCR, and are maintained at the same temperature, full compensation is obtained.

Of course, the simplest way of compensating lead-wire effect is to use $(G_f)_{\text{eff}}$ value for the gauge factor as suggested by Eq. (7.30), while calculating strain from the output voltage change.

Example 7.9

A 120Ω strain gauge having $G_f = 2.0$ and located 100 m away from the measuring bridge is used to measure strain. The lead wire is 29 SWG copper which has a resistance of $43.4 \Omega/100 \text{ m}$. What value of the gauge factor should be used to calculate strain?

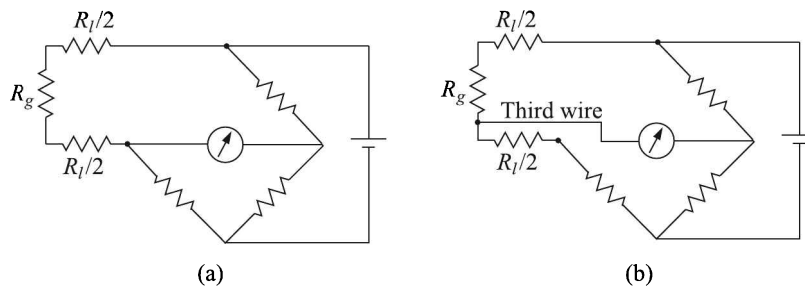


Fig. 7.16 (a) Resistance of long lead-wires interfering in the bridge measurement, and (b) three-wire connection to correct for lead-wire effects.

Solution

The total resistance of the lead wire is

$$R_l = 2 \times 100 \times \frac{43.4}{100} = 86.8 \Omega$$

Therefore, from Eq. (7.29)

$$\alpha = \frac{86.8}{120} = 0.72$$

$$(G_f)_{\text{eff}} = \frac{G_f}{1 + \alpha} = \frac{2.0}{1.72} = 1.16$$

Example 7.10

An electrical resistance strain gauge of resistance 120Ω has a gauge factor of 2. It is bonded to a steel specimen (modulus of elasticity, $E = 2 \times 10^{11} \text{ N/m}^2$) for measuring strain. Estimate

- Strain induced in the specimen if a tensile stress of $60 \times 10^6 \text{ N/m}^2$ is applied on the specimen.
- Change in the electrical resistance of the gauge due to the tensile stress as given in (a).
- Change in the electrical resistance of the gauge if there is an increase of temperature by 40°C .

Assume the following data:

| | |
|---|--|
| Temperature coefficient of resistance of the gauge | $= 20 \times 10^{-6} \text{ per } ^\circ \text{C}$ |
| Thermal coefficient of linear expansion of the gauge | $= 16 \times 10^{-6} \text{ per } ^\circ \text{C}$ |
| Thermal coefficient of linear expansion of the steel specimen | $= 12 \times 10^{-6} \text{ per } ^\circ \text{C}$ |

Solution

Given:

$$G_f = 2 \quad E = 2 \times 10^{11} \text{ N/m}^2 \quad \alpha_{T|\text{gauge}} = 16 \times 10^{-6} \text{ per } ^\circ \text{C} \quad \alpha_R = 20 \times 10^{-6} \text{ per } ^\circ \text{C}$$

$$R = 120 \Omega \quad \sigma = 60 \times 10^6 \text{ N/m}^2 \quad \alpha_{T|\text{steel}} = 12 \times 10^{-6} \text{ per } ^\circ \text{C}$$

$$(a) \text{ Strain} = \frac{\sigma}{E} = \frac{60 \times 10^6}{2 \times 10^{11}} = 300 \mu\text{-strain}$$

(b) $\Delta R = G_f \varepsilon R = (2)(300 \times 10^{-6})(120) = 0.072 \Omega$

(c) Change in the resistance owing to the change in temperature is

$$\Delta R = R \alpha_R \Delta T = (120)(20 \times 10^{-6})(40) = 0.096 \Omega$$

Strain in the gauge due to the differential expansion of steel and the gauge material is

$$\Delta \varepsilon = (\alpha_T|_{\text{steel}} - \alpha_T|_{\text{gauge}}) \Delta T = [(12 - 16) \times 10^{-6}](40) = -1.6 \times 10^{-4}$$

Note: Since the steel, to which the gauge is bonded, expands less, the gauge is in compression, rather than in tension.

Corresponding change in the resistance of the gauge is

$$\Delta R' = G_f \Delta \varepsilon R = (2)(-1.6 \times 10^{-6})(120) = -0.0384 \Omega$$

Therefore, the total change in the resistance is

$$(0.096 - 0.0384) = 0.0576 \Omega$$

Bridge Excitation Voltage

Significant temperature changes may be produced in the gauge itself through $i^2 R$ Joule heating if proper excitation voltage is not chosen to drive the bridge circuit. Of course, the resultant temperature of the gauge depends not only on the applied power but also on the ability of the gauge to dissipate heat. The ability of heat dissipation depends on the thermal conductivity of the material on which the gauge is mounted.

The rule of thumb is to apply a power density of 7.5 to 1500 mW per cm^2 of the gauge element depending on the required accuracy of measurement as well as the thermal conductivity and thickness of the substrate. Table 7.4 gives an idea about the requirement of power density for different situations.

Table 7.4 Power density requirement with respect to accuracy of measurement and substrate

| Required accuracy | Substrate | | Power density (ρ_E) (mW/cm ²) |
|-------------------|----------------------|-----------|---|
| | Thermal conductivity | Thickness | |
| High | Good | Thick | 300–750 |
| | | Thin | 150–300 |
| | Poor | Thick | 75–150 |
| | | Thin | 7.5–30 |
| Average | Good | Thick | 750–1500 |
| | | Thin | 150–750 |
| | Poor | Thick | 150–300 |
| | | Thin | 15–75 |

Now, the power generated within a gauge is

$$P_g = E_g i_g$$

It is easy to see that because of the symmetry of a Wheatstone bridge, if E_i is the excitation voltage and i is the bridge current, the current through any arm is $i/2$ and the voltage developed across any arm is $E_i/2$. Therefore,

$$P_g = \left(\frac{E_i}{2}\right) \left(\frac{i}{2}\right) = \frac{1}{4}E_i i \quad (7.32)$$

The net bridge resistance R_{bridge} as seen by the power supply, when all arms have the same resistance R_g , is¹⁰ R_g . So, the current through the bridge is

$$i = \frac{E_i}{R_{\text{bridge}}} = \frac{E_i}{R_g} \quad (7.33)$$

From Eqs. (7.32) and (7.33), we get

$$P_g = \frac{E_i^2}{4R_g} \quad (7.34)$$

The required power density is given by

$$\rho_E = \frac{P_g}{A_g} \quad (7.35)$$

where A_g is the area of cross-section of the gauge. Using Eqs. (7.34) and (7.35), we get the following expression for the maximum excitation voltage

$$E_i|_{\text{max}} = 2\sqrt{\rho_E R_g A_g} \quad (7.36)$$

Example 7.11

A constantan strain gauge of dimension 3 mm × 10 mm and resistance 120 Ω is fixed on a thin aluminium substrate. Calculate the maximum excitation voltage of the bridge if a measurement of high accuracy is needed.

Solution

The gauge parameters are

$$R_g = 120 \Omega \qquad A_g = 0.3 \times 1.0 = 0.3 \text{ cm}^2$$

The thermal conductivity of the aluminium substrate is good, but it is thin. So, from Table 7.4, we get

$$\rho_E|_{\text{max}} = 300 \text{ W/cm}^2$$

Therefore from Eq. (7.36), we get

$$E_i|_{\text{max}} = 2\sqrt{(0.3)(120)(0.3)} \cong 10 \text{ V}$$

¹⁰ $R_{\text{bridge}} = 2R_g \parallel 2R_g = R_g$.

Calibration of Strain Gauges

In a dynamic measurement, a strain gauge attached to a bridge produces an output voltage when stressed. Depending upon the nature of the bridge—quarter- or half- or full—this output voltage is related to the input strain, gauge-factor and the bridge excitation voltage [see Eqs. (7.22), (7.23), (7.24)]. If the last two factors are kept constant, the output voltage has a particular linear dependence on the strain for a fixed type of bridge. Thus, the output voltage can be calibrated in terms of known strains which, in turn, can be generated by known stresses, such as by putting known weights on the medium.

In static measurements, however, the change in resistance of the strain gauge is measured and the strain is calculated therefrom once the gauge factor and the strain gauge resistance are known [see Eq. (7.2)]. The strain gauge resistance can be measured by varying the resistance of the third or fourth arm of the bridge and the resistance of this arm may be calibrated in terms of the strain.

Alternatively, the strain gauge itself can have a parallel variable resistance as shown in Fig. 7.17. Whenever the strain gauge changes resistance owing to the generation of stress

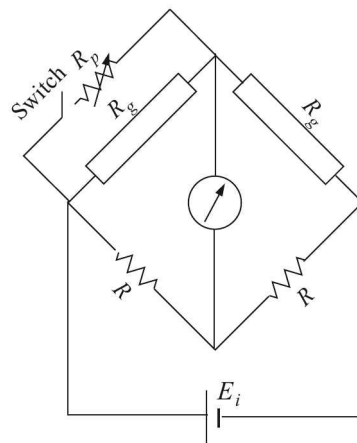


Fig. 7.17 Parallel or shunt resistance arrangement for calibration of strain-gauge.

in the measuring medium, the bridge is unbalanced because of a change in current in the two branches of the bridge. The balance can be restored by adjusting the parallel resistance R_p . This R_p can be calibrated in terms of strain as follows.

The switch of the shunt resistance is initially open and the bridge is balanced with no strain on strain gauges. Then if there is a strain, the bridge is off balance, switch is closed and a shunt resistance of value R_p is used to restore balance of the bridge. What is the value of the strain?

The change in resistance ΔR of the first arm of the bridge due to inclusion of R_p is

$$\Delta R = R_g - (R_g \parallel R_p) = \frac{R_g^2}{R_g + R_p} \quad (7.37)$$

The corresponding value of the strain is

$$\begin{aligned}\varepsilon &= \frac{1}{G_f} \cdot \frac{\Delta R}{R_g} = \frac{1}{G_f} \cdot \frac{R_g^2}{R_g(R_g + R_p)} \quad [\text{using Eq. (7.37)}] \\ &= \frac{R_g}{G_f(R_g + R_p)}\end{aligned}\quad (7.38)$$

Equation (7.38) can be rewritten in terms of the parallel resistance as

$$R_p = \frac{R_g(1 - G_f\varepsilon)}{G_f\varepsilon} \quad (7.39)$$

Equation (7.39) shows a nonlinear relation between R_p and ε , which is disadvantageous. However, for all practical purposes $G_f\varepsilon \ll 1$ and hence

$$R_p \cong \frac{R_g}{G_f\varepsilon}$$

which shows that R_p and ε^{-1} have a linear relationship. Thus, the shunt resistance may be calibrated in terms of ε^{-1} .

Note: Although the variation of a strain gauge resistance can be accurately measured, the derivation of the corresponding value of strain depends upon knowing the value of the gauge factor G_f accurately. This is a rather difficult proposition because of the following reasons:

1. The value of G_f cannot be determined unless the gauge is bonded to a specimen. Even if it is bonded, the value of G_f can be calculated from the measured value only from a theoretically calculated value of strain.
2. Once a gauge is bonded to a specimen, it cannot be removed. Therefore, the value of G_f supplied by the manufacturer is not a measured value for that particular gauge, but is based on an average value of that type of gauge. That means, the precision of the supplied value depends on the statistical quality control of the product. Typically, the precision is $\pm 1\%$.
3. For semiconductor gauges, G_f is not strictly constant; it varies with the strain, though the variation is small.
4. For all gauges, G_f may not remain constant if the gauge is used beyond a prescribed limit of cycle of measurement.

These difficulties notwithstanding, we can achieve very high accuracy with strain gauge measurements if we calibrate the gauge *in situ*.

Example 7.12

The static calibration of strain gauge bridges is carried out by connecting a standard resistor shunted across one of the arms. If the nominal value of the strain gauge resistance is $120\ \Omega$ and the gauge factor is 2.0, calculate the shunt resistance needed to obtain equivalent microstrain levels of (a) 300, and (b) 2000.

Solution

$1 \mu\text{-strain} = 1 \mu\text{m}/1 \text{ m} = 1 \times 10^{-6} \text{ m/m}$. Therefore, 300 and 2000 $\mu\text{-strains}$ correspond to 0.0003 m/m and 0.002 m/m respectively. Plugging in these values in Eq. (7.39) we get the values of shunt resistance as

$$(a) \text{ For } 300 \mu\text{-strain, } R_p = \frac{120(1 - 2 \times 0.0003)}{2 \times 0.0003} = 199,880\Omega \simeq 200\text{k}\Omega$$

$$(b) \text{ For } 2000 \mu\text{-strain, } R_p = \frac{120(1 - 2 \times 0.002)}{2 \times 0.002} = 29,880\Omega \simeq 30\text{k}\Omega$$

7.3 Fibre-optic Strain Gauges

Fibre-optic strain gauges are of recent origin. With proper modification of optical fibres, tiny interferometers or gratings are constructed. The strain, which generates a small displacement, is measured with the help of these interferometers or gratings. They can be of three types:

1. Fabry-Pérot interferometer type
2. Bragg grating type
3. Brillouin scattering type

Fabry-Pérot Interferometer Type

A Fabry-Pérot interferometer¹¹ is a linear optical resonator which consists of two reflecting mirrors (with small transmittivity) and is often used as a high-resolution optical spectrometer¹².

As can be seen from Fig. 7.18, the incident light suffers multiple reflection in the gap between the two mirrors and thus the emerging light rays have path differences. As a result, these rays interfere constructively and destructively to produce fringes. The region between the two mirrors is referred to as the *cavity*.

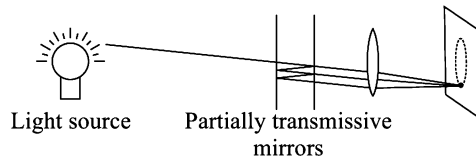


Fig. 7.18 Fabry-Pérot interferometer schematic.

Depending on their construction, the Fabry-Pérot interferometers, used in strain measurements, are of two types

1. Extrinsic Fabry-Pérot interferometric (EFPI) sensor
2. Intrinsic Fabry-Pérot interferometric (IFPI) sensor

Extrinsic Fabry-Pérot interferometric (EFPI) sensor

The construction of the extrinsic Fabry-Pérot interferometric sensor is shown in Fig. 7.19.

¹¹aka Fabry-Pérot resonator.

¹²See, for example, *Fundamentals of Optics*, FA Jenkins and HE White, McGraw-Hill, Section 14.10.

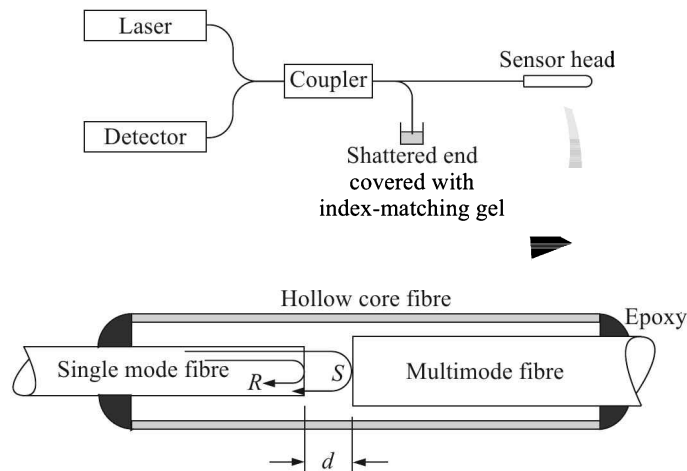


Fig. 7.19 Extrinsic Fabry-Pérot interferometric sensor.

The cavity is formed between a single-mode fibre¹³ and a reflecting target sealed inside a hollow core tube. The reflecting target may be a multimode fibre. The single-mode fibre acts as the lead-in/lead-out fibre while the multimode fibre acts solely as a reflector. Since the cavity is external to the lead-in/lead-out fibre, the interferometer is *extrinsic*. The hollow core tube acts as a guide tube in addition to protecting the cavity from external elements. Light, entering through the single-mode fibre, is partially reflected from the glass-air interface as R . This is called the *reference reflection*. The transmitted light travels through the cavity, is reflected from the air-glass interface and enters the single mode fibre again as S which is called the *sensing reflection*. These reflections then interfere in the single-mode fibre. The output depends on the difference in the optical path lengths between the two interfering waves. The effects of subsequent reflections inside the cavity are negligible. The detected output intensity I_{det} is approximately given by

$$I_{\text{det}} \approx A^2 \left[1 + \frac{2rt}{r + 2d \tan[\sin^{-1}(N_A)]} \cos\left(\frac{4\pi d}{\lambda}\right) + \left(\frac{rt}{r + 2d \tan[\sin^{-1}(N_A)]}\right)^2 \right]$$

where

- A is the reference reflection coefficient
- t is the transmission coefficient of the air-glass interface
- r is the radius of the core of the fibre
- d is the length of the air-gap
- N_A is the numerical aperture of the single-mode fibre
- λ is the wavelength of the incident radiation.

The typical variation in the detected intensity as a function of air-gap length is shown in Fig. 7.20.

Small displacements that result in operation around the quiescent or Q-point of the sensor lead to a near linear variation of the I_{det} vs. the length of the air gap, i.e. the displacement of the target. For larger displacements, I_{det} may vary over several sinusoidal periods. In this

¹³For information on optical fibres, see Section 6.3 at page 206.

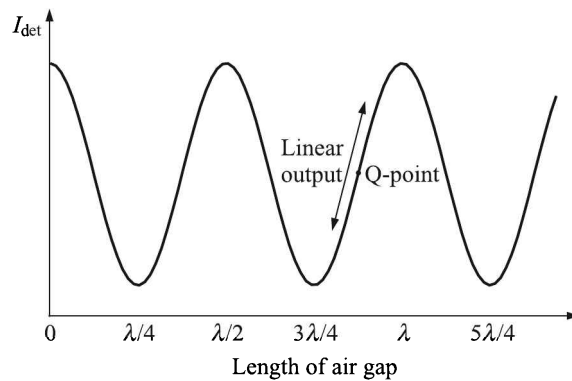


Fig. 7.20 Variation of the detected intensity as a function of air-gap length.

case, a *fringe* in I_{det} is defined from a maximum to the next maximum (or from a minimum to the next minimum) and each fringe corresponds to a change in the cavity length by $\lambda/2$. If Δd is the change in the cavity length, the strain is calculated from the relation

$$\varepsilon = \frac{\Delta d}{d} \quad (7.40)$$

where the gauge length d is defined as the distance between the input and reflecting fibres at no strain condition. Various signal demodulation schemes have been developed for EFPI-based sensors. A few of them are discussed below.

Quadrature phase shifted (QPS) EFPI. The QPS EFPI demodulation scheme allows relative measurement of the change in air gap length. In this scheme, two different cavities with outputs 90° out of phase or *in quadrature* are fabricated. Figures 7.21(a) and 7.21(b) show the arrangement of cavities and signal outputs in this scheme.

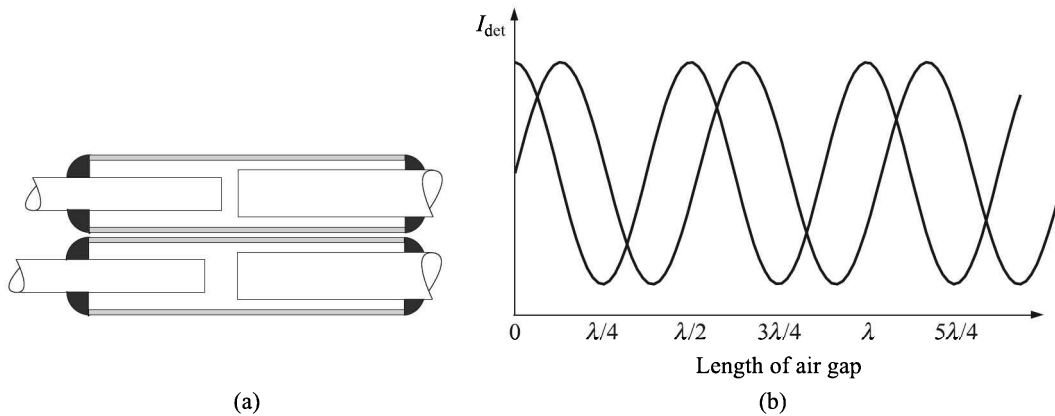


Fig. 7.21 QPS EFPI: (a) Arrangement of cavities, and (b) signal output.

Since the system consists of two different cavities resonating in quadrature with each other, at any given instant, at least one of the cavities operates in its most sensitive region. By

monitoring the phase lead-lag of the two signals, information about the direction can be unambiguously obtained. Thus, the measurement is independent of the initial operating point.

The disadvantages of this scheme include

1. Necessity of complex fringe counting methods
2. Difficulty in fabrication of the cavities
3. Maintenance of the quadrature phase shift under repeated strain conditions.

Dual wavelength method. Here, only one cavity is used but it is illuminated with radiations of two wavelengths λ_1 and λ_2 . The corresponding phase shifts are

$$\Delta\phi_1 = \frac{4\pi d}{\lambda_1}$$

$$\Delta\phi_2 = \frac{4\pi d}{\lambda_2}$$

where d is the length of the air gap. The relative phase difference $\Delta\phi$ between the two wavelengths is, therefore,

$$\Delta\phi = \Delta\phi_1 - \Delta\phi_2 = 4\pi d \frac{\lambda_1 - \lambda_2}{\lambda_1 \lambda_2} = 4\pi d \frac{\Delta\lambda}{\lambda_1 \lambda_2}$$

The range is limited to $0 < \Delta\phi < \pi$ radians. The typical range for this scheme is 40 μm . For a given range, monitoring the lead-lag of the phase of the two signals yields directional information.

White light interferometry. The EFPI in this case is illuminated with a broadband source and the output is analysed by a spectrometer. Waves, for which the phase difference between the reference and sensing reflections is a multiple of 2π , interfere constructively and show up as peaks on the output spectrum. The gap length is then determined by using the relation

$$d = \frac{\lambda_1 \lambda_2}{2(\lambda_1 - \lambda_2)} \quad (7.41)$$

where λ_1 and λ_2 are the wavelengths of any two subsequent peaks in the optical spectrum.

One of the many advantages of this technique is the elimination of complex fringe counting methods. Another major advantage is the absolute gap length detection. The use of the spectrometer slows the frequency response of the system to 5 Hz. In addition, to obtain accurate information, the system has to be operated in a transmissive mode with a high *finesse*¹⁴ EFPI cavity which complicates fabrication. The system described above can have a range of 0 to 500 μm with a resolution of 1 μm .

The EFPI can measure strain with a resolution of < 1 microstrain and has a dynamic range of > 80 dB. Moreover, since the cavity is external to the fibre, transverse strain components that tend to influence the response of intrinsic sensors have negligible effect on EFPI sensors.

¹⁴Defined as its free spectral range divided by the FWHM (full width at half-maximum) bandwidth of its resonances. It is fully determined by the resonator losses and is independent of the resonator length. The *finesse* is related to the Q -factor where $Q = (\text{finesse} \times \text{resonance frequency}) / (\text{free spectral range})$.

Intrinsic Fabry-Pérot interferometric (IFPI) sensor

The intrinsic Fabry-Pérot interferometric sensor works in a similar way as the EFPI though its construction is different. A schematic diagram of the system is shown in Fig. 7.22.

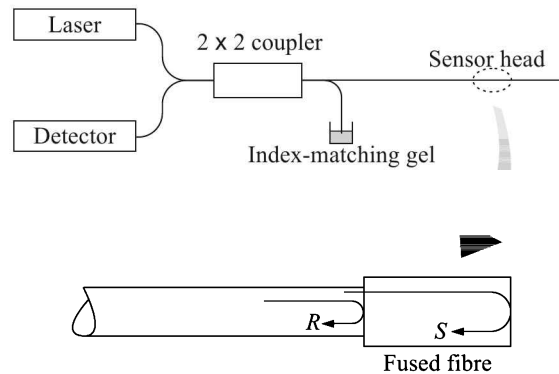


Fig. 7.22 The intrinsic Fabry-Pérot interferometric sensor system.

A laser diode is used as the optical source to one of the input arms of the bi-directional 2×2 coupler. The Fabry-Pérot cavity is formed internally fusing a small length of a single-mode fibre connected to one of the legs of the coupler. The cavity can also be constructed by introducing two Fresnel or other reflectors along the length of the single-mode fibre. Since the cavity is located within the fibre, it is called the *intrinsic interferometer*.

The reference R and sensing S reflections interfere in the single-mode fibre to generate a sinusoidal intensity variation, the wavelength of which depends on the cavity length. The cavity having been formed within the fibre, changes in the refractive index of the fibre owing to the strain can significantly alter the phase of the sensing signal S .

Like the EFPI sensor, the IFPI sensor also has a nonlinear output that complicates the measurement of strains of large magnitude. This can be overcome by operating the sensor at the linear Q-point as before.

The disadvantages of the IFPI sensor are:

1. They are highly susceptible to temperature changes and transverse strain components. In embedded applications, the sensitivity to all strain components can result in complex signal output.
2. The fabrication of IFPI sensors is more complicated than EFPI sensors since the cavity needs to be formed within the same fibre by some special procedure.
3. They suffer from drift in the output signal caused by variation in the state of polarisation of the input light.

Nevertheless, in proper conditions, the resolution of IFPI sensors is about 1 microstrain and the dynamic operating range is greater than 80 dB.

Interference-based sensors though have a simple output variation, they suffer from limited sensitivity to strain. Grating-based sensors have recently become popular because they provide wavelength-encoded output signals that can typically be demodulated to derive information about the perturbation under investigation.

Fibre Bragg Grating Sensor

A fibre Bragg grating (FBG) is a periodic (or aperiodic) perturbation of the effective refractive index in a short segment (a few millimetres or centimetres) of the core of an optical fibre [Fig. 7.23(a) and (b)]. This variation of the refractive index essentially generates a wavelength specific dielectric mirror that reflects particular wavelengths of light and transmits all others as shown in Fig. 7.23(c). Therefore, a fibre Bragg grating can be used as an inline optical notch filter¹⁵ to block certain wavelengths, or as a wavelength-specific reflector.

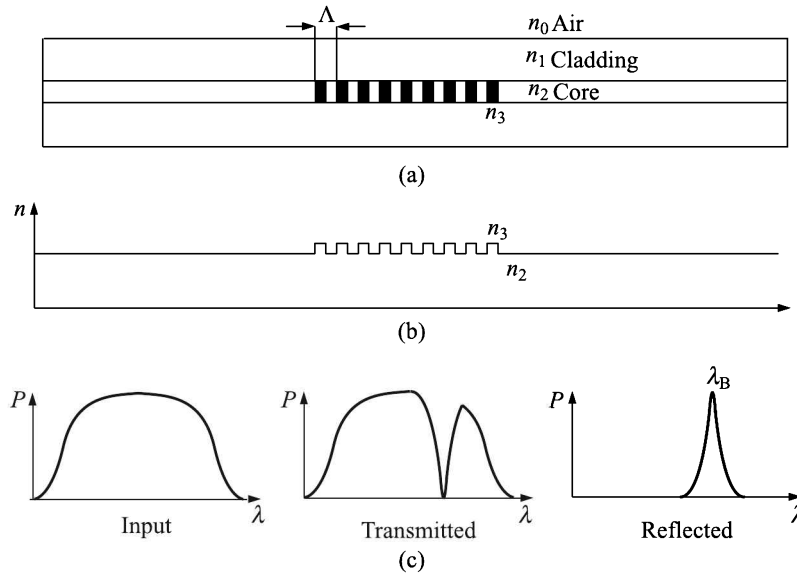


Fig. 7.23 Fibre Bragg grating: (a) structure, (b) refractive index profile, and (c) spectral response.

Theory

The perturbation in the refractive index leads to the reflection of light (back propagation along the fibre) in a narrow range of wavelengths, for which the Bragg condition¹⁶ is satisfied:

$$\frac{2\pi}{\Lambda} = 2 \cdot \frac{2\pi n_{\text{eff}}}{\lambda_B}$$

or

$$\lambda_B = 2n_{\text{eff}}\Lambda \quad (7.42)$$

where Λ is the grating period (see Fig. 7.23)

λ_B is the Bragg wavelength

n_{eff} is the effective refractive index $[= (n_3 + n_2)/2]$ in the fibre.

¹⁵See Section 16.2 at page 779.

¹⁶See Section 14.9 at page 701.

Essentially, the condition means that the wavenumber of the grating matches the difference of the (opposite) wave vectors of the incident and reflected waves. That is,

$$\mathbf{k}_i + \mathbf{k}_r = \mathbf{k}_B \equiv \frac{2\pi}{\Lambda}$$

In that case, the complex amplitudes corresponding to reflected field contributions from different parts of the grating are all in phase so that they can add up constructively. This is a kind of phase matching. Even a weak index modulation (with an amplitude of e.g. 10^{-4}) is sufficient for achieving nearly total reflection, if the grating is sufficiently long (e.g. a few millimetres). Light at other wavelengths, not satisfying the Bragg condition, is nearly not affected by the Bragg grating, except that some side lobes frequently occur in the reflection spectrum (Fig. 7.24).

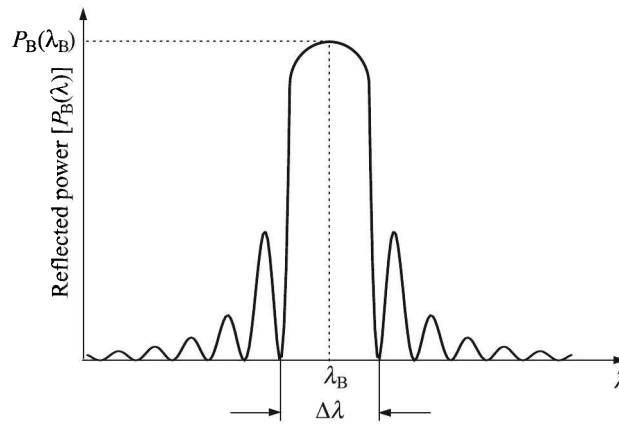


Fig. 7.24 The reflected power of an FBG vs. wavelength.

The wavelength spacing between the first minima, or the bandwidth $\Delta\lambda$ is given by,

$$\Delta\lambda = \left[\frac{2(n_3 - n_2)\eta}{\pi} \right] \lambda_B = \left[\frac{2(n_3 - n_2)\eta}{\pi} \right] \cdot 2n_{\text{eff}}\Lambda \quad (7.43)$$

where η is the fraction of power in the core.

Equation (7.43) shows that the bandwidth is proportional to the grating period Λ as well as the refractive indices. As the wavelength of maximum reflectivity depends not only on the Bragg grating period but also on temperature and mechanical strain, Bragg gratings can be used as temperature and strain sensors. Transverse stress, as generated, e.g. by squeezing a fibre grating between two flat plates, induces birefringence and thus polarisation-dependent Bragg wavelengths.

Grating structures

The refractive index and the grating period are the two parameters through which the structure of the FBG can vary (Fig. 7.25).

Grating period. The grating period can be uniform [Fig. 7.25(a)], chirped [Fig. 7.25(b)], tilted [Fig. 7.25(c)] or distributed in a superstructure [Fig. 7.25(d)].

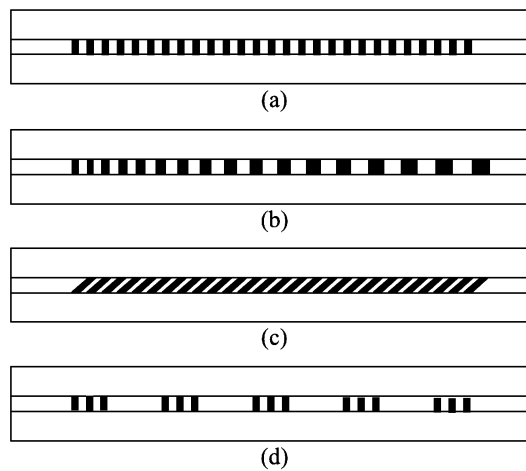


Fig. 7.25 The structure of an FBG of uniform refractive index: (a) uniform, (b) chirped, (c) tilted, and (d) superstructure.

A linear variation in the grating period is called a *chirp*. The chirp has the effect of broadening the reflected spectrum.

In a tilted FBG, the variation of the refractive index is at an angle to the optical axis. The angle of tilt has an effect on the reflected wavelength and bandwidth.

For a standard FBG, the grating period is of the same size as the Bragg wavelength as defined in Eq. (7.42). So, a grating that is required to reflect at 1200 nm is constructed with a grating period of 400 nm, using a refractive index of 1.5. However, longer periods help achieve much broader responses than those of standard FBGs. These superstructure gratings have grating periods of about 100 μm to 1 mm and are therefore much easier to construct.

Refractive index profile. The refractive index profile again can be uniform or apodised.¹⁷ In an apodised grating the refractive index graded to approach zero at the end of the grating. Apodised gratings offer significant improvement in the suppression of side-lobes while maintaining reflectivity and a narrow bandwidth. The two functions typically used to apodise an FBG are Gaussian and raised-cosine as shown in Fig. 7.26.

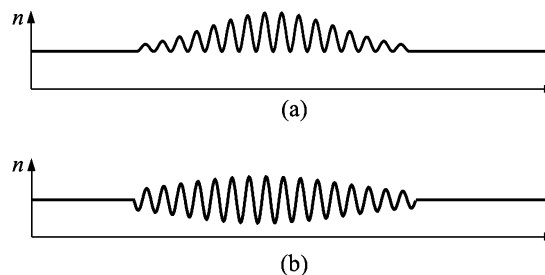


Fig. 7.26 Refractive index profile of apodised FBGs: (a) Gaussian, and (b) raised-cosine.

¹⁷ Apodise means remove or smoothen a sharp discontinuity in something.

Fabrication

The periodic variation of refractive index into the core of an optical fibre is *inscribed* or *written* using an intense ultraviolet source such as a KrF or ArF excimer¹⁸ laser or other type of UV laser to fabricate an FBG.

The photosensitivity of the core glass is utilised to write a variation in its refractive index. Silica glass has a weak photosensitivity, whereas germanosilicate glass exhibits a much stronger effect. A significant further enhancement in the photosensitivity can be achieved by loading the fibre with hydrogen (hydrogenated fibres). The amount of change of the refractive index depends on the exposure intensity and duration.

Two main processes—interference and masking—are used.

Interference process. An UV laser writing beam at 244 or 248 nm, is split into two parts of approximately the same intensity by a beam splitter. The two beams are made to interfere with each other and create an interference pattern of periodic intensity distribution. The interference pattern is focussed on a portion of the Ge-doped fibre whose protective coating has been removed. The refractive index of the photosensitive fibre changes according to the intensity of light that it is exposed to. By varying the incident angle of the writing beam, the grating period Λ can be altered. We know, the Bragg wavelength λ_B is directly related to the grating period.

Masking process. An appropriate photomask is placed between the UV light source and the photosensitive fibre. The shadow of the photomask then determines the grating structure. This method is specifically suitable for the manufacture of chirped FBGs which cannot be fabricated by using an interference pattern.

Instrumentation

The instrumentation of an FBG sensor is similar to that of an IFPI (Fig. 7.22). Instead of an IFPI sensor, the FBG sensor is used in one of the output arms of the bi-directional 2×2 coupler.

Fibre Brillouin Scattering Sensor

If a short pulse of laser beam is launched on a fibre, the light is scattered as the pulse passes down the fibre owing to its interactions with

1. Density and composition fluctuation (Rayleigh¹⁹ scattering)
2. Lattice vibration in the acoustic mode (Brillouin²⁰ scattering)
3. Molecular vibrations (Raman²¹ scattering)

¹⁸The term *excimer* is abbreviated from *excited dimer*. Typically, a combination of an inert gas (argon, krypton, or xenon) and a reactive gas (fluorine or chlorine) is used in an excimer laser. Under the appropriate conditions of electrical stimulation, a pseudo-molecule, called a dimer, is created. The dimer can exist only in an excited state and can generate laser light in the ultraviolet range.

¹⁹John William Strutt, 3rd Baron Rayleigh (1842–1919) was an English physicist. He earned the Nobel Prize for Physics in 1904. Rayleigh scattering explains why the sky is blue.

²⁰Léon Nicolas Brillouin (1889–1969) was a French physicist. His contributions to quantum mechanics, radio wave propagation in the atmosphere, solid state physics and information theory are significant.

²¹Chandrasekhara Venkata Raman (1888–1970) was an Indian physicist who was a recipient of the Nobel Prize for Physics in 1930 for the discovery of the inelastic scattering of light named after him.

The spectrum of the backscattered radiation is schematically shown in Fig. 7.27. The effects of scattering are classified by the relation between wavelengths of the incident and scattered radiations.

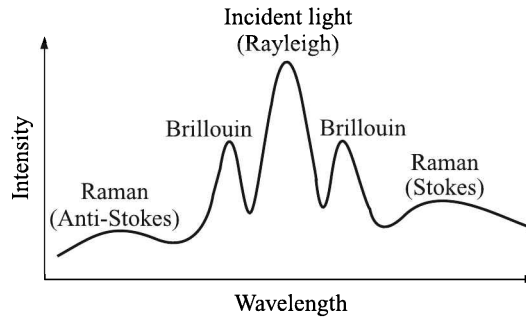


Fig. 7.27 Spectrum of backscattered light after pulsed illumination with a laser light source.

If these frequencies or wavelengths are equal, the phenomenon is called *unshifted scattering*, i.e. Rayleigh or elastic scattering. But if these frequencies differ, the term *shifted or inelastic scattering* is used. Examples are the Raman and Brillouin scattering.

The counterpropagating Brillouin scattering wave drains energy from the forward-moving input pulse. To satisfy the requirements of energy conservation, there occurs a frequency shift between the incident pulse frequency and the Brillouin scattering wave, which, in general, is on the order of tens of GHz. Since the frequency shift of a Brillouin gain spectrum is sensitive to strain (as well as temperature), it has been found to be useful in the construction of fibre-optic sensors for measuring the strain.

The magnitude of the longitudinal strain, in particular, affects the Brillouin frequency shift. This is because, the acoustic wave frequency induced by the incident photon is different under different strain conditions. Thus, the longitudinal strain distribution over a length can be measured from the Brillouin scattering effect. We may note here that it is difficult to measure the longitudinal strain distribution by other techniques.

Figure 7.28 shows the typical set-up of a Brillouin fibre-optic time-domain reflectometry.

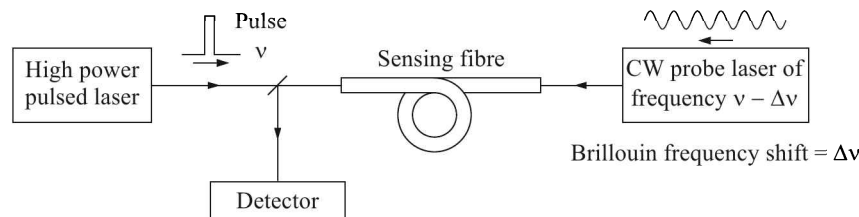


Fig. 7.28 Schematic set-up of optical time-domain reflectometry based on Brillouin scattering.

We have already stated that in practice, the Brillouin peaks are separated from the launch wavelength by only a few tens of GHz. Therefore, it is rather difficult to separate these components from the Rayleigh signal. A special technique, called *coherent detection technique* is resorted to for this purpose.

For coherent detection, the fibre is interrogated from one end by a powerful pulsed source and from the other end by a continuous-wave (CW) source. The frequency difference between the two radiations is adjusted to be equal to the Brillouin frequency shift. The subtracted frequency is amplified to retrieve the Brillouin peak. A typical spectrum, obtainable by this technique, is shown schematically in Fig 7.29.

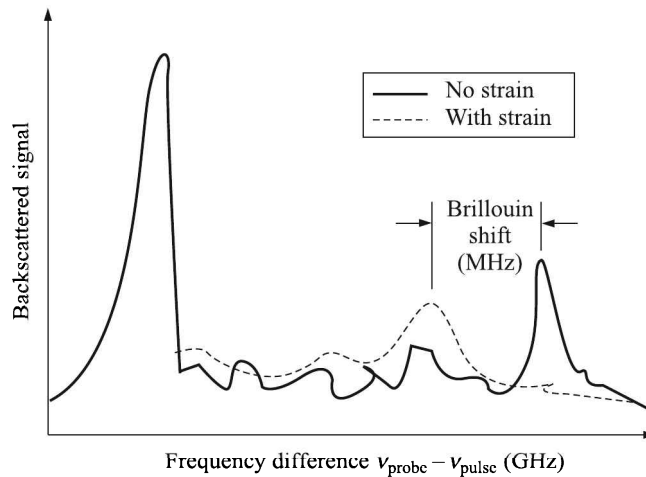


Fig. 7.29 A typical Brillouin scattering shift spectrum under different longitudinal strain conditions.

The solid curve in Fig. 7.29 represents no longitudinal strain while the dotted curve represents a longitudinal strain. The Brillouin shift is normally about a few hundred MHz which the current technology allows us to detect. However, a very stable laser frequency has to be used for the detection to be successful. Tunable diode lasers or Q-switched Nd:YAG lasers offer that stability, power and narrow bandwidth.

Advantages of Fibre-optic Strain Gauges

Fibre-optic strain gauges offer many advantages over the classical resistance strain gauges. Some of them are:

1. A single gauge can cover large distances as is necessary in distributed measurements.
2. A single gauge can monitor strain gradient over a sample.
3. They are free from electromagnetic interference (EMI) or radiofrequency interference (RFI) disturbances.
4. They can be used in explosive or hazardous environments because no electric current flows through them.
5. They are light-weight structures and amenable to remote measurements.

Although the sensors themselves are inexpensive, the instrumentation is costly.

Review Questions

- 7.1 (a) Define gauge factor (G_f) for a strain gauge. Compare some of the important characteristics of metallic and semiconductor type strain gauges.
- (b) Explain with a circuit diagram the principle of operation of a strain measurement system having arrangement for temperature compensation.
- (c) In an equal-arm Wheatstone bridge, the single active gauge has nominal resistance 120 ohms, and is made of 'Advance' having a thermal coefficient of expansion of $30 \times 10^{-6} \text{ m/m } ^\circ\text{C}$ and temperature coefficient of resistance of $12 \times 10^{-6} \text{ ohm/ohm } ^\circ\text{C}$. The other three arms are fixed resistors having negligible temperature coefficients. If the gauge is bonded to steel having a thermal coefficient of expansion of $13 \times 10^{-6} \text{ m/m } ^\circ\text{C}$, calculate the bridge output for a 60°C rise in specimen temperature, if the gauge current is 25 mA.
- 7.2 Discuss the principle of operation of the strain gauge. What is the gauge factor? A resistance strain gauge with a gauge factor of 2 is fastened to a steel member which is subjected to a strain of 1×10^{-6} . If the original resistance value of the gauge is 130Ω , calculate the change in resistance.
- 7.3 (a) Describe, in brief, two types of wire-wound strain gauges, mentioning their typical size, resistance, maximum excitation voltage, and construction material.
- (b) Show, from theoretical considerations, why the gauge factor G_f for most gauges is nearly 2.
- (c) The resistance of a strain gauge is 120Ω and $G_f = 2$. It is connected to a current-sensitive Wheatstone bridge in which resistances on all arms are 120Ω . If the input voltage is 4 V and the resistance of the galvanometer is 100Ω , calculate the detector current in μA for 1 μ -strain.
- 7.4 (a) Describe, in brief, a piezoresistive-type strain gauge, mentioning its merits and demerits.
- (b) Name two materials which are used to construct piezoelectric transducers.
- 7.5 (a) Explain in detail the constructional features and the theory of operation of wire-type strain gauge.
- (b) Name some commonly used materials for wire/foil gauge and associated bonding techniques.
- (c) A strain gauge of 120Ω nominal resistance is fixed on a structural member subjected to a strain of $50 \mu\text{m/m}$. If the gauge factor is 2.5, what is the change in resistance of the gauge?
- 7.6 (a) Define the gauge factor of a metallic strain gauge.
- (b) A strain gauge with nominal resistance of 200Ω and $G_f = 2.0$ is fixed on one flat surface of a short column of $2 \text{ cm} \times 2 \text{ cm}$ cross-sectional area. The column is subjected to an axial force of 100 N. The strain gauge forms one arm of a bridge with other arms all equal to 200Ω . Find the open-circuit output of the bridge excited by 10 V. Given, Young's modulus of elasticity = $2.1 \times 10^{11} \text{ N/m}^2$.

- 7.7 (a) What is meant by gauge factor? Derive the expression for the gauge factor.
 (b) What do you mean by dummy gauge?
 (c) A resistance wire strain gauge having nominal resistance of $350\ \Omega$ is subjected to strain of 500 microstrain. Find the change in the value of resistance, neglecting the piezoelectric effect.
 (d) What are the advantages of semiconductor strain gauges over metallic strain gauge?
- 7.8 (a) What is a strain gauge? Give the classification of strain gauge.
 (b) Deduce the expression for the gauge factor of a strain gauge $G = 1 + 2\mu + \frac{\Delta\rho/\rho}{\Delta l/l}$, where G is the gauge factor, μ is the Poisson's ratio and $\frac{\Delta\rho/\rho}{\Delta l/l}$ is the change in resistance due to piezo-resistive change.
 (c) A strain gauge with a gauge factor of 2 and fastened to a metallic member is subjected to a stress of $1000\ \text{kg/cm}^2$. The modulus of elasticity of the metal is $2 \times 10^6\ \text{kg/cm}^2$. Calculate the % change in the resistance of the strain gauge. What is the value of the Poisson's ratio?
 (d) Write down the working principle of semiconductor used to measure strain.
- 7.9 Consider the quarter bridge strain measuring circuit shown in Fig. 7.30. Voltage e_o is amplified by a differential amplifier with a gain of 2500 and a CMRR of 80 dB. Find the actual and the indicated strain when the output V_o is 4.5 V. Given, $G_f = 2.0$, $R = 200\ \Omega$, $E = 12\ \text{V}$.

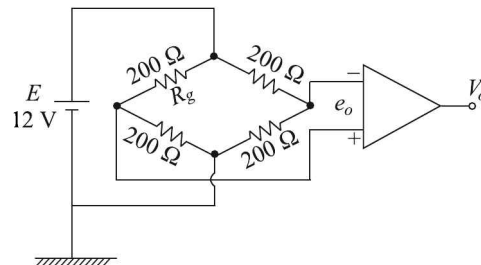


Fig. 7.30

- 7.10 Two strain gauges of resistance values $120\ \Omega$ each and gauge factor $G = 2.0$ are mounted on a cantilever beam as shown in Fig. 7.31. The gauges are connected to a bridge which uses two more fixed resistances of $120\ \Omega$ each, and the excitation voltage is $E = 1.0\ \text{V}$.

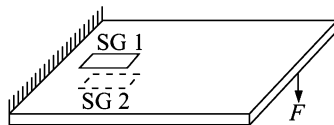


Fig. 7.31

- (a) Show the bridge arrangement for perfect temperature compensation.
- (b) The unbalanced voltage of the bridge is fed to a digital voltmeter having a resolution of $1 \mu\text{V}$. Find the minimum microstrain it can detect.
- 7.11 Two identical strain gauges of resistance 120Ω and gauge factor $+2.0$ each, are attached to a steel block of Poisson's ratio $\nu = 0.3$ as shown in Fig. 7.32. Find the unbalance voltage per unit microstrain applied to the steelblock as shown.

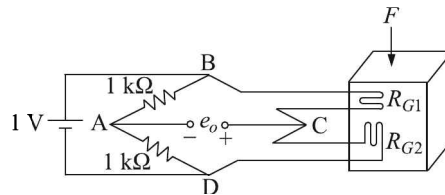
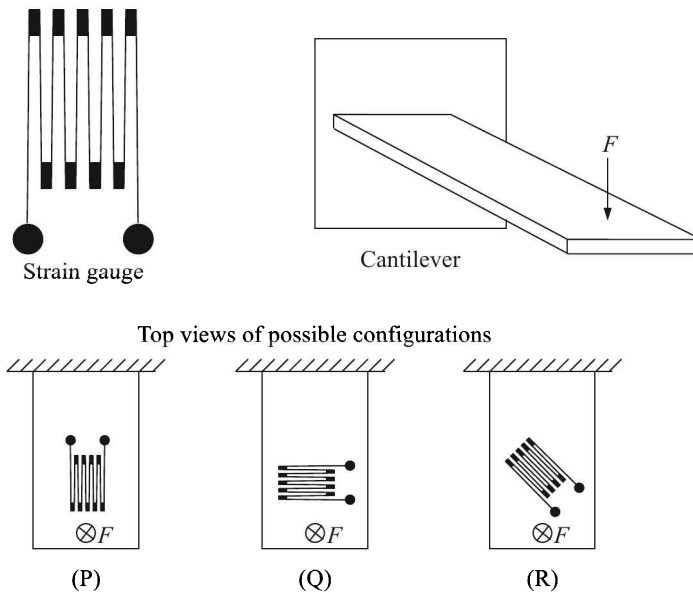


Fig. 7.32

- 7.12 Indicate the correct choice:
- (a) A semiconductor strain gauge
- has a much higher gauge factor than that of a metal wire gauge
 - employs piezoelectric property of undoped silicon
 - does not require temperature compensation
 - exhibits very little gauge factor variation as compared to that of metal wire gauges
- (b) A unity ratio quarter bridge strain measuring circuit produces an output of 1 mV for a strain of 500 microstrain when the bridge excitation is 4 volts. The gauge factor of the element is then
- 1
 - 2
 - 3
 - 4
- (c) A strain gauge has a gauge factor $G = -100$. The type of the strain gauge is
- Unbonded metal type
 - Bonded metal foil type
 - p -type semiconductor
 - n -type semiconductor
- (d) Among the bonded metal strain gauges, the foil type is more popular than the wire type because
- Error due to transverse strain is much less in foil type
 - Gauge factor is much higher in foil type
 - The foil type is much more insensitive to temperature variation
- (e) A strain gauge is attached to a bar of 20 cm which is subjected to a tensile force. The nominal resistance of strain gauge is 100Ω . The changes in resistance and elongation in the bar measured are 0.35Ω and 0.2 mm respectively. The gauge factor of the strain gauge is

- (i) 2
(ii) 3.5
(iii) 10
(iv) 100
- (f) All metal resistive strain gauges have a gauge factor (GF) nearly 2.5 because
- Young's modulus is the same for all metals and alloys
 - Poisson's ratio is the same for all metals and alloys
 - The conductivity of the metals changes with applied strain in the elastic region in the same way
 - The conductivity of the material is independent of the applied strain
- (g) A strain gauge of resistance $120\ \Omega$ and gauge factor 2.0 is at zero strain condition. A $200\ \text{k}\Omega$ fixed resistance is connected in parallel with it. Then the combination will represent an equivalent strain of
- $+5290\ \mu\text{m}/\text{m}$
 - zero
 - $-123.8\ \mu\text{m}/\text{m}$
 - $-300\ \mu\text{m}/\text{m}$
- (h) The figure below shows various configurations of bonding a strain gauge to a cantilever subjected to a bending force F .



Which configuration gives the maximum change in resistance for this force?

- P
- Q
- R
- All have equal change in resistance

-
- (i) A strain gauge has a nominal resistance of $600\ \Omega$ and a gauge factor of 2.5. The strain gauge is connected to a dc bridge with three other resistances of $600\ \Omega$ each. The bridge is excited by a 4 V battery. If the strain gauge is subjected to a strain of $100\ \mu\text{m/m}$, the magnitude of the bridge output will be
- (i) 0 V
 - (ii) $250\ \mu\text{V}$
 - (iii) $500\ \mu\text{V}$
 - (iv) $750\ \mu\text{V}$

Pressure Measurement

Pressure is easily converted to force by allowing it to act on an area. Therefore, pressure measurement essentially reduces to force measurement, except for the high vacuum region where some special methods not related to force measurement are necessary. A force is measured by balancing it with a known or calibrated force. Consider the measurement of weight of a substance, which is the force exerted on it by the earth's attraction. While measuring it by a spring balance, the opposing force exerted by the spring balances the weight. The amount of elastic force generated within the spring depends on the displacement of its, say, end point. The two opposing forces balance, and therefore the displacement constitutes a measure of the weight. In a common balance, the downward force acting on the mass is directly balanced by counterpoising it with an equal mass on the other scale pan. The same principle is applied to pressure measurement which is based on

1. Comparison with known dead-weights acting on known areas, or
2. Deflection of elastic elements subjected to unknown pressure.

Manometers and piston (or dead-weight) gauges are examples of the first kind while elastic deflection devices assume many forms. These elastic devices which convert applied pressures to forces and then to displacements are collectively called *force summing devices*.

But before we discuss how pressure is measured by different methods, we need to consider a few definitions as well as different units of pressure.

8.1 Definitions

While specifying pressures, three terms are used, namely

1. Absolute pressure
2. Gauge pressure
3. Vacuum pressure

We explain below in what connotations the terms are used.

Absolute Pressure

Absolute pressure is the pressure that includes the atmospheric pressure which is acting on all substances unless they are placed in evacuated enclosures. The quantity p in Eq. (8.1) corresponds to the absolute pressure acting on the piston of the DWG.

In those countries that continue to use FPS units, an 'a' is added to the unit descriptor to indicate the absolute pressure, the abbreviation for pounds *per square inch (absolute)* being *psia*.

The difference between the absolute and gauge pressures will be clear from Fig. 8.1.

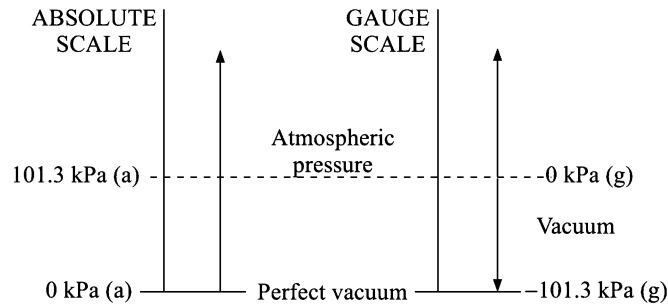


Fig. 8.1 Absolute and gauge pressure scales.

Gauge Pressure

The quantity $(p - p_A)$ in Eq. (8.2) is called the *gauge pressure*. Clearly, the gauge pressure is the pressure that does not include the atmospheric pressure and therefore, is lower than the actual fluid pressure.

Since the atmospheric pressure varies from place to place and also, from time to time, dial type pressure gauges are usually calibrated to indicate the gauge pressure rather than the absolute pressure. The absolute pressure can be obtained by adding the local atmospheric pressure to the indicated gauge pressure.

The gauge pressure is indicated by adding a 'g' to the unit descriptor. Therefore, the pressure unit *pounds per square inch (gauge)* is abbreviated as *psig*. When using SI units, it is customary to add 'gauge' to the units used, such as 'kPa gauge' to indicate gauge pressure.

The relation between the two kinds of pressure values is summed up in Table 8.1.

Table 8.1 Difference between absolute pressure and gauge pressure

| <i>Absolute pressure</i> | <i>Gauge pressure</i> |
|---|--|
| 1. Reference point is absolute vacuum | 1. Reference point is atmospheric pressure |
| 2. Indicated by an affix 'a', such as kPa (a) | 2. Indicated by an affix 'g', such as kPa (g) |
| 3. Value at mean sea level is 101.3 kPa (a) | 3. Value at mean sea level is 0 kPa (g) |
| 4. Always has a positive value | 4. Value below atmospheric pressure is negative and indicates a vacuum condition |
| 5. Equals (gauge pressure + atmospheric pressure) | 5. Equals (absolute pressure - atmospheric pressure) |

Example 8.1

Calculate the gauge pressure and absolute pressure in kg/cm^2 at the depth of 20 m in a water tank.

Solution

Assuming that the density ρ of water is 1000 kg/m^3 , the gauge pressure is

$$p_{\text{gauge}} = h\rho = (25)(1000) \text{ kg/m}^2 = 2.5 \text{ kg/cm}^2$$

And, the absolute pressure is

$$p_{\text{abs}} = p_{\text{gauge}} + p_{\text{atmos}} = 2.5 + 1.03 = 3.53 \text{ kg/cm}^2$$

Vacuum Pressure

The pressure corresponding to that lower than the atmospheric pressure is often expressed as a *vacuum of* so many Torr¹. Example 8.2 will explain what is meant by that.

Example 8.2

If the measurement indicates a vacuum of 150 Torr, what are the gauge and absolute pressures?

Solution

The vacuum of 150 Torr actually indicates 150 Torr below atmospheric pressure. So,

$$\begin{aligned} p_{\text{gauge}} &= -150 \text{ Torr} \\ p_{\text{abs}} &= p_{\text{atmos}} - 150 = 760 - 150 = 610 \text{ Torr} \end{aligned}$$

Note: The specification is *vacuum of* 150 Torr, not *pressure of* 150 Torr. A vacuum is thus always a *negative* quantity if expressed as a gauge pressure.

8.2 Pressure Units and Their Conversions

The basic SI unit of pressure is pascal (Pa) which is defined as a force of 1 newton acting on area of 1 m^2 , or

$$1 \text{ Pa} = 1 \text{ N/m}^2$$

Since pascal is a small unit, it is customary to use kPa or MPa. Quite a number of other units are used to report pressure readings. A conversion table helps to interrelate them. Table 8.2 lists such conversions.

Example 8.3

Convert the pressure of 20 kg/cm^2 to (a) mmWG, (b) mmHg and (c) bar units.

Solution

Assuming that the given pressure is gauge pressure,

(a) We know that $1 \text{ kg/cm}^2 = 10000 \text{ mmWG}$. Therefore,

$$20 \text{ kg/cm}^2 = 2 \times 10^5 \text{ mmWG}$$

¹The pressure exerted by a 1 mm column of Hg.

Table 8.2 Conversion factors for pressures in various systems of units

| | Pa | mbar | Torr | at ^a | atm ^b | psi |
|----------------------------|--------------------|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Pa | 1 | 1×10^{-2} | 7.5×10^{-3} | 1.02×10^{-5} | 9.87×10^{-6} | 14.5×10^{-5} |
| mbar | 1.0×10^2 | 1 | 7.5×10^{-1} | 1.02×10^{-3} | 9.87×10^{-4} | 14.5×10^{-3} |
| Torr | 1.33×10^2 | 1.33 | 1 | 1.36×10^{-3} | 1.32×10^{-3} | 19.3×10^{-3} |
| at ^a | 9.80×10^4 | 9.80×10^2 | 7.36×10^2 | 1 | 9.68×10^{-1} | 14.5 |
| atm ^b | 1.01×10^5 | 1.01×10^2 | 7.60×10^2 | 1.03 | 1 | 14.7 |
| psi | 6.8×10^3 | 0.68 | 51.8 | 6.9×10^{-2} | 6.8×10^{-2} | 1 |
| mmWG/ mmWC ^c | 9.8 | 9.8×10^{-2} | 7.35×10^{-2} | 10^{-4} | 9.67×10^{-5} | 1.42×10^{-3} |

^a Technical atmosphere

^b Physical atmosphere

^c Full forms are *mm water gauge* or *mm water column*. 1 mmWG or mmWC is the pressure required to support a 1 mm column of water at normal temperature.

(b) 1 mmHg = 13.546 mmWG. Therefore,

$$\begin{aligned} 20 \text{ kg/cm}^2 &= 2 \times 10^5 \text{ mmWG} = \frac{2 \times 10^5}{13.546} \\ &= 1.48 \times 10^4 \text{ mmHg} \end{aligned}$$

(c) 1 bar = 1.03 kg/cm². Therefore,

$$20 \text{ kg/cm}^2 = \frac{20}{1.03} = 19.42 \text{ bar}$$

With this background information, we now move on to the methods of pressure measurement.

8.3 Comparison with Known Dead-weights

Dead-weight Gauge

Direct pressure measurement is rarely done with dead-weight gauges; they are mainly used to calibrate direct reading dial-type pressure gauges. Hence, we will discuss this gauge from that perspective.

The gauge consists of an accurately machined hollow cylinder and a close-fitting piston with a top platform, and another similar cylinder connected to the former through a reservoir (Fig. 8.2). The cylinders and piston are honed to micron tolerances and their cross-sectional areas are accurately known. The gauge to be calibrated is connected to the second cylinder, while on the platform of the former accurately known weights in the form of discs can be placed. By the piston arrangement shown at the bottom, the fluid pressure is gradually increased until the piston-weight arrangement just freely floats.

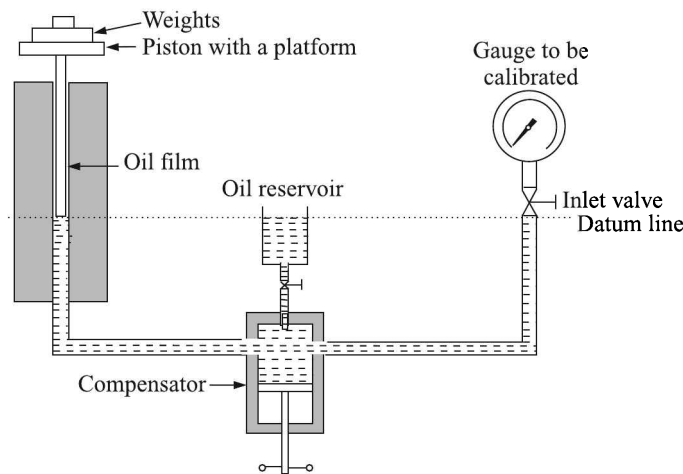


Fig. 8.2 Calibration of a dial-type gauge by the deadweight gauge.

If p is the fluid pressure

p_A is the atmospheric pressure acting on the weights

M_t is the tare (or mass of the piston assembly)

M_c is the extra mass placed on top of the piston to calibrate pressure

A is the equivalent area of the piston cylinder corrected for friction, pressure level and temperature

g is the acceleration due to gravity

then

$$p = p_A + \frac{(M_t + M_c)g}{A} \quad (8.1)$$

$$p - p_A = \frac{(M_t + M_c)g}{A} \quad (8.2)$$

Equation (8.2) assumes that g is acting vertically downwards, which demands a strict verticality of the piston and the cylinder. The fluid film between the cylinder and the piston generates a frictional force which can be reduced by rotating or vibrating the piston, because dynamic friction is always less than the static friction.

Dead-weight gauges are pretty accurate, the maximum attainable accuracy being about 0.01%. M_t in Eq. (8.2) is the tare (or weight of the piston assembly). Obviously, the minimum pressure that such a gauge can measure is $M_t g/A$.

Example 8.4

A pressure gauge in the range of 0 to 100 kg/cm² requires to be calibrated against a dead-weight gauge. The tare of the DWG is 5 kg and the cross-sectional area of the base region of the piston is 7.5 cm². If the calibration is to be checked in steps of 10 kg/cm², design a suitable set of standard weights. Assume, the frictional forces are negligible.

Solution

For the 10 kg/cm² calibration, the weight required to be put according to Eq. (8.2) is

$$M_c = (p - p_A)A - M_t = (10)(7.5) - 5 = 70 \text{ kg}$$

For the 20 kg/cm² calibration, the weight required to be put is

$$M_c = (p - p_A)A - M_t = (20)(7.5) - 5 = 145 \text{ kg}$$

Similarly, for 30 and 40 kg/cm² calibration we need 220 and 295 kg weights. A set of 10, 20, 30 and 40 kg/cm² pressure standards is sufficient to generate pressures from 50–100 kg/cm².

The zero of the pressure gauge cannot be calibrated by the DWG because then we need a –5 kg weight! It should be marked at the place where the gauge pointer rests after dismounting it from the DWG.

- Note:*
1. The calibration weights can be made lighter by choosing a piston of smaller cross-sectional area. But that will be done at the expense of the accuracy of measurement, because a thin piston will be vulnerable to deformation.
 2. The calibrated pressures will be (see Section 8.1 at page 281) rather than absolute pressures because we have used $(p - p_A)$ in our calculation.

Manometers

Unlike dead-weight gauges, manometers are self-balancing, deflection-type rather than null-type instruments, and they have a continuous rather than stepwise readout. But their disadvantages are they are large, cumbersome, and not well suited for integration into automatic control loops. Therefore, manometers are usually found in the laboratory or used as local indicators. Depending on the reference pressure used, they can indicate absolute, gauge, and differential pressures.

Two useful variations of the form of the manometer are shown in Fig. 8.3. The most important advantage of well-type manometers is that here the reading of a single leg gives the desired pressure reading. A vernier scale can be used here to determine the mercury column height more accurately.

The barometric type, where the end of the reading tube is evacuated and sealed, gives not only absolute pressure readings, but also the arrangement of height adjustment of the mercury level of the well to the zero mark eliminates the small error caused by the movement of mercury out of the well to the reading limb.

Manometers are of comparable accuracy to deadweight gauges at lower pressures. But higher pressure measurements with manometer are impractical because of the length of liquid columns involved.

Characteristics of manometer fluid

1. Manometer fluid should not wet the wall of the container.
2. It should not absorb gas or chemically react with it.
3. It should be of reasonably high density so that the pressure balancing column stays within a desirable limit.

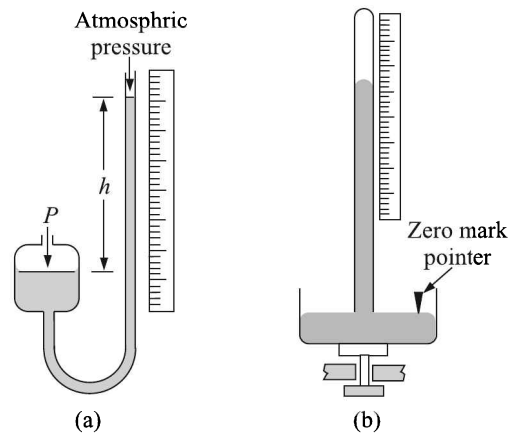


Fig. 8.3 Manometers of different kinds: (a) well-type manometer, and (b) barometer.

4. It should have low vapour pressure at the operating temperature.
5. It should freely move in the limbs of the manometer.
6. It should not be compressible.

Advantages of manometers

1. Manometers are considered as standards of pressure by many standardising institutions.
2. They are simple and low cost devices.
3. Suitable sensors such as capacitance or sonar devices can be used to provide better precision in their readouts.

Sources of error in using manometers

1. Surface tension of manometer fluids, causing capillary effects, may give wrong readings.
2. Thermal expansion of the fluid as well as of the readout scale may cause error.
3. Compressible fluid may change calibration.
4. Evaporated fluid at low pressure and high temperature can interfere in measurement.

Inclined-limb manometer

If one of the limbs of a manometer is inclined at an angle θ , its sensitivity increases as can be seen from the following analysis. In such an arrangement, with no pressure difference Δp applied between the limbs of the manometer, the meniscus in each limb is at the same level A. If now a pressure difference Δp is applied, the tube meniscus goes up by a height h_t while the well meniscus goes down by h_w as shown in Fig. 8.4.

Thus,

$$\Delta p = (h_t + h_w)\rho g \quad (8.3)$$

where, ρ is the density of the manometric fluid and g is the acceleration due to gravity.

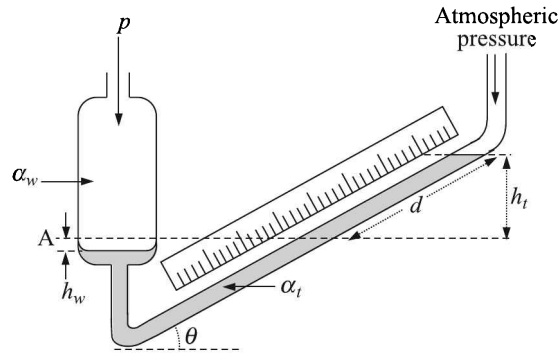


Fig. 8.4 Inclined-limb manometer.

Now, if α_t and α_w are cross-sectional areas of the tube and well respectively, and d is the distance moved by the meniscus in the tube, we get from the continuity of volume,

$$\alpha_w h_w = \alpha_t d = \alpha_t \frac{h_t}{\sin \theta} \quad [\because h_t = d \sin \theta]$$

Thus,

$$h_t + h_w = \left(1 + \frac{\alpha_t}{\alpha_w \sin \theta}\right) h_t = \left(\sin \theta + \frac{\alpha_t}{\alpha_w}\right) d \quad (8.4)$$

From Eqs. (8.3) and (8.4),

$$\begin{aligned} \Delta p &= d \left(\sin \theta + \frac{\alpha_t}{\alpha_w}\right) \rho g & (8.5) \\ &\cong d \rho g \sin \theta & [\text{since } \alpha_w \gg \alpha_t] \\ &= h_t \rho g \end{aligned}$$

Thus, the pressure difference can be measured by measuring either h_t or d . Since, $h_t = d \sin \theta$, if θ is 30° , $d = 2h_t$ for the same Δp . Which is why, measuring d will increase the sensitivity of the instrument.

Example 8.5

A manometer uses transformer oil of specific gravity 0.864 as measuring liquid. The scale is graduated in mm of water. If one leg is a 2 mm bore tube and the other a 20 mm well, calculate the angle to the horizontal at which the tube and scale must be inclined to give 4 mm scale deflection for a pressure of 1 mm head of water. Assume, 1 mm of water = 9.81 Pa.

Solution

$$\begin{aligned} \text{Here} \quad \Delta p &= 9.81 \text{ Pa} & d_t &= 2 \text{ mm} = 0.002 \text{ m} \\ d_w &= 20 \text{ mm} = 0.02 \text{ m} & d &= 4 \text{ mm} = 0.004 \text{ m} \\ \text{Specific gravity} &= 0.864 & \Rightarrow \rho &= 0.864 \times 1000 \text{ kg/m}^3 \end{aligned}$$

From Eq. (8.5)

$$\begin{aligned}\sin \theta &= \frac{\Delta p}{d\rho g} - \left(\frac{d_t}{d_w}\right)^2 \\ &= \frac{9.81}{(0.004)(0.864 \times 1000)(9.81)} - (0.1)^2 \\ &= 0.2935\end{aligned}$$

or
$$\theta = \sin^{-1}(0.2935) = 16.22^\circ$$

Example 8.6

The ratio of cross-sectional area between the limb and the well of a well-type mercury manometer is 1:20. If it is used to measure the pressure drop caused by a water flow across an orifice, what will be the error in the measurement if the normal practice of reading the limb value is followed?

Solution

Since the volume of mercury pushed down in the well will equal that pushed up in the limb, If α_w and α_l are the cross-sectional areas of the well and limb, and h_w and h_l are the change of heights of mercury columns in the well and the limb respectively, we have

$$\alpha_w h_w = \alpha_l h_l$$

or
$$h_w = \frac{\alpha_l}{\alpha_w} h_l = \frac{h_l}{20}$$

Thus, the actual pressure drop Δp is given by

$$\Delta p = h_l + h_w = h_l \left(1 + \frac{1}{20}\right)$$

If only h_l is read, instead of measuring Δp , $\Delta p/[1 + (1/20)]$ will be measured. Hence, the per cent error ε will be

$$\varepsilon = \left(\Delta p - \frac{\Delta p}{1 + (1/20)}\right) \times \frac{100}{\Delta p} = 4.76\%$$

Ring-balance manometer

The ring-balance manometer consists of a hollow ring, half-filled with a manometric fluid. Along the horizontal diameter of the ring, a bar is fixed on a central knife-edge that rests on a rigid surface (see Fig. 8.5). A pointer and a counter-weight are fixed perpendicular to the bar. The ring has two pressure inlets at the top. When pressures at two inlets are equal, the ring experiences no torque and therefore the pointer is horizontal, indicating a zero pressure differential. In case, $p_1 > p_2$, the ring tilts in one direction owing to the torque it experiences and the pointer indicates a differential pressure. At this position, the torque applied by the pressure differential is balanced by the counter torque generated by the displaced counter-weight.

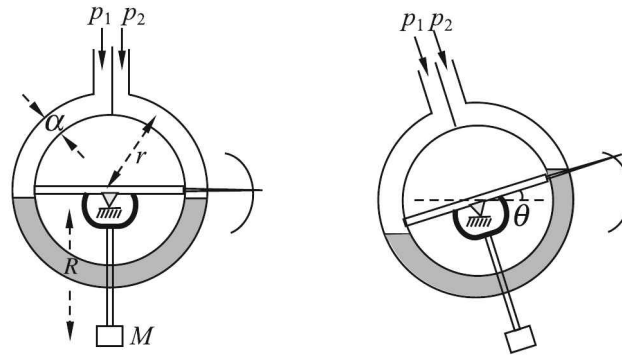


Fig. 8.5 Ring-balance manometer.

Thus,

$$(\Delta p \alpha) r = MgR \sin \theta$$

where M is the mass of the counter-weight

R is the radius of movement of the counter-weight

Δp is the pressure differential

α is the area of cross-section of the ring

θ is the angle of deflection, and

r is the radius of the ring

Therefore,

$$\Delta p = \frac{MgR \sin \theta}{\alpha r} \cong \frac{MgR}{\alpha r} \theta \quad [\text{for small } \theta]$$

Since all factors, other than the angle of deflection, are constant for the manometer,

$$\Delta p = K \theta$$

where K is a constant. It may be noted that the deflection for a given pressure differential will remain the same whatever be the manometric fluid. But, care should be taken not to apply a large enough pressure differential that will tilt the ring in such a way that the fluid on one side reaches the bottom of the ring. In that case, the gas on one side will bubble through the fluid to reach the other side thus disturbing not only the pressure differential but also mixing the two gases.

It is apparent from the construction of the manometer that the hose connected to the two inlets should be flexible enough to allow the ring to rotate rather freely. Such hoses are usually unable to withstand high pressures. So, only differentials between moderate pressures can be measured by this manometer. Depending on the design, the minimum and maximum spans of these manometers are 50 Pa and 125 kPa respectively. At higher pressures, the flexible leads become a source of error and require maintenance.

Inverted bell manometer

In an inverted bell manometer, the bell is placed upside down on top of a tube conveying the lower pressure and a sealing liquid separates the high and low pressure sides as shown in

Fig. 8.6. The inverted bell is loaded by a spring on top. Owing to the pressure differential, the inverted bell makes a translational movement which can be converted to a pointer movement by a mechanical arrangement. Alternatively, an electrical signal can be generated through a variable reluctance pick-up which can be fixed in the spring area.

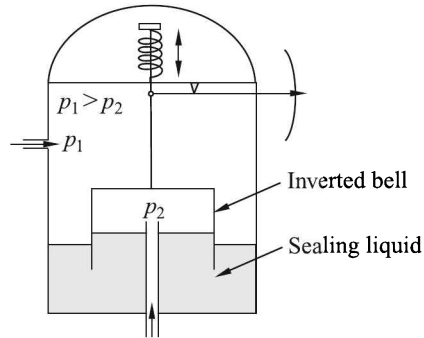


Fig. 8.6 Inverted bell manometer.

If the low pressure side is connected to a vacuum line with a suitable sealing liquid, the inverted bell manometer will measure absolute pressure. It is easily seen that the displacement of the bell varies linearly with the differential pressure. Because, the upward force F_u acting on the bell is

$$F_u = (p_1 - p_2)\alpha$$

where, α is the area of cross-section of the bell. The downward force F_d offered by the spring is

$$F_d = k\Delta x$$

where k is the spring constant and Δx is the displacement of the bell. At equilibrium,

$$k\Delta x = (p_1 - p_2)\alpha$$

$$\Rightarrow \Delta x = \frac{\alpha}{k}(p_1 - p_2) \quad (8.6)$$

Now, if l is the length of the pointer and θ is the angular deflection of the pointer,

$$\Delta x = l\theta$$

$$\Rightarrow \theta = \frac{\alpha}{kl}(p_1 - p_2) \quad (8.7)$$

With a suitable design, the instrument can measure pressure differential up to 38 cm of water (~ 3.72 kPa). A double bell-type manometer (Fig. 8.7) is sometimes used to achieve a high degree of precision in differential pressure measurement. It can be successfully used in measuring Δp in orifice flow meters².

Here the bell is suitably shaped so that $\Delta x \propto \sqrt{\Delta p}$. That eventually gives a linear relationship between volume flow rate Q and meter deflection³.

²See Section 11.2 at page 449.

³See Eq. (11.5) at page 448.

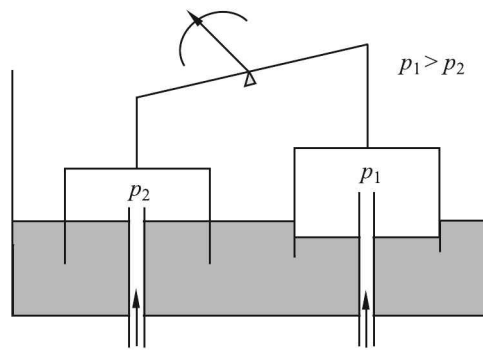


Fig. 8.7 Double bell-type manometer.

8.4 Force-summing Devices

Force-summing devices, which convert the applied pressures to displacements, are primary transducers while generated displacements may be measured by secondary transducers. We first discuss the primary transducers, i.e. force-summing devices.

Commonly used force-summing devices are:

1. Diaphragms
2. Bellows
3. Bourdon gauge

These are discussed below.

Diaphragms

The diaphragm elements are made of circular metal discs or flexible elements such as rubber, plastic or leather. The material the diaphragm is chosen in such a way that its elasticity can be utilised directly or as a transmitter of pressure to an opposing element such as a spring. Diaphragms made of metal discs utilise the elastic characteristics, while those made of flexible elements are transmitters of pressure to another elastic element.

Different kinds of diaphragms are shown in Fig. 8.8.

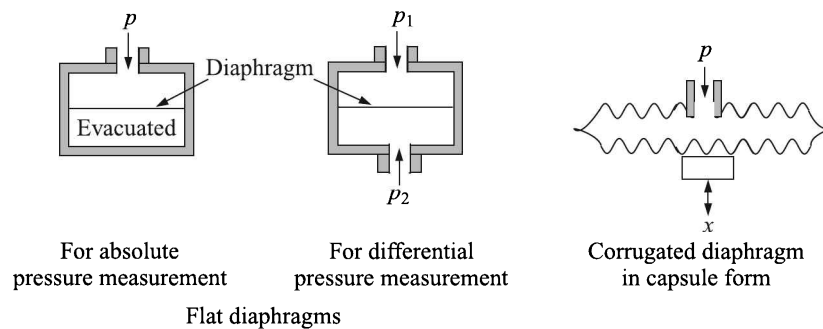


Fig. 8.8 Diaphragms of different kinds.

Corrugated diaphragms give a larger displacement and two of them can be conveniently soldered together at the outer edges to form capsules. For a given pressure, a capsule generates a displacement which is twice as large as that of a single diaphragm.

Theoretical relations between displacement of diaphragms and pressure have been derived. They show that if d is the thickness of the diaphragm and x is its deflection owing to the application of a pressure Δp

$$x \propto \Delta p \quad \text{for } x < \frac{d}{3}$$

Therefore, one may experimentally calibrate the displacement against the pressure.

The diaphragm sensors are very sensitive (0.01 MPa) to rapid pressure changes. The metal type can help measure pressures up to 7 MPa, while the elastic type can have ranges from 0–0.1 kPa to 0–2.2 MPa when connected to capacitive transducers or differential pressure sensors. They are also very versatile and are commonly used in very corrosive environments or extreme over-pressure situations.

Example 8.7

The diaphragm element of a pressure gauge is a circular foil of steel (Young's modulus $E = 2 \times 10^{11} \text{ N/m}^2$, Poisson's ratio $\nu = 0.3$) which is firmly clamped around its circumference. The radius a and thickness t of the element are 2.5 mm and 1.1 mm respectively. On the application of uniform pressure p , the deflection y at any radial position r , measured from the centre, is given by the expression

$$y = \frac{3p(1 - \nu^2)(a^2 - r^2)^2}{16Et^3}$$

- Find the maximum design pressure if the allowable deflection of the element is limited to 0.3 times its thickness.
- Schematically show the variations of deflection, the radial and tangential stresses from the centreline to the edge of the diaphragm element.

Solution

(a) Since the maximum deflection will occur at the centre, i.e. $r = 0$, and $y_{\max} = 0.3t$ we get from the given expression

$$\begin{aligned} p_{\max} &= \frac{16Et^3(0.3t)}{3(1 - \nu^2)a^4} \\ &= \frac{16(2 \times 10^{11})0.3(1.1 \times 10^{-3})^4}{3(1 - 0.3^2)(25 \times 10^{-3})^4} \\ &= 2.568 \text{ MN/m}^2 \end{aligned}$$

(b) It can be shown⁴ that for small centre deflections, say $y < 0.5t$, of a clamped diaphragm of radius a (m) and thickness t (m), the maximum radial stress near the edges becomes

$$\sigma_r = \frac{3pa^2}{4t^2} \text{ N/m}^2$$

⁴See, *Strain gauges: kinds and uses*, HKP Neubert, Macmillan (London) pp. 141–3.

where p (N/m^2) is the pressure. The tangential stress is given by

$$\sigma_t = \frac{3p\nu a^2}{4t^2} \text{ N/m}^2$$

where ν is Poisson's ratio. The stresses as well as the deflection are shown schematically in Fig. 8.9.

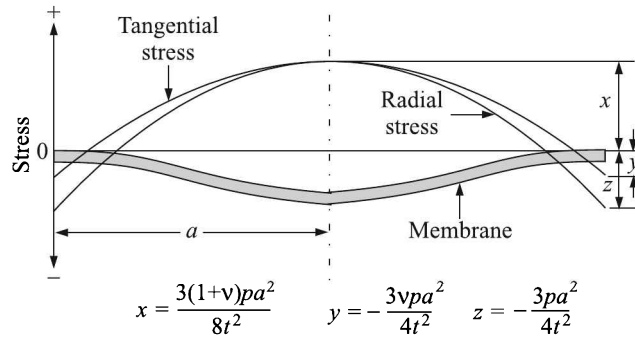


Fig. 8.9 Deflection of a circular diaphragm along with the radial and tangential stress distributions under small loads.

Bellows

Bellows are thin-walled cylindrical shells with deep convolutions and are sealed at one end (Fig. 8.10). The sealed end suffers axial displacement when pressure is applied at the open end.

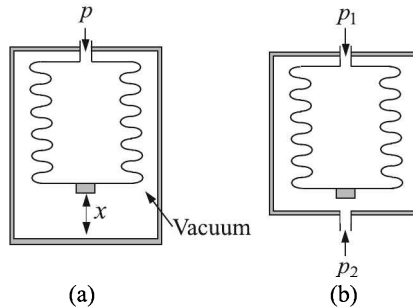


Fig. 8.10 Schematic view of bellows: (a) for absolute pressure measurement, and (b) for differential pressure measurement.

The absolute pressure can be measured by evacuating either the exterior or interior space of the bellows and then measuring the pressure at the opposite side. Bellows can only be connected to an ON/OFF switch or potentiometer and are used at low pressures, < 0.2 MPa, with a sensitivity of 1.2 kPa.

Example 8.8

Figure 8.11 shows the arrangement of differential bellows to measure absolute pressure. Each bellows has a natural length of 40 cm, an effective area of 1000 mm^2 and stiffness of 1 N/mm . Bellows A is evacuated and contains a spring S of stiffness 3 N/mm .

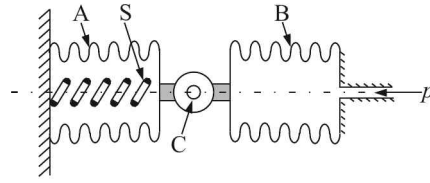


Fig. 8.11 Arrangement of differential bellows (Example 8.8). A, B: bellows, S: spring and C: output point.

- Supposing the bellows are to be compressed to a length of 30 mm when a pressure of $p = 120 \text{ kN/mm}^2$ absolute is applied to B, find the required natural length of the spring.
- Find the displacement of the output point C for a change of 12 kN/mm^2 in the applied pressure p .

Solution

- The force f on the spring of effective area A , caused by the applied pressure p is

$$f = pA = (120 \times 10^3)(1000 \times 10^{-6}) = 120 \text{ N}$$

If x is the displacement (i.e. compression) of the spring of stiffness k that balances f , then

$$x = \frac{f}{k} = \frac{120}{3} = 40 \text{ mm}$$

Therefore, the required natural length of the spring is $(40 + 40) = 80 \text{ mm}$.

- The change in pressure Δp is 12 kN/mm^2 . So, the change in force Δf is

$$\Delta f = \frac{\Delta p}{A} = \frac{12 \times 10^3}{1000 \times 10^{-6}} = 12 \text{ N}$$

This force is balanced by the elastic reaction created by the two bellows and the spring, their total stiffness being $(2 \times 1 + 3) = 5 \text{ N/mm}$. Therefore, the displacement of the output point C is

$$\frac{12}{5} = 2.4 \text{ mm}$$

Bourdon Gauge

The Bourdon⁵ gauge is one of the most common pressure sensors in use. All the various forms of Bourdon gauges have a common feature—they are constructed of tubes of non-circular cross-section [Fig. 8.12(a)]. Bourdon tubes can be of

- C-type [Fig. 8.12(b)]
- Spiral type [Fig. 8.12(c)]
- Twisted-tube type [Fig. 8.12(d)]
- Helical type [Fig. 8.12(e)]

⁵Named after Eugene Bourdon (1808–1884), a French watchmaker and engineer, who invented the device.

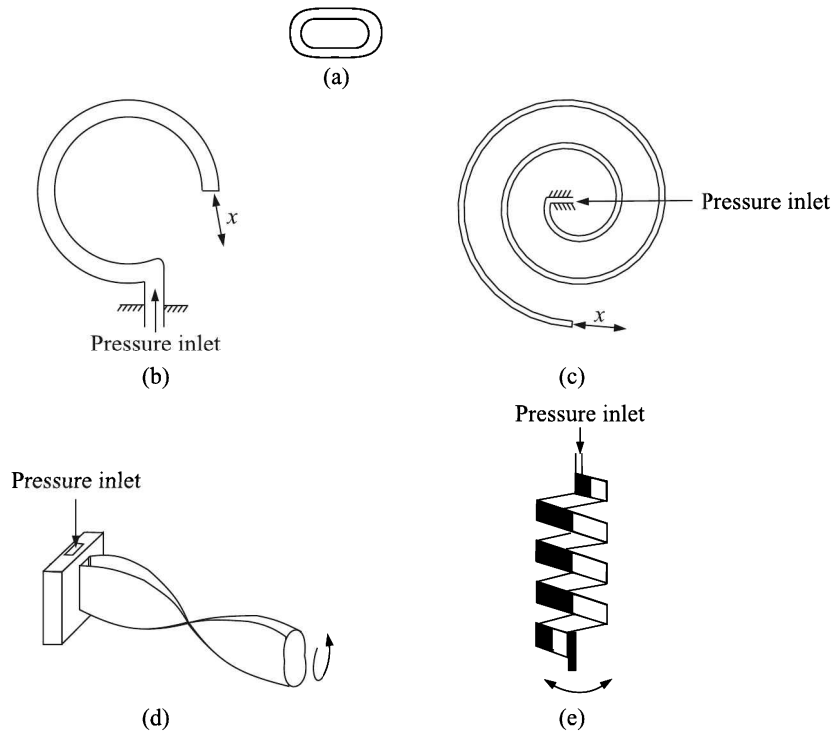


Fig. 8.12 (a) Non-circular cross-section of Bourdon tubes. Bourdon tubes of different shapes: (b) C-type, (c) spiral type (d) twisted-tube type, and (e) helical type.

The simplest form of the Bourdon gauge comprises a C-shaped metal tube (Fig. 8.13).

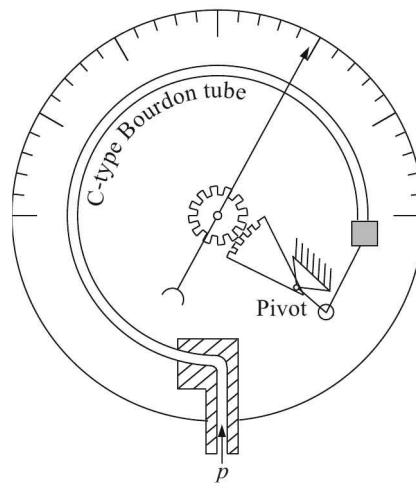


Fig. 8.13 Schematic diagram of a pressure gauge using a C-type Bourdon tube.

One end of the tube is sealed, and the other end is connected to the source of pressure that is being measured. The end that pressure is applied to is mounted in such a way that it cannot

move. When pressure is applied to the inside of the tube, the sealed end of the tube will tend to straighten out. This causes a small amount of movement of the sealed end in an arc, and this motion is converted into the rotation of a segment of a gear by a connecting link which is usually adjustable. A small diameter pinion gear is on the pointer shaft, so the motion is magnified further by the gear ratio.

Adjustments. The following components of the gauge provide adjustments for calibration of the pointer to indicate the desired range of pressure:

1. Positioning of the indicator card behind the pointer
2. Initial pointer shaft position
3. Linkage length and initial position

Differential pressure can be measured by gauges containing two different Bourdon tubes, with connecting linkages of the sealed end.

Alternatively, the movement can be converted to an electrical signal with the help of, say, an LVDT [see Fig. 8.16(b)]. Mechanical dial-type gauges incorporating Bourdon tubes are most common.

Range. Bourdon gauges can operate under a pressure range from 0.1 to 700 MPa. C-type tubes can be used in pressures up to 700 MPa. Their minimum recommended pressure range is 30 kPa, i.e. they are not sensitive enough for pressure differences less than 30 kPa.

Construction materials. Tube materials can be changed accordingly to suit the required process conditions. Usually stainless steel is used to construct tubes for ordinary ranges; for higher ranges phosphor bronze is used.

Bourdon tubes measure *gauge pressure*, relative to ambient atmospheric pressure, as opposed to absolute pressure; vacuum is sensed as a reverse motion. They are portable and require little maintenance. However, they can only be used for static measurements and have low accuracy.

8.5 Secondary Transducers

Displacements produced by force-summing devices are converted to readable format by means of secondary transducers. Secondary transducers can be

1. Mechanical
2. Resistive
3. Inductive
4. Capacitive
5. Photoelectric
6. Piezoelectric
7. Hall effect based
8. Vibrating element type
9. Surface acoustic wave type

Mechanical Transducers

The displacement generated by force-summing devices are, till today, generally used to rotate mechanically a pointer in front of a dial. In these mechanical pressure sensors, a Bourdon tube, a diaphragm or a bellows element detects the pressure and causes a corresponding movement. Aneroid barometers or aneroid sphygmomanometers⁶ use capsule-type diaphragms while large industrial pressure gauges use Bourdon tubes (Fig. 8.13) as force-summing devices. Diaphragms are preferred for common applications because they require less space.

Example 8.9

A water line carrying water at a pressure of 10 kg/cm^2 is running at the ground level. For the ease of reading, a Bourdon tube pressure gauge is mounted at a height of 2 m to measure the line pressure. What will be the error in the measurement?

Solution

Since the gauge is mounted 2 m above the water line, it will read less by an amount equal to a 2 m head than the water line pressure. Therefore,

$$\begin{aligned} \text{Pressure read by the gauge} &= 10 \text{ kg/cm}^2 - \text{pressure exerted by a column of 2 m of water} \\ &= 10 - \frac{2 \times 1000}{10^2} = 9.8 \text{ kg/cm}^2 \end{aligned}$$

Thus, the error in measurement will be 2%.

Note: This situation demands a zero elevation⁷ to get correct readings from the gauge.

Resistive Transducers

The resistive secondary transducers are generally of two types—strain gauge and potentiometer.

Strain gauge type

Usually wire-wound or foil type strain gauges are used for this purpose. Two such gauges, one on the top and another at the bottom side can be bonded to a diaphragm and can be connected to two arms of a Wheatstone bridge, making it essentially a half-bridge arrangement. But the problem with these types of strain gauges is that the bond between the diaphragm and the wire filament degrades with time causing changes in calibration. The search for improved strain gauges to measure pressure, and of course strain, resulted in the introduction of bonded thin-film and later diffused semiconductor strain gauges⁸. The general arrangement of using a strain gauge for measuring a differential pressure is shown in Fig. 8.14.

The normal range of these transducers is as low as 5 mm of Hg to as high as 200,000 psig (1400 MPa). The error in measurement typically lies between 0.1% of the span and 0.25% FSD.

⁶A pressure gauge for measuring blood pressure.

⁷See Section 12.1 at page 502.

⁸See Section 7.2 at page 245 for a detailed discussion.

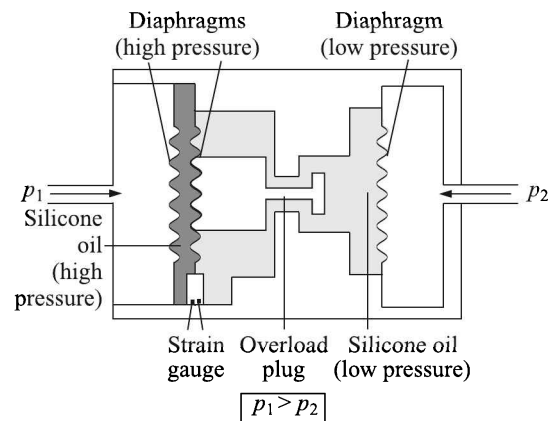


Fig. 8.14 Schematic diagram of a differential pressure gauge using a strain gauge.

Potentiometer type

The potentiometric secondary transducer consists of a precision potentiometer, the wiper arm of which is mechanically linked to a force-summing device like the bellows, Bourdon tube or diaphragm. We have shown such an arrangement in Fig. 8.15 using bellows as the force-summing device. The movement of the bellows causes the wiper arm to move across the potentiometer. The potentiometer, in turn, converts the mechanically detected deflection of the force-summing device into a resistance measurement using a Wheatstone bridge circuit.

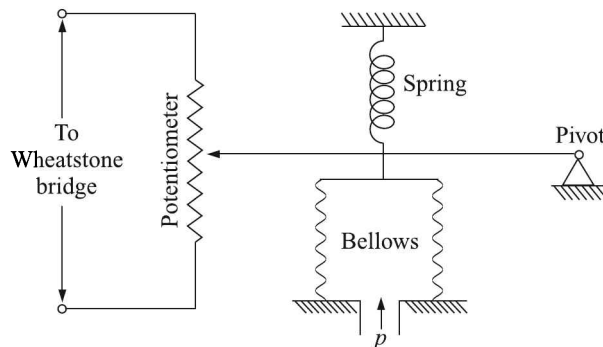


Fig. 8.15 Potentiometer type secondary transducer using a bellows force-summing device.

Apart from the threshold problem of the mechanical nature of the linkage between the wiper arm and the force-summing device, temperature effects introduce errors into this type of measurement. However, potentiometric transducers can be made extremely small and installed in a small space like that offered by the housing of a 10 cm dial pressure gauge. They provide a strong output signal that can be read without further amplification. Above all, they are inexpensive.

Pressure gauges having potentiometric secondary transducers can measure pressures between 5 and 10,000 psig (35 kPa and 70 MPa) with an accuracy lying between 0.5% and 1% of the FSD.

Inductive Transducers

As shown in Fig. 8.16, LVDTs can be used in conjunction with bellows or Bourdon tubes to measure pressure. The diagrams are self-explanatory.

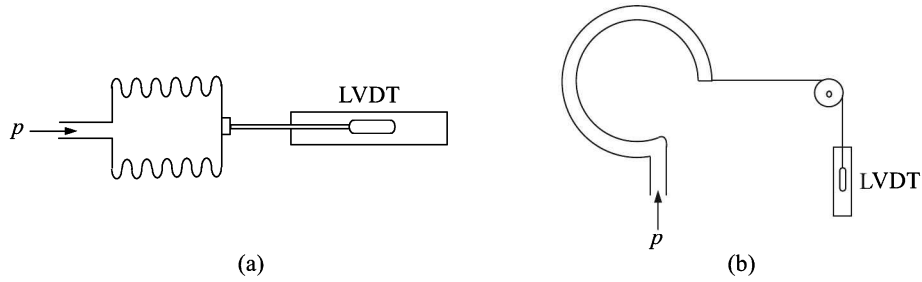


Fig. 8.16 LVDT as secondary transducer used in conjunction with (a) bellows, and (b) Bourdon tube.

Another arrangement, in which a diaphragm alters the reluctance of the flux path of two coils (Fig. 8.17) on application of a pressure difference, is often used.

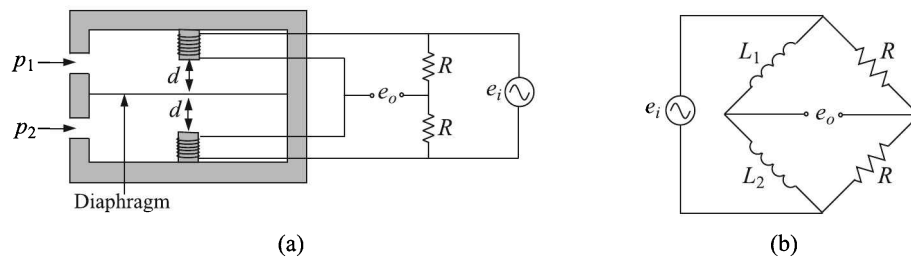


Fig. 8.17 Inductive secondary transducer for pressure measurement: (a) schematic diagram of the transducer and the circuit, and (b) the equivalent bridge.

Here the coils having equal number of turns are connected to two arms of a bridge, the other two arms of which consist of a pair of equal resistances R . For $p_1 = p_2$, the bridge is balanced and the output voltage $e_o = 0$. But when $p_1 \neq p_2$, the bridge will generate an output voltage. The following analysis shows that this voltage varies linearly with the displacement of the diaphragm. And, since the diaphragm displacement varies linearly with the applied pressure within a certain limit, the output maintains a linear relationship with the differential pressure. Now, initial self-inductance of a coil = N^2/R_0 where N is its number of turns and R_0 is the reluctance of the flux path. If the diaphragm which is made of magnetic material, suffers a small displacement x owing to a pressure differential Δp , then the two reluctances on either side of the diaphragm can be written as

$$R_1 = R_0 + K(d - x)$$

$$R_2 = R_0 + K(d + x)$$

where d is the initial distance between the diaphragm and the coil and K is a constant.

As a result, the two coils will have the following values of self-inductance,

$$L_1 = \frac{N^2}{R_o + K(d - x)} \quad (8.8)$$

$$L_2 = \frac{N^2}{R_o + K(d + x)} \quad (8.9)$$

With the bridge connection as shown in Fig. 8.17, the output voltage is

$$e_o = \left(\frac{1}{2} - \frac{L_2}{L_1 + L_2} \right) e_i \quad (8.10)$$

Substituting the values of L_1 and L_2 from Eqs. (8.8) and (8.9) into Eq. (8.10) we get on simplification,

$$e_o = \frac{Kx}{2(R_o + Kd)} e_i \quad (8.11)$$

Thus, $e_o \propto x$ other terms being constant, and if the condition of linear response from a diaphragm is satisfied, $e_o \propto \Delta p$.

The range of inductive transducers is generally 0 to 100 kPa.

Capacitive Transducers

A special construction [Fig. 8.18(a)] is used for capacitive pressure transducers. Spheroidal depressions of a depth of about 0.025 mm are ground into two glass discs which are held face to face with a taut thin stainless steel diaphragm in between. The depressions are coated with gold to form two fixed plates of the differential capacitor and the diaphragm forms the movable plate. So, this is essentially a differential arrangement of capacitors as shown in Fig. 8.18(c). We have already seen in Section 6.2 at page 189 that such an arrangement produces an output voltage that is proportional to the displacement of the movable plate. Though the plates are not strictly parallel here, it can be shown that the linear relationship remains valid for small displacements of the diaphragm.

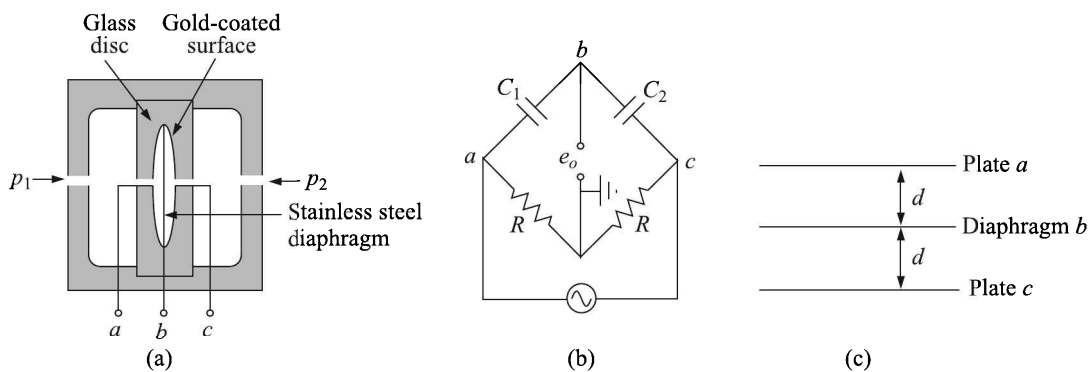


Fig. 8.18 Capacitive pressure transducer: (a) construction, (b) bridge arrangement, and (c) capacitor configuration.

Measurements are made with the help of a bridge circuit [Fig. 8.18(b)] where the two capacitors form the two arms of the bridge. In this arrangement, displacements between 10^{-8} mm and 10 mm can be sensed with an accuracy of 0.1% and therefore, very accurate pressure measurements at lower range are possible. The input frequency of the supply voltage should be around 2.5 kHz.

- Note:*
1. A differential pressure can be measured with this arrangement. To measure pressure of just one gas, the other side of the diaphragm may be evacuated and sealed.
 2. The arrangement is eminently suitable for measuring differential pressures of *gases*.
 3. We did not consider the relative permittivity of the gases in our analysis at page 189. So, if there are two different gases with different relative permittivities on the two sides of the diaphragm, the calibrations done with one gas may not hold good.

Photoelectric Transducers

The arrangement is shown in Fig. 8.19(a). With the application of pressure, the diaphragm is depressed with a consequent reduction in the width of the aperture and, therefore, a reduction in the intensity of light falling on the photon detector. This causes a variation in the output voltage owing to the variation of current in the circuit.

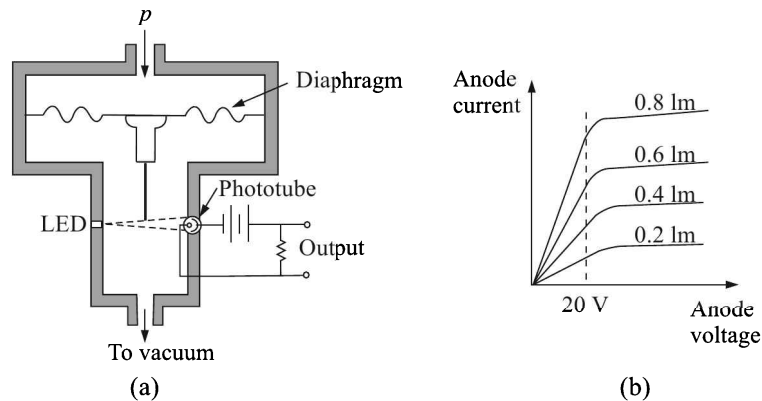


Fig. 8.19 (a) Photoelectric secondary transducer, and (b) phototube characteristics.

A typical phototube⁹ characteristics plot is shown in Fig. 8.19(b) which indicates that if the anode voltage is higher than 20 V, the anode current is linearly proportional to the intensity of illumination.

Obviously, such arrangements need a highly stabilised power supply to the lamp and since the photocurrent is of the order of μA , a stable amplifier is necessary for the signal. The other disadvantage is that a rather appreciable displacement of the diaphragm is necessary to produce a detectable current.

⁹See page 158.

Piezoelectric Transducers

We have already discussed¹⁰ what piezoelectricity is. We consider here how it is utilised in pressure measurement.

Piezoelectricity phenomenon can be utilised in the following three ways to measure pressure:

1. By measuring the electrostatic charge
2. By measuring piezoresistivity¹¹
3. By measuring the resonant frequency¹²

Depending on which way is used, the sensor can be called *electrostatic*, *piezoresistive*, or *resonant*.

Electrostatic piezoelectric transducer. We know that the charge developed across a piezoelectric crystal is proportional to the force applied on it and therefore, it can be used to measure pressure which is nothing but the force per unit area. But the fundamental difference between electrostatic piezoelectric sensors and such devices as strain gauges is that the electric signal generated by the piezoelectric crystal decays rapidly. This characteristic makes electrostatic piezoelectric sensors unsuitable for measurement of static pressures. But they can be used for dynamic pressure measurement. The appearance of a typical electrostatic piezoelectric pressure sensor is shown in Fig. 8.20. The electrostatic piezoelectric sensors are small and rugged. The pressure can be applied in the longitudinal or transverse direction. In either case, a high output voltage, proportional to the pressure, is generated. Their high speed response—30 kHz with peaks to 100 kHz—makes them suitable for measuring transient phenomena like rapidly changing pressures from blasts, explosions or other sources of shock and vibration.

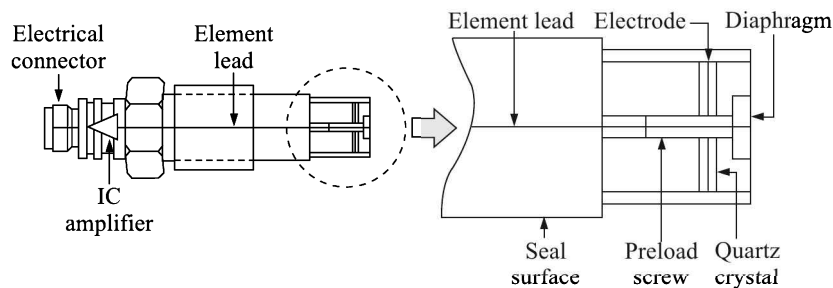


Fig. 8.20 Electrostatic piezoelectric pressure sensor.

The output of these dynamic pressure sensors is often expressed in units of *relative* pressure ('psir' instead of psig) because the output depends on the initial condition of the crystal. Their range is generally 0.1 to 10,000 psir (0.7 kPa to 70 MPa). The typical accuracy is 1% FSD.

Piezoresistive transducer. The resistivity of semiconductors depends on the force applied on them. The piezoresistive pressure sensors operate on this principle of measurement of

¹⁰See page 130.

¹¹See page 150.

¹²See Circuit analysis at page 140.

the resistivity of semiconductors. Like a strain gauge, in a piezoresistive sensor four pairs of semiconducting resistors are bonded onto a diaphragm. But unlike the construction of a strain gauge sensor, here the diaphragm itself is a semiconductor (typically silicon) wafer and the sensing resistors are diffused into it during the growth of the wafer.

If a measurement of the absolute pressure is intended, the bonding process is carried out in vacuum and the cavity behind the diaphragm is kept evacuated. If a relative measurement is required, the cavity behind is either ported to a reference pressure or to the atmosphere.

Piezoresistive sensors can be used to measure pressures between 3 psi and 14,000 psi (21 kPa and 100 MPa). We note that since piezoresistivity does not decay with time, we can measure static pressures with the help these sensors.

Resonant transducer. Resonant piezoelectric pressure sensors, which also can measure static pressures, utilise the variation of the resonant frequency of piezoelectric transducer crystals when a force is applied to them. The sensor, which can be in the form of a suspended beam, can be made to oscillate through the inverse piezoelectric effect by applying an oscillating electric field. The change in the resonant frequency is related to the applied pressure as

$$p = A \left(1 - \frac{\nu_p}{\nu_0} \right) - B \left(1 - \frac{\nu_p^2}{\nu_0^2} \right) \quad (8.12)$$

where A , B are constants and ν_p , ν_0 are resonant frequencies corresponding to applied pressure p and zero pressure respectively.

These transducers can be used to measure pressures between 0 and 900 psia (0 and 6 MPa) in different spans.

Advantage and disadvantages of piezoelectric transducers

Table 8.3 lists the advantages and disadvantages of piezoelectric pressure sensors.

Table 8.3 Advantages and disadvantages of piezoelectric pressure sensors

| <i>Advantages</i> | <i>Disadvantages</i> |
|---|--|
| 1. Small in size, lightweight, and very rugged. | 1. The electrostatic type cannot measure static or absolute pressure for more than a few seconds. |
| 2. The electrostatic transducer may cover a dynamic pressure range of $1:10^5$ and frequency range from 2 Hz to 1 MHz with almost no phase shift. | 2. All types are sensitive to temperature changes and, therefore, require proper temperature compensation. |
| 3. Outputs are quite large (see Example 5.3). | |
| 4. Special units operate up to 350°C . | |

Hall Effect Transducer

The Hall effect can be used as a secondary transducer to measure pressure by coupling it with a force-summing device like bellows.

A magnetic assembly is attached to a bellows assembly (Fig. 8.21). As the bellows expands and contracts with the application of pressure, the magnetic assembly is moved. If the sensor is placed in close proximity to the assembly, an output voltage proportional to pressure input can be achieved.

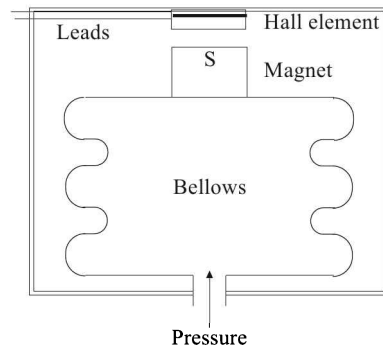


Fig. 8.21 Use of Hall effect transducer in pressure sensing.

Vibrating Element Transducers

The vibrating element may be a resonant wire or a cylinder. Let us consider only the resonant wire, the resonant cylinder being very similar. We know that the frequency of vibration of a resonant wire is given by

$$\nu = \frac{1}{2\pi} \sqrt{\frac{T}{m}}$$

where ν is the frequency of vibration, T is the tension, and m is the mass per unit length of the wire. The resonant wire secondary pressure transducer utilises the change in the frequency of vibration of a resonant wire with its tension to measure pressure. The change in the tension of the resonant wire with pressure is caused by a force-summing device like a diaphragm. A schematic drawing of the arrangement is shown in Fig. 8.22.

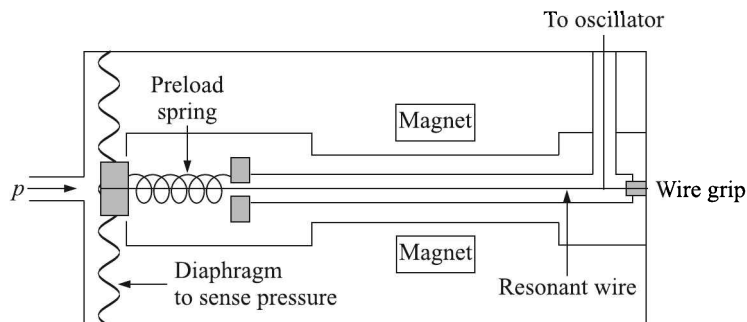


Fig. 8.22 Resonant wire secondary pressure transducer.

An oscillating voltage, supplied by an oscillator, is fed to the wire which is placed between the pole-pieces of a magnet. This causes the wire to vibrate. The oscillator frequency can be tuned to cause a resonant vibration. A change in the process pressure, sensed by the

diaphragm, changes the wire tension, which in turn changes the resonant frequency of the wire. A digital counter can detect the frequency shift and calibrate it in terms of the pressure.

We have already talked about the resonant piezoelectric transducer which is used for pressure measurement. Such transducers have the elegance of exciting oscillation just by placing the sensor in an oscillating electric field. The inverse piezoelectric effect causes the sensor to oscillate. In a resonant wire transducer, however, an oscillating current has to be passed through the wire and the wire needs to be placed in a static magnetic field to excite oscillation. This is distinctly a disadvantage. But because of the fragile nature of piezoelectric sensors, they cannot withstand very high changes of frequency. As a result, they cannot be used to measure high pressure changes.

Advantages and disadvantages

The advantages and disadvantages of vibrating element pressure transducer are given in Table 8.4.

Table 8.4 Advantages and disadvantages of vibrating element pressure sensor

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|---|
| <ol style="list-style-type: none"> 1. Generates an inherently digital signal which is amenable to its acquisition by the microprocessor-based instrumentation. 2. Because the change in frequency can be detected quite precisely, the resonant wire transducer is suitable for measurement of low differential pressures as well as high gauge pressures. 3. The detectable pressure range is typically from 10 mm of Hg to 6,000 psig (42 MPa) with a typical accuracy of 0.1% of the span. | <ol style="list-style-type: none"> 1. Sensitive to temperature variation, shock and vibration. 2. Nonlinear output, though can be suitably compensated for by software. |

Surface Acoustic Wave Transducer

Surface acoustic wave velocities are strongly affected by stresses applied to the piezoelectric substrate on which the wave is propagating. A SAW pressure sensor is therefore created by making the SAW device onto a diaphragm (Fig 8.23).

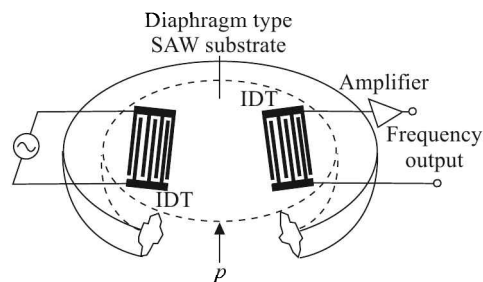


Fig. 8.23 Surface acoustic wave transducer for pressure measurement.

Unfortunately, the device is also affected by a change of temperature. The temperature drifts that tend to interfere with SAW pressure sensors can be compensated by placing a reference SAW device close to the measuring SAW on the same substrate and mixing the two signals. One sensor acts as a temperature detector, whose proximity to the pressure sensor ensures that both are exposed to the same temperature. However, the temperature sensor SAW must be isolated from the stresses that the pressure SAW experiences.

SAW pressure sensors can be wireless. They are low cost, rugged, very small in size and lightweight.

8.6 Vacuum Measurement

Although the vacuum literally means *a space from which the air has been completely or partly removed*¹³, it is with the latter meaning that pressures lower than the atmospheric pressure are often referred to as vacuum and the corresponding measurement, vacuum measurement.

The vacuum is usually measured in Torr which equals the pressure exerted by a 1 mm column of mercury. Table 8.5 will give an idea about the extent of low pressures which can be measured with the help of gauges we have already discussed.

Table 8.5 Low pressure measurement limits of conventional gauges

| Gauge | Manometers, bellows | Bourdon tube | Diaphragm gauge |
|-----------------------------------|---------------------|--------------|-----------------|
| Lowest measurable pressure (Torr) | 0.1 | 10 | 10^{-3} |

It is apparent from Table 8.5 that conventional gauges are no good for measurement of low pressures or degrees of vacuum. Such measurements call for different techniques. We discuss here a few common low pressure gauges which employ some such techniques. These gauges can broadly be divided into three categories:

1. Mechanical
2. Thermal
3. Ionisation based

Mechanical Vacuum Gauges

Of all the mechanical vacuum gauges, we will consider only the following:

1. McLeod gauge
2. Knudsen gauge
3. Molecular momentum gauge
4. Viscous friction gauge

¹³ *Pocket Oxford Dictionary*, Oxford University Press, 9th Ed. (2004) Delhi.

McLeod gauge

To measure a low pressure, we can isolate a sample of the low pressure gas, compress it to a known extent and measure the resultant pressure with a simple manometer. Precisely, this principle is followed in pressure measurement by a McLeod¹⁴ gauge.

The arrangement is shown in Fig. 8.24. Initially the mercury is lowered to a fixed mark in the tube by lowering the mercury reservoir, thus admitting the gas at unknown pressure p_i into the gauge. Then the reservoir is raised when the mercury level goes up sealing off a gas sample of known volume V in the bulb and the capillary tube A. The reservoir is raised slowly (so that the gas does not get heated) until the mercury level in the capillary B reaches the '0' mark.

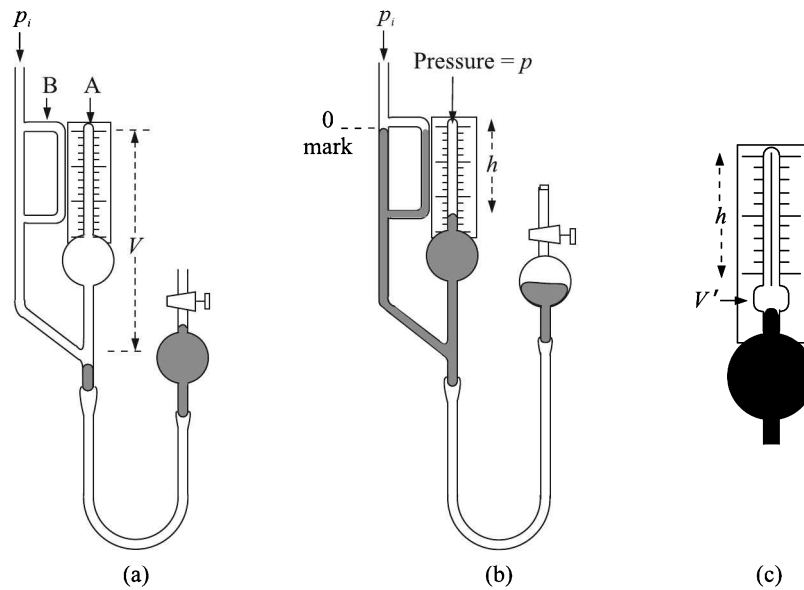


Fig. 8.24 Low pressure measurement by McLeod gauge: (a) drawing a gas sample, (b) compressing the gas sample, and (c) including another volume V' to linearise the scale at relatively high pressure.

- If, α is the area of cross-section of the capillary tube A
 h is the height of the gas column trapped in the capillary tube A when mercury level in B has reached the '0' mark
 p is the pressure of the trapped gas in A
 ρ is the density of mercury
 g is the acceleration due to gravity

then we have

$$p = p_i + h\rho g \quad (8.13)$$

$$p_i V = p\alpha h \quad (8.14)$$

¹⁴Named after its inventor Herbert McLeod (1841–1923), a British chemist.

From Eqs. (8.13) and (8.14) we have

$$p_i V = (p_i + h\rho g)\alpha h = p_i \alpha h + \rho g \alpha h^2 \quad (8.15)$$

$$\Rightarrow p_i = \frac{\rho g \alpha h^2}{V - \alpha h} \quad (8.16)$$

$$\approx \frac{\alpha \rho g h^2}{V} \quad \text{if } V \gg \alpha h \quad (8.17)$$

Other quantities on the right-hand side of Eq. (8.17) being constant for the instrument, the pressure can be calculated by measuring h . Two precautions need to be taken while using the gauge:

1. If the gas contains vapours which may condense on compression, there will be error in the measurement.
2. When employed to calibrate other gauges, a liquid-air trap should be used between the McLeod gauge and the gauge to be calibrated to prevent passage of mercury vapour.

Equation (8.17) shows that p_i varies as h^2 .

Linearisation. The relation can be made linear by the following procedure:

1. An additional volume V' [see Fig. 8.24(c)] is to be added to the capillary tube, and
2. The mercury level, after compression of the sampled gas, has to remain below V' .

Obviously, the second condition is fulfilled at relatively high pressure of the sample. Then Eqs. (8.15) and (8.17) can be rewritten as

$$\begin{aligned} p_i V &= p(V' + \alpha h) \\ &= (p_i + h\rho g)(V' + \alpha h) \\ &= p_i(V' + \alpha h) + h\rho g(V' + \alpha h) \\ \Rightarrow p_i &= \frac{h\rho g(V' + \alpha h)}{V - (V' + \alpha h)} \\ &\cong \frac{h\rho g V'}{V - V'} \quad \text{because } V' \gg \alpha h \end{aligned} \quad (8.18)$$

Equation (8.18) shows that with this modification of the capillary tube, the p_i vs h relation becomes linear.

Though McLeod gauges allow measurement of absolute pressure, the procedure is time-consuming and therefore they are mostly used for calibration purposes. They can measure pressure up to 10^{-4} Torr, below which the mercury vapour interferes.

Example 8.10

To find the constant of a McLeod gauge, it was connected in parallel with an accurate water manometer. When the manometer showed 13.6 mm pressure, the McLeod gauge recorded a reading of 20 cm. What is the pressure of the gas container which shows a reading of 30 mm by the same McLeod gauge?

Solution

Since the specific gravity of mercury is nearly 13.6, 13.6 mmWG corresponds to 1 mm of Hg = 1 Torr. From Eq. (8.17)

$$p \equiv Kh^2$$

we get

$$\begin{aligned} K &= \frac{1.0}{(200)^2} \\ &= 2.5 \times 10^{-5} \text{ Torr/mm}^2 \end{aligned}$$

Therefore the McLeod gauge reading of 30 mm corresponds to

$$p = (2.5 \times 10^{-5})(30)^2 = 2.25 \times 10^{-2} \text{ Torr}$$

Example 8.11

A McLeod gauge has a volume of the bulb equal to 100 cm³ and a capillary of diameter 1 mm. Calculate the pressure indicated by a reading of 3 cm. What error would result if the capillary volume is assumed to be negligible compared to the volume of the bulb?

Solution

Given, $V = 100 \text{ cm}^3$, $d = 1 \text{ mm}$, and $h = 3 \text{ cm}$. Therefore, the area of cross-section of the capillary is

$$\alpha = \frac{\pi d^2}{4} = \frac{\pi(0.1)^2}{4} = 7.854 \times 10^{-3} \text{ cm}^2$$

From Eq. (8.16), we get

$$\begin{aligned} p_i &= \frac{\alpha h^2}{V - \alpha h} \text{ cm of Hg} \\ &= \frac{(7.854 \times 10^{-3})(3)^2}{100 - (7.854 \times 10^{-3})(3)} \times 10 \text{ Torr} \\ &= 7.07 \times 10^{-3} \text{ Torr} \end{aligned}$$

If $\alpha h \ll V$,

$$p_i = \frac{(7.854 \times 10^{-3})(3)^2}{100} \times 10 = 7.069 \times 10^{-3} \text{ Torr}$$

Therefore,

$$\% \text{ error} = \frac{(7.07 - 7.069) \times 10^{-3}}{7.07 \times 10^{-3}} \times 100 = 0.014$$

Knudsen gauge

If we have

1. Two surfaces at different temperatures, and
2. The separation between the surfaces is less than the mean free path of gas molecules in the intervening space

then the motion of molecules exerts a mechanical force between two surfaces. The second condition ensures that the pressure is low enough so that the molecules suffer less number of collisions among themselves and thus can acquire sufficient kinetic energy. The molecules striking the hotter surface will rebound with a higher momentum than those striking the cooler surface generating a mechanical force. The Knudsen¹⁵ gauge is constructed on this principle.

Figure 8.25 shows the internal arrangement of the Knudsen gauge. V_1 and V_2 constitute two movable vanes of a rectangular frame which is suspended by a quartz fibre Q . A mirror M is attached to the frame. S_1 and S_2 are two electrically heated stationary plates kept opposite to the movable vanes. The entire arrangement is kept within an evacuated vessel with an inlet for the low pressure sample.

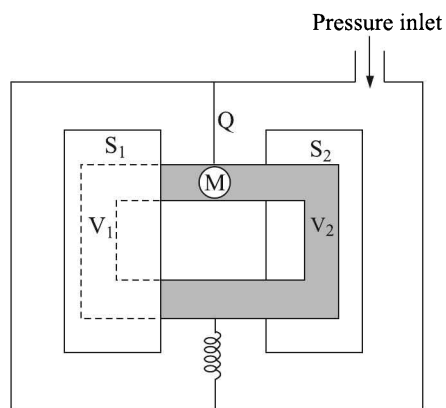


Fig. 8.25 Schematic arrangement of the Knudsen gauge. S_1 and S_2 are two electrically heated stationary plates, V_1 and V_2 are two movable vanes of a rectangular frame, Q is a quartz fibre and M is a mirror.

At sufficiently low pressure within the vessel, gas molecules rebound from the heated stationary plates with greater momentum than from the cooler movable vane, thus imparting a net force (couple) on the vanes. As a result, the suspended assembly suffers a rotation which can be measured by a lamp and scale arrangement.

Knudsen figured out that the pressure inside the vessel p_i can be calculated from a knowledge of the absolute temperatures of the movable and stationary plates T_m and T_s respectively, from the relation

$$p_i = \frac{KF}{\sqrt{T_s/T_m - 1}}$$

where K is a constant for the gauge and F is the force which can be evaluated from the deflection and the elastic constants of the suspension fibre. This formula, however, was derived by Knudsen with the assumption that the distance between vanes and stationary plates is small compared to the mean free path of the gas molecules and, therefore, this gauge cannot be employed to measure pressure above 10^{-3} Torr when the mean free path becomes comparable with the distance between plates.

The Knudsen gauge is an absolute gauge for measurement of pressure in the range 10^{-8} to 10^{-3} Torr.

¹⁵Martin Hans Christian Knudsen (1871–1949) was a Danish physicist.

While the Knudsen gauge can measure absolute pressures at very high vacuum, it is more suitable for laboratories than industries. Another gauge, known as *molecular momentum gauge*, which is based on the transfer of momentum, is used in industries though it does not offer a high accuracy.

Molecular momentum gauge

The molecular momentum vacuum gauge consists of two cylinders. One of the cylinders is rotated at a constant speed of 3600 rpm. Gas molecules from the process sample come in contact with the rotating cylinder, experience a change in momentum, and are set in motion in the direction of the rotation. With the acquired momentum from the spinning cylinder, the gas molecules strike the restrained cylinder and transfer the momentum to it. The impact drives the restrained cylinder to a distance proportional to the momentum transferred. Obviously, the transferred momentum, and hence, the displacement of the restrained cylinder, is proportional to the number of gas molecules present in the gas. The number of gas molecules, in turn, is proportional to the pressure of the gas. The pointer attached to the restrained cylinder can be calibrated in terms of the pressure of the gas.

The momentum transferred by gas molecules is proportional to not only the number and velocity of gas molecules, but also the molecular weight of gas molecules. Therefore, the pressure calibration of the gauge becomes gas specific. For air, the range of the gauge is from 10^{-3} Torr to 20 Torr (1.3×10^{-4} kPa to 2.7 kPa) while for hydrogen, the maximum is 280 Torr (37 kPa). The accuracy of the gauge varies between $\pm 5\%$ and $\pm 25\%$ with accuracy decreasing at lower pressures. Additional inaccuracies are introduced by temperature variations, vibrations, and presence of dirt, vapours, etc. in the gas.

Viscous friction gauge

At high vacuum, when the mean free path of gas molecules becomes comparable to the container dimensions, viscosity of gases depends on the pressure of the gas. This phenomenon is utilised to construct the viscous friction vacuum gauge in which a magnetic ball is kept levitated in the gas by applying two opposing magnetic fields—one from a permanent magnet and another produced by current through coils (Fig. 8.26).

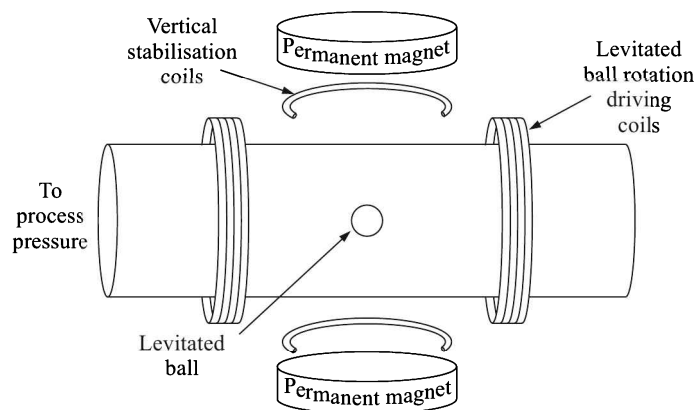


Fig. 8.26 Viscous friction vacuum gauge.

The ball can be made to spin by sending current through another pair of driver coils. The ball is first driven to spin at around 425 rotations per second. Then the driver is turned off and the rotational speed of the ball falls owing to the viscous friction offered by the gas present. The degree of vacuum is determined by measuring the length of time taken by the spinning ball to drop from 425 to 405 revolutions per second. The higher the vacuum, the lower the viscous friction and consequently the more time is taken by the ball to drop to the lower speed of rotation.

Usually the ball is made of magnetic stainless steel. Therefore, the gauge can be used to measure pressure in the presence of corrosive vapours. It can also be used at temperatures up to 400°C. Since the viscosity varies from gas to gas, the gauge needs to be calibrated for an individual gas. Then the accuracy is about $\pm 1.5\%$. However, if it is not so calibrated, the accuracy is around $\pm 4\%$.

This gauge can measure vacua down to 10^{-7} Torr.

Thermal Gauges

According to the kinetic theory of gases, the mean free path λ of gas molecules confined within a temperature enclosure is given by

$$\lambda = \frac{1}{\pi n \sigma^2} \quad (8.19)$$

where n is the molecular density and σ is the molecular diameter of the gas. An outcome of Eq. (8.19) is that the thermal conductivity K of a gas is independent of its pressure and is given by the relation

$$K = \varepsilon \mu c_V \quad (8.20)$$

where μ is the coefficient of viscosity
 c_V is the specific heat of the gas at constant pressure
 ε is a constant

At low pressures, however, the mean free path becomes comparable to the dimensions of the confining vessel and is almost constant. This phenomenon annuls the pressure-independent relation between the thermal conductivity, viscosity and specific heat of the gas given by Eq. (8.20) and makes the total loss of heat energy from a unit area directly proportional to pressure, p . This phenomenon is utilised in the construction of the thermal gauges, namely

1. Thermocouple gauge
2. Pirani gauge
3. Convectron gauge

Obviously, these gauges, except the convectron gauge, cannot be used to measure pressures above 1 Torr.

Thermocouple gauge

This gauge consists of a heated surface (H) and a thermocouple (C) in contact with it. The arrangement [Fig. 8.27(a)] is kept within a closed vessel and the surface is kept hot by passing a constant current through it, while the output of the thermocouple is measured by a high impedance microvoltmeter.

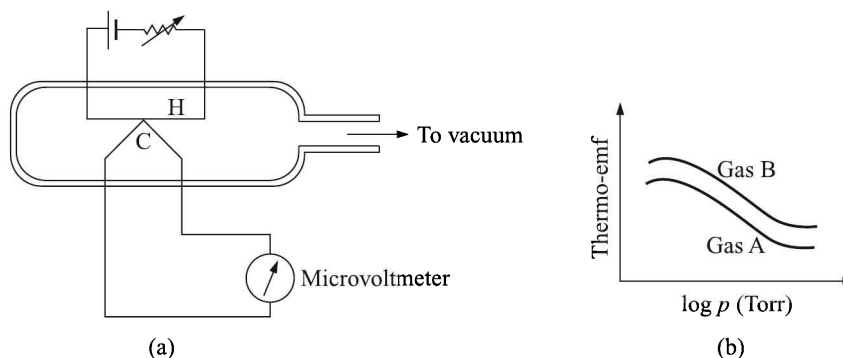


Fig. 8.27 (a) Thermocouple vacuum gauge, and (b) thermo-emf-pressure characteristics.

Heat is transferred from the hot surface to the cold envelope by conduction and radiation. While heat conduction is proportional to the pressure of the gas within the envelope, the radiation loss depends upon the emissivity of the hot surface as well as the temperature difference between the hot surface and the cold enclosure. Thus, a surface of low emissivity can be used to minimise the radiation loss. The plot of thermocouple voltage versus pressure assumes a form [Fig. 8.27(b)] in which the linear region falls between 10^{-4} and 1 Torr. Hence, the thermocouple gauge is used to measure pressure in this range.

Pirani gauge

A Pirani¹⁶ gauge is nothing but a heated resistive element placed in an envelope, the element constituting one arm of a Wheatstone bridge. The resistive element is taken in the form of four coiled tungsten or platinum filaments connected in parallel [Fig. 8.28(a)].

Such a construction is adopted to increase the contribution of the conduction process in the heat loss mechanism. As the pressure inside the envelope goes down, heat conduction decreases thus increasing the temperature of the resistive element with a consequent increase in its resistance. In a current-sensitive bridge [Fig. 8.28(b)], the output current thus becomes a function of pressure of gas within the envelope, the functional form having been shown in Fig. 8.28(c).

Alternatively, the temperature of the resistance wire can be held constant by adjusting the current passing through the element. Then, the change in the current will be proportional to the pressure of gas inside the gauge. This temperature-compensated design has the advantage of dispensing with the necessity of a bridge. Also, a constant temperature increases the life of the element.

The range of this gauge is 10^{-4} to 10^{-2} Torr where the pressure versus current curve is nearly linear. A Pirani gauge will not work above 1 Torr when the thermal conductivity of the gases no longer changes with pressure. Also, within its range of operation, the gauge has to be calibrated for the individual gas being measured because the thermal conductivity of each gas is different. Nevertheless, Pirani gauges are inexpensive, convenient and reasonably accurate. The typical accuracy is 2% in the calibrated region.

¹⁶Named after its inventor Marcello Stefano Pirani (1880–1968), a German physicist.

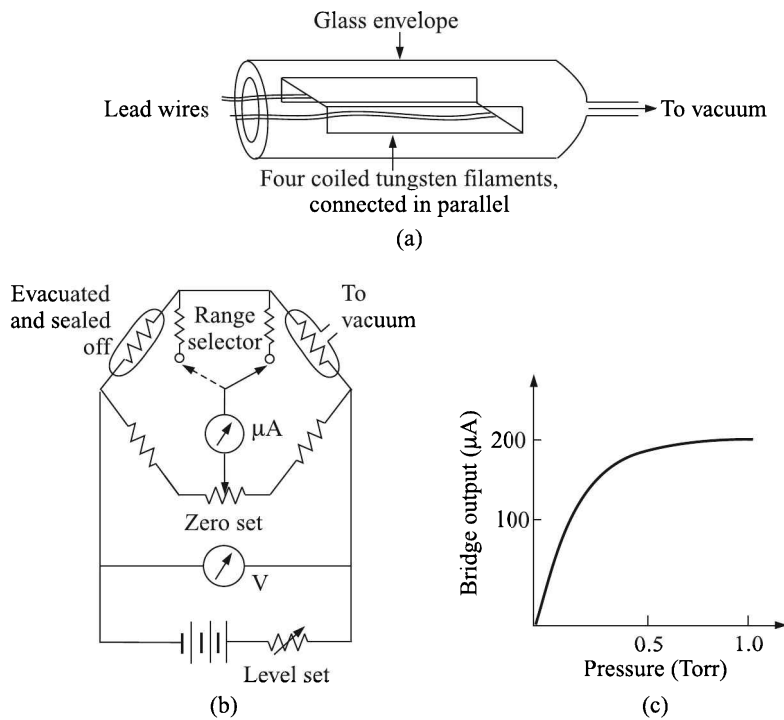


Fig. 8.28 (a) Pirani vacuum gauge, (b) bridge arrangement, and (c) pressure vs. bridge current plot.

Convectron gauge

In convectron gauges, which are very similar to Pirani gauges, cooling effects owing to both conduction and convection are detected. At higher vacua, the response depends on the conductivity of the gas, while at lower vacua the convective cooling by gas molecules predominates. The sensor used is a temperature-compensated, gold-plated tungsten wire. Its measurement range is from 10^{-3} to 1000 Torr. Except for its expanded range, its advantages and shortcomings are the same as those of Pirani gauges.

Ionisation Gauges

Ionisation gauges measure vacuum by measuring the current produced by ionised gas molecules. The gas molecules are ionised by impinging electrons on them. Three types of ionisation gauges are available:

1. Hot cathode ionisation gauge
2. Cold cathode ionisation gauge
3. Alphasatron ionisation gauge.

Hot cathode ionisation gauge

The hot cathode ionisation gauge comprises a heated filament, a grid and an anode. The filament acts as a cathode, the grid with a negative potential acts as ion collector and the

anode collects electrons (Fig. 8.29). The hot filament breaks up impinging gas atoms into ions and electrons. The ions get collected at the grid producing an ion current i_i , while electrons collected at the anode produce an electron current i_e .

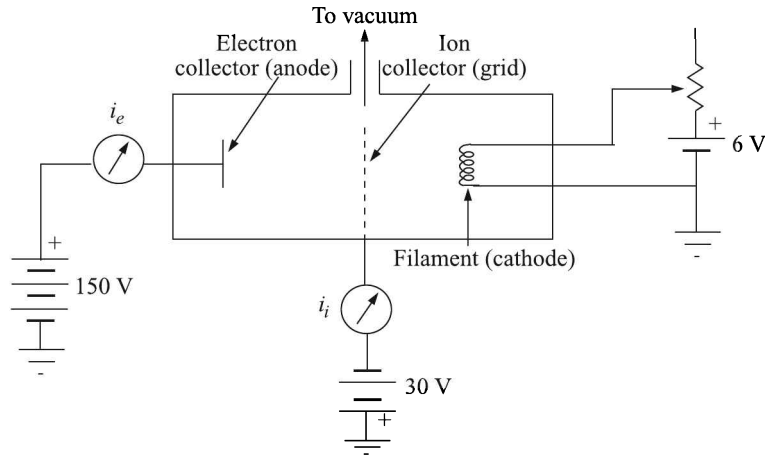


Fig. 8.29 Hot cathode ionisation gauge.

Obviously, the higher the pressure of the gas, the more the number of ions, i.e. $p_i \propto i_i$, where p_i is the input pressure. A higher number of ions, in turn, captures more number of electrons emitted from the filament. Therefore, $p_i \propto 1/i_e$. The two currents are related to the input pressure as

$$p_i = \frac{i_i}{Si_e}$$

where S is a constant, called the *sensitivity of the gauge*.

In early configurations, a filament remained at the centre with a grid surrounding the filament and a collector surrounding the grid (Fig. 8.30).

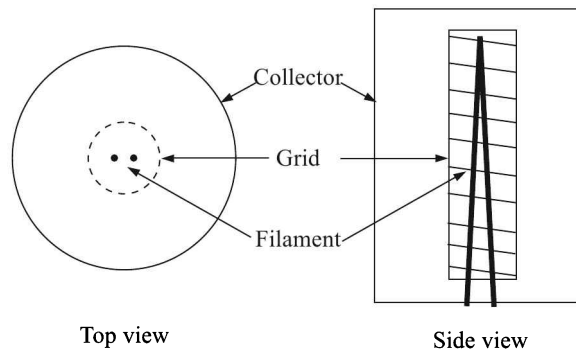


Fig. 8.30 Original triode configuration of the hot cathode ionisation gauge.

Although a tube of this configuration is very dependable, it cannot measure pressures below about 1×10^{-7} Torr, because of a background current unrelated to pressure, called the *X-ray limit*. What happens is that electrons from the filament striking the grid cause X-rays to

be emitted. The X-rays, in turn, strike the collector, where they release photoelectrons. The resulting photoelectric current, independent of pressure, is indistinguishable from the incoming ion current. This interference sets the lower limit of pressure measurement.

Bayard-Alpert geometry. In 1950, the X-ray limit problem was solved by inverting the geometry of the triode tube. The collector in the form of a wire of small diameter was put at the centre. The grid surrounded the collector and finally the filament was put outside the grid. This geometry (Fig. 8.31) gave a much smaller cross-section for collecting X-rays, reducing the X-ray limit by more than 3 decades. Now pressure could readily be measured to 1×10^{-10} Torr.

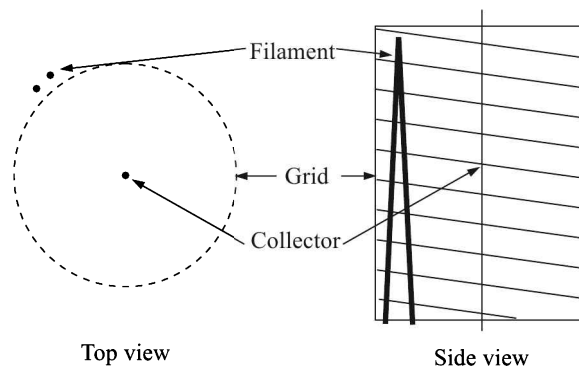


Fig. 8.31 Bayard-Alpert configuration of the hot cathode ionisation gauge.

This Bayard-Alpert geometry constitutes the majority of hot cathode ionisation gauges sold today. Additional designs to further limit the currents generated by X-ray through various shielding schemes have been implemented. But their complexity limits their applications to cases where the measurement of pressures below 10^{-10} Torr is required.

The Bayard-Alpert ionisation gauge can measure pressures from 10^{-10} to 10^{-2} Torr. Newer instruments use a modulated electron beam to extend this range. The modulated electron beam is synchronously detected to give two values for the ion current. At pressures below 10^{-3} Torr, the two values almost coincide while at higher pressures, the ratio between the two readings increases linearly, allowing the gauge to measure pressures up to 1 Torr.

However, because it decomposes gases due to high filament temperature, the gauge is not suitable for use in a decomposable gas environment. Also, an inrush of air may burn out the filament.

Cold cathode ionisation gauge

As the name implies, in a cold cathode ionisation gauge electrons are not produced by heating a filament. Here, a high electric field of about 4 kV is applied between the cathode and the anode to draw electrons out (Fig. 8.32).

A magnetic field of about 1500 gauss, applied around the tube, causes the electrons to spiral their way to the anode. This spiralling of electrons increases their probability of collision with the gas molecules thereby increasing the ionisation. The typical range of this gauge is from 10^{-10} to 10^{-2} Torr.

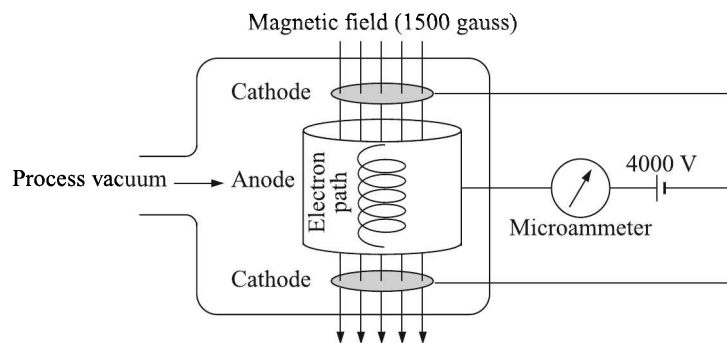


Fig. 8.32 Cold cathode ionisation gauge.

Having no hot filament, the cold cathode gauge can safely be used in a decomposable gas environment and for the same reason, is unaffected by the inrush of air.

Alphatron vacuum gauge

As already discussed, the hot cathode gauges suffer from problems such as

1. The cathode may get oxidised if oxygen is present. Also, its high temperature may initiate undesirable chemical reaction with other suitable gases.
2. Analyte gases may dissociate owing to the high temperature of the cathode.

The cold cathode ionisation gauges, on the other hand, are rather costly. Alphatron ionisation gauges provide a cheap alternative in a decomposable gas environment. Here a suitable α -emitter is used to ionise the gas and a positively-biased electrode is used to collect only the negative ions that are generated by the impact with the positively charged α -particles (Fig. 8.33).

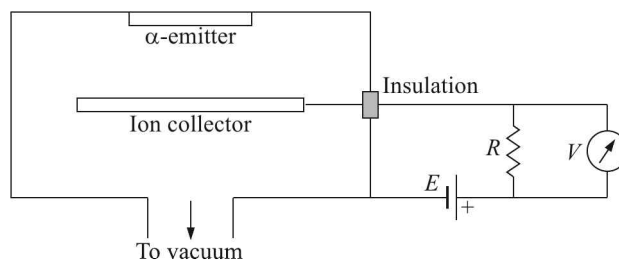


Fig. 8.33 Schematic diagram of alphatron vacuum gauge.

The generated current is proportional to the degree of vacuum inside the gauge. It flows through the resistor R and its value lies between 10^{-13} A and 10^{-9} A. A high impedance voltmeter is used to measure the voltage V across the resistor. The gauge has a linear characteristic between 10^{-3} Torr and 10^3 Torr. However, the calibration differs for different gases.

It may be mentioned here that a miniature variant of this gauge is used in household smoke detectors.

The usable pressure ranges and some areas of applications of different vacuum gauges are summed up in Fig. 8.34.

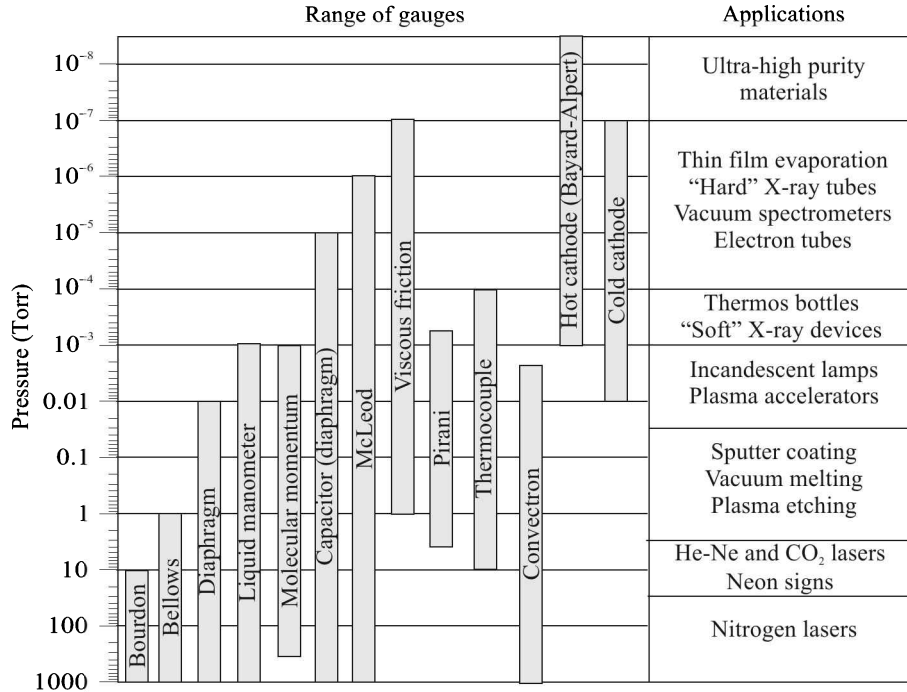


Fig. 8.34 Pressure ranges of vacuum gauges and their areas of applications.

8.7 Accessories

Of the many accessories needed in the pressure measurement, we will deal with only protective devices and pressure switches.

Protective Devices

The essential protective accessories for pressure gauges can be classified into five categories, namely:

1. Overpressure protector
2. Shut-off valves
3. Siphons
4. Snubbers
5. Chemical seals

Overpressure protector

Most pressure measuring instruments are provided with overpressure protectors that operate at overpressures of 50 to 200% of the range. One such method of protection for a C-type Bourdon

tube gauge is shown in Fig. 8.35. This way of stopping the flexure at a desired overpressure generally works for the majority of gauges. Where higher overpressures are expected and their nature is temporary (pressure spikes of short duration—seconds or less), snubbers (see later) can be installed. These filter out spikes, but cause the measurement to be less responsive. If excessive overpressure is expected to be of longer duration, the sensor can be protected by installing a pressure relief valve. However, this will result in a loss of measurement when the relief valve is open.

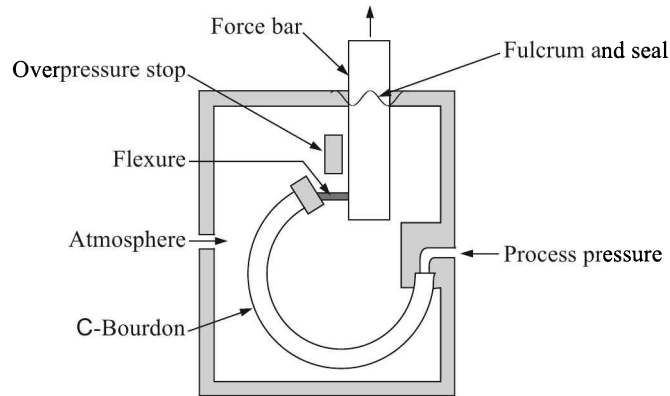


Fig. 8.35 Overpressure protection of Bourdon tube gauge.

Valve manifolds

Valve manifold is a standard accessory for pressure measurement and differential pressure transmitter. A pressure measuring instrument installed with a valve manifold allows a calibration or replacement of the instrument without the necessity of plant shutdown. Depending upon the application, three types of valve manifold are in use:

1. Two-valve manifold
2. Three-valve manifold
3. Five-valve manifold

Two-valve manifold. The two-valve manifold is used for pressure transmitters only. The typical two-valve manifold consists of one block valve and one drain or test valve (Fig. 8.36).

If it is necessary to calibrate the pressure transmitter, the block valve may be closed and the drain valve opened. Then the drain valve outlet is connected to the pressure generator to let a test pressure in.

Three-valve manifold. The three-valve manifold is used either for differential pressure measurement/transmission or for test/drain functions.

As shown in Fig. 8.37(a), the typical 3-valve manifold used in differential pressure measurement/transmission consists of two block valves and one equaliser valve.

To check the zero of the differential pressure measuring instrument/transmitter, we just need to close the block valves and open the equaliser valve.

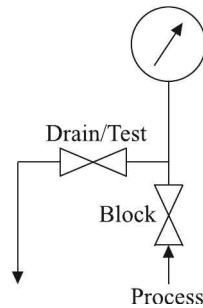


Fig. 8.36 Two-valve manifold.

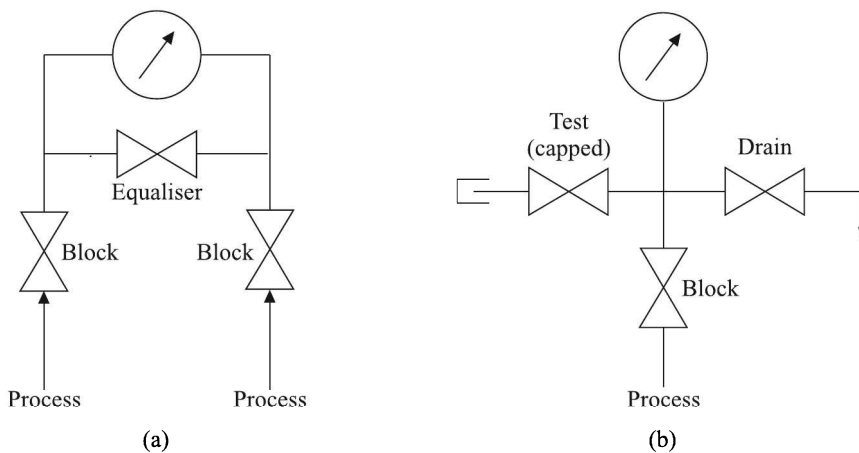


Fig. 8.37 Three-valve manifold: (a) for measurement of differential pressure, and (b) for test/drain.

Some manufacturers have modified the three-valve manifold by providing a capped test connection [Fig. 8.37(b)] for the common test/drain functions. Here, the block valve is used to isolate the process, the drain valve to discharge the trapped process fluid from the instrument to some safe container while the test valve can be used to calibrate the instrument by applying an external test pressure.

Five-valve manifold. Like the common three-valve manifold, five-valve manifold is also used for differential pressure transmitter. The typical five-valve manifold consists of two block valves, one equaliser valve and two vent or test valves (Fig. 8.38). To check the zero of the transmitter, we just need to close the block valves and open the equaliser valve. To calibrate the transmitter for 3 or 5 point calibration, after the pressure is equalised we need to connect the test valve to a pressure generator. Two drain valves are necessary to drain fluid from the two sides of the equaliser valve.

The five-valve manifold is the most common valve manifold for differential pressure transmitters.

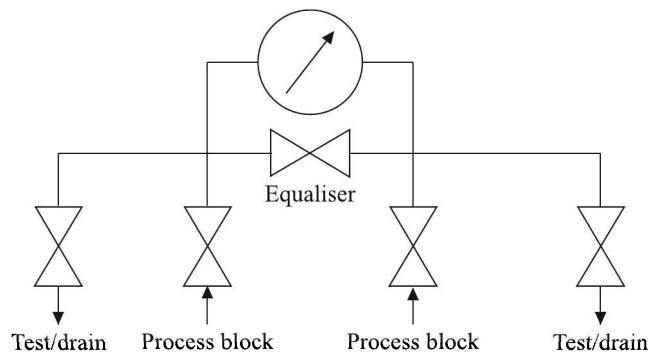


Fig. 8.38 Five-valve manifold.

Siphons

In some applications, it is desirable to prevent the process fluid to come in contact with the sensing element directly. The temperature of the fluid may damage the sensor, or the condensed fluid/solid may plug the sensor connectivity. Siphons [Fig. 8.39(a)] inserted between the sensor and the process can protect the gauge.

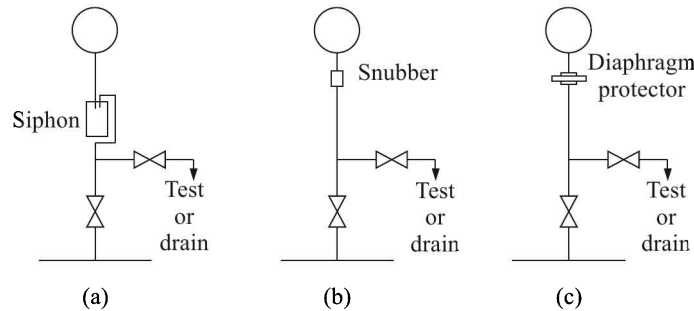


Fig. 8.39 Installation of protective accessories: (a) siphons, (b) snubbers, and (c) diaphragm protectors (or chemical seals).

Snubbers

If an unprotected pressure sensor is connected to, say, a positive displacement pump or a compressor, its pointer will cycle continuously which is undesirable. To bust or to average out such pressure spikes, it is necessary to install snubbers between the process and the pressure gauge [Fig. 8.39(b)]. They are also called *pulsation dampers*.

Corrosion-resistant porous metal filters, fixed or variable pistons or restrictions like a needle valve are used for this purpose (Fig. 8.40). But they invariably delay the pressure reading by a few seconds.

Chemical seals

Chemical seals [Fig. 8.39(c)] protect the gauge from plugging up in viscous or slurry service, and prevent corrosive, noxious or poisonous process materials from reaching the sensor. They

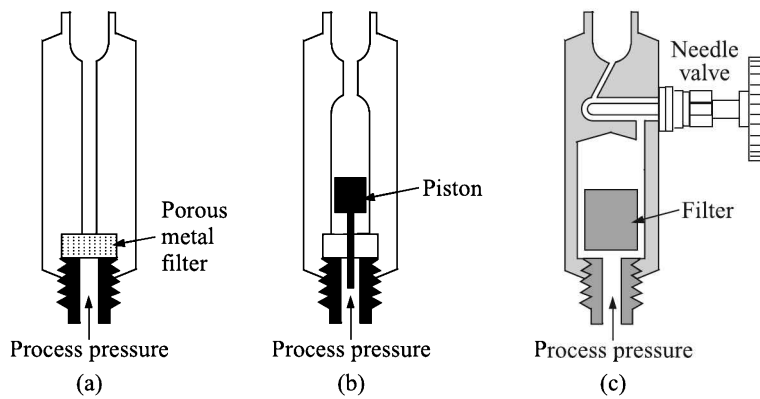


Fig. 8.40 Snubbers: (a) porous metal filter type, (b) variable piston type, and (c) needle valve restriction type.

also keep the process fluid from freezing or gelling in a dead-ended sensor cavity. The seal protects the gauge by placing a diaphragm between the process and the gauge. This is why they are often called the *diaphragm protectors*.

The cavity between the gauge and the diaphragm is filled with a stable, non-corrosive fluid of low thermal expansion and low viscosity. Depending on the temperature, different substances are used to fill the cavity. Table 8.6 gives an idea about the substances used.

Table 8.6 Substances used to fill the cavity between the gauge and the diaphragm

| <i>Temperature</i> | <i>Substance used</i> |
|--------------------|---|
| High | Sodium-potassium eutectic |
| Moderate | Mixture of glycerine and water |
| Low | Ethyl alcohol, toluene, or silicone oil |

The pressure gauge can be suitably located for better operator visibility if the chemical seal is connected to the gauge by a capillary tube. However, long or large bore capillaries increase the volume of the filling fluid. That is a source of error because it makes the pressure reading susceptible to the variation of the ambient temperature. Long and small bore capillaries, on the other hand, may cause a slow response.

The spring rate of the diaphragm in the chemical seal can cause measurement errors when detecting low pressures and in vacuum service (because gas bubbles dissolved in the filling fluid might come out of solution). For these reasons, pressure repeaters often are preferred to seals in such service.

Pressure Switches

Pressure switches are installed to control pressure of a process at a set pressure. They serve to energise or de-energise electrical circuits as a function of whether the process pressure is normal or abnormal.

Single pole, double throw (SPDT) snap action switches are standard. There the switch is provided with one normally closed (NC) and one normally open (NO) contact. Alternatively,

the switch can be configured as double pole double throw (DPDT), when two SPDT switches, each of which can operate a separate electric circuit, are used. The diagram of a normally open contact is shown in Fig. 8.41.

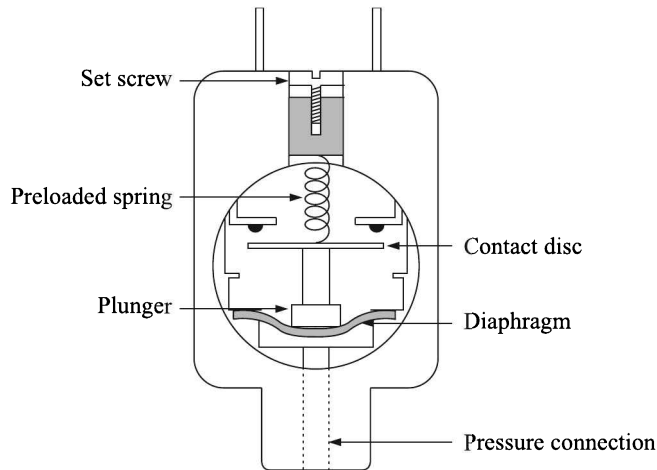


Fig. 8.41 Normally open (NO) pressure switch.

Pressure enters through the connection and acts on the diaphragm. If the force resulting from this pressure is greater than the force exerted by the preloaded compression spring, the plunger moves taking with it the contact disc, which closes the circuit between the contacts. When the pressure falls again by an amount greater than the hysteresis, the switch opens again.

For a normally closed switch, the action of the contacts is reversed. By turning the setting screw, the pressure switch can be adjusted to a set point within its pressure range.

Figure 8.42 illustrates the terminology used to describe pressure switch functionality and performance. When the pressure reaches the set point, the switch signals an *abnormal* condition.

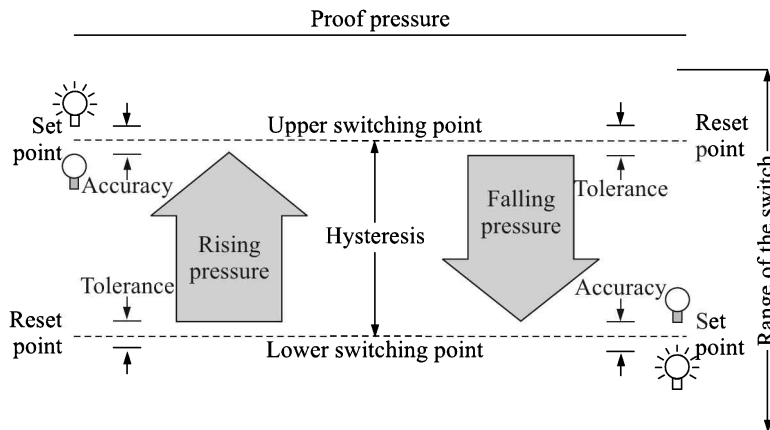


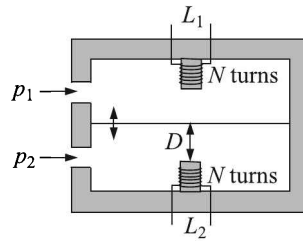
Fig. 8.42 Terminology of the pressure switch function.

It does not return to the *normal* state (called the *reset* or *reactivation* point) until the pressure goes down by the *hysteresis* (also called the *dead-band* or *differential*). The precision of the actuation at the set point is called its *accuracy*, while that of the reset point is called *tolerance*. The maximum pressure that can be applied to a pressure switch without causing an irreparable damage to it is called the *proof pressure*. It is usually 150% of the rated maximum system pressure of the pressure sensing element.

Review Questions

- 8.1 (a) What is meant by a 'force-summing device'? Briefly discuss about the construction and other aspects of commonly used force-summing devices.
- (b) Show how a differential pressure measuring arrangement comprising a suitable force-summing device and an inductive secondary transducer can produce a linear response.
- (c) Explain, with the help of a diagram, the principle of operation of a McLeod gauge.
- 8.2 (a) Show by a properly labelled diagram how the displacement of a force-summing device is converted to an electric signal by a photoelectric transducer.
- (b) What is the common basic element in all the various forms of Bourdon tube?
- 8.3 Explain the charge generator model of the piezoelectric accelerometer.
- 8.4 Define 'gauge' pressure. Show three different constructions of elastic pressure sensing elements. Assuming that the stresses in a metallic diaphragm are linearly proportional to the deflection, show an arrangement to get an electrical signal proportional to stress applied to the diaphragm.
- 8.5 Describe with diagram the principle of operation of vibrating element pressure sensor.
- 8.6 A piezoelectric pressure transducer with a normal compression disc configuration has 7.84 pC/torr sensitivity and 200 pF inherent capacitance. Calculate its natural frequency and voltage sensitivity.
- 8.7 (a) Explain the properties of a quartz crystal which make it suitable for its use in an oscillator.
- (b) Mention merits and demerits of the crystal oscillator.
- 8.8 Explain the working principle of a ring balance manometer. Also mention its advantages and limitations.
- 8.9 (a) Define the following terms: (i) absolute pressure, (ii) gauge pressure, (iii) differential pressure, and (iv) atmospheric pressure.
- (b) Describe the construction and operation of a Pirani gauge. Also mention its advantages and disadvantages.
- 8.10 (a) Discuss the use of elastic diaphragm for pressure measurement. What secondary transducers are generally used with elastic diaphragms?
- (b) Describe the use of an inclined U-tube manometer. What are the characteristics of the liquids used in manometer?

-
- 8.11 (a) Define d and g coefficients by which the sensitivity of a crystal is identified and obtain interrelationship between them.
- (b) Derive the expression for impulse response of piezoelectric transducers and sketch the response curves.
- 8.12 (a) Explain with a neat sketch the most important advantage of a well-type manometer over a simple U-tube one.
- (b) A well-type manometer has its 'capillary diameter-to-well diameter ratio' as 1:20. It is required to measure a pressure differential of 1 pascal. What should be the approximate height of the mercury column in the capillary?
- (c) What are over-range and under-range protections in a pressure measuring instrument? Explain with an example.
- 8.13 (a) Why is a McLeod gauge considered to be a standard for measurement of pressure in the vacuum range?
- (b) What are the limitations of a McLeod gauge?
- (c) What modification is done to linearise the scale of the McLeod gauge at higher pressures?
- 8.14 (a) Define Torr.
- (b) Why is the inclined tube manometer used?
- (c) What are the selection criteria of manometric fluid?
- 8.15 Describe the construction and operation of Pirani gauge. Why does such gauge require empirical calibration curve?
- 8.16 (a) Explain the working of a ring balance type manometer.
- (b) What are the over-range and under-range protections in a pressure measuring instrument? Explain with an example.
- (c) What errors occur during the pressure measurement by a C-type Bourdon tube? How can they be removed?
- (d) What are the important characteristics of the elastic members?
- 8.17 A differential pressure sensor shown in the figure below has a diaphragm as the primary sensor and a variable inductance transducer as the secondary element. When the pressures p_1 and p_2 are equal, the diaphragm maintains an airgap of D for the variable reluctances. The pressure difference is confined to a range where the deflection d is proportional to $p_1 - p_2$ and the reluctances are linearly proportional to airgap variation and is approximated as $R = R_0 + K(D \pm d)$. Two coils each of N turns are wound around the central limb to form variable inductances L_1 and L_2 .
- (a) Find the expressions for the inductances L_1 and L_2 .
- (b) Form a suitable ac bridge with these two inductances connected in push-pull and two proportional equal resistances and show that the open circuit output is proportional to d and $p_1 - p_2$.



- 8.18 A quartz crystal of dimensions $10 \text{ mm} \times 10 \text{ mm} \times 1 \text{ mm}$ is subjected to a displacement signal of $10^{-8} \sin 100t \text{ m}$. Find the voltage generated. Given: charge sensitivity $d = 2 \text{ pC/N}$, Young's modulus $E = 8.6 \times 10^{10} \text{ N/m}^2$, permittivity $\epsilon = 42 \times 10^{-12} \text{ F/m}$.
- 8.19 A piezoelectric transducer having diameter = 8 mm, thickness = 4 mm, charge sensitivity = $2 \times 10^{-12} \text{ C/N}$, dielectric constant = $4 \times 10^{-11} \text{ F/m}$ and modulus of elasticity = $8.6 \times 10^{10} \text{ N/m}^2$ is used for the measurement of small displacement. For an input displacement of 10^{-9} m , determine
- the force to which it is subjected,
 - the capacitance of the transducer,
 - the charge generated,
 - the voltage developed.
- 8.20 A piezoelectric transducer with a sensitivity of 2.0 pC/N having a capacitance of 1600 pF and a leakage resistance of $10^{12} \Omega$ is connected to a charge amplifier as shown in the Fig. 8.43.

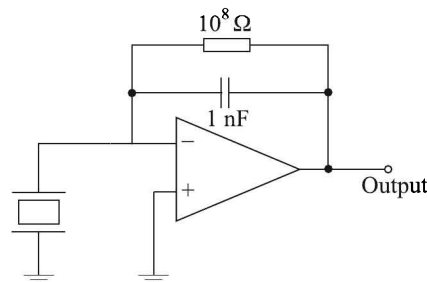


Fig. 8.43

If a force of $0.1 \sin 10t \text{ N}$ is applied to the transducer, the output amplitude of the charge amplifier is

- 0.141 mV
 - 0.232 mV
 - 1.414 mV
 - 1.732 mV
- 8.21 A well of cross-sectional area a_w is connected to an inclined tube of cross-sectional area a_t to form a differential pressure gauge as shown in Fig. 8.44. When $p_1 = p_2$ the common liquid level is denoted by A. When $p_1 > p_2$, the liquid level in the well is depressed to B, and the level in the tube rises by l along its length such that the difference between the tube and well levels is h_d . The angle of inclination θ of the tube with the horizontal is

- $\sin^{-1} \left[\frac{l}{h_d} - \frac{a_w}{a_t} \right]$
- $\sin^{-1} \left[\frac{h_d}{l} + \frac{a_t}{a_w} \right]$
- $\sin^{-1} \left[\frac{h_d}{l} - \frac{a_t}{a_w} \right]$
- $\sin^{-1} \left[\frac{h_d}{l} + \frac{a_w}{a_t} \right]$

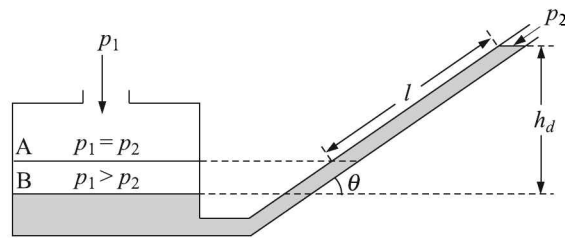


Fig. 8.44

8.22 Indicate the correct choice:

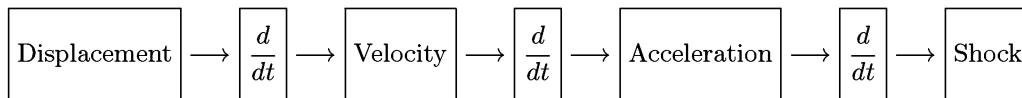
- (a) The least suitable transducer for static pressure measurement is
 - (i) semiconductor strain gauge
 - (ii) variable capacitor transducer
 - (iii) metal wire strain gauge
 - (iv) piezoelectric transducer
- (b) The operation of a Pirani gauge is based on
 - (i) ionisation of gas at low pressures
 - (ii) variation of volume with pressure
 - (iii) variation of viscosity with pressure
 - (iv) variation of thermal conductivity of gas with pressure
- (c) Bellows expansion is done usually against a spring. The spring is provided to
 - (i) increase sensitivity
 - (ii) increase operating range
 - (iii) increase linearity
 - (iv) decrease hysteresis effect
- (d) A piezoelectric transducer is directly connected through a cable to an electronic voltmeter. The minimum operating frequency of measurement is 1000 Hz. If the connecting cable length is doubled, the new minimum operating frequency is
 - (i) 500 Hz
 - (ii) 1000 Hz
 - (iii) 2000 Hz
 - (iv) 4000 Hz
- (e) The primary standard for calibrating vacuum is
 - (i) McLeod gauge
 - (ii) Dead-weight tester
 - (iii) Thermocouple gauge
 - (iv) Knudsen gauge

- (f) A pressure gauge used to measure vacuum indicates a gauge pressure of 5 kPa. If the atmospheric pressure is 100 kPa, the absolute pressure is
- (i) 105 kPa
 - (ii) 0.05 kPa
 - (iii) 95 kPa
 - (iv) 20 kPa
- (g) An elastic transducer is used to measure pressure in a vessel and it indicates a pressure of 3.2 bar. Atmospheric pressure is 1.01 bar. The absolute pressure in the vessel in bar is
- (i) 1.01
 - (ii) 2.19
 - (iii) 3.20
 - (iv) 4.21
- (h) Which of the following gauges can measure the lowest vacuum pressure?
- (i) McLeod gauge
 - (ii) Pirani gauge
 - (iii) Ionisation gauge
- (i) A Pirani gauge sensor is used to measure pressures of the order of
- (i) 10 MPa
 - (ii) 1 MPa
 - (iii) 100 Pa
 - (iv) 1 Pa
- (j) A quartz piezoelectric type pressure sensor has a built-in charge amplifier. The sensor has a sensitivity of $1 \mu\text{V}/\text{Pa}$. It is subjected to a constant pressure of 120 kPa. The output of the transducer at the steady state is
- (i) 0 mV
 - (ii) 100 μV
 - (iii) 120 μV
 - (iv) 120 mV
- (k) A pressure sensor has the following specifications: sensitivity at the design temperature = 10 V/MPa, zero drift = 0.01 V/°C, sensitivity drift = 0.01 (V/MPa)/°C. When the sensor is used in an ambient 20 °C above the design temperature, the output from the device is 7.4 V. The true value of the pressure will be
- (i) 0.71 MPa
 - (ii) 0.68 MPa
 - (iii) 0.65 MPa
 - (iv) 0.61 MPa

-
- (l) A quartz clock employs a ceramic crystal with a nominal resonance frequency of 32.768 kHz. The clock loses 30.32 s every 23 days. The actual resonance frequency of the crystal is
- 32.7685 kHz
 - 32.768 kHz
 - 32.7675 kHz
 - 32.7670 kHz
- (m) A mercury barometer reads h mm Hg with the temperature of the mercury at T °C. The barometer reading corrected for the standard temperature 0 °C with β denoting the volumetric expansion coefficient of mercury in °C⁻¹, is
- $\frac{h}{1 + \beta T}$
 - $h(\beta + T)$
 - $h(1 + \beta T)$
 - $h(\beta - T)$
- (n) A Pirani gauge measuring vacuum pressure works on the principle of
- change in ionising potential
 - change in thermal conductivity
 - deformation to elastic body
 - change in self-inductance
- (o) Liquid column manometers have the operating range of
- 2 to 700 MPa
 - 0 to 70 MPa
 - Up to 0.2 MPa
 - Up to 1000 MPa
- (p) The cross-sectional area of Bourdon tube is
- circular
 - elliptical
 - rectangular
 - none of these
- (q) Which accessory is commonly used in the installation of a pressure measuring device under pulsating condition?
- Diaphragm seal
 - Siphon
 - Snubber
 - 3-valve manifold

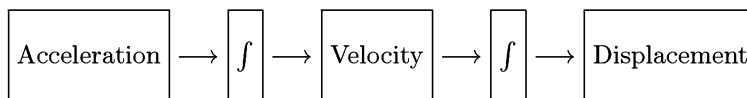
Acceleration, Force and Torque Measurement

We know that the velocity of a moving object is the time derivative of its displacement and the acceleration is the time derivative of its velocity. The time derivative of acceleration is defined as the *shock*. The relations are depicted below:



So, instead of trying to measure acceleration directly, we could obtain it by successively differentiating the displacement twice. If the displacement vs. time signal is a smooth function, it is, of course, not a difficult task. But in practice the rate of change of displacement varies arbitrarily and the functional form is rarely smooth. So, finding out the slope of the signal at every instant is indeed a tall order.

Now, let us consider the inverse procedure as shown below:



It involves at each step integration which means finding the area under a signal vs. time curve. This task is rather easy, whatever be the shape of the curve. Of course, none will try to find displacement by successively integrating the acceleration twice, but the velocity at any instant can be found with ease in this way.

Alternative unit. Though the SI unit of acceleration is m/s^2 , it is a common practice to measure accelerations in multiples of the acceleration due to gravity g which equals 9.81 m/s^2 . For example, it is often said that a passenger car negotiates a corner at about $2g$ acceleration while the acceleration in a space shuttle is about $10g$.

9.1 Acceleration Measurement

All acceleration measuring instruments or accelerometers convert acceleration a to a force F by allowing it to act on an inertial mass M . The mass, resting on a spring of stiffness k , causes a displacement x of the spring end. Mathematically speaking

$$F = -Ma \quad \text{[for inertial mass]}$$

$$F = kx \quad \text{[for the spring]}$$

$$\Rightarrow \quad a = -\frac{kx}{M}$$

So, by measuring the relative displacement of the mass, the acceleration can be measured.

Basically, an accelerometer involves measurement of the inertial displacement of a mass just not held by a spring, but by a spring and a dashpot. Let us first see how this movement is related to acceleration of the system to which the dashpot is attached.

Principle of Acceleration Measurement

In Fig. 9.1 we have a mass M that is free to move vertically. The mass is connected to the base of the housing by a spring of stiffness k that is in its relaxed state and a dashpot or damper of damping constant D . The whole assembly is rigidly fixed to a body that moves with an acceleration a upwards. At any instant, when the body moves a distance x_i , the displacement of the mass in the direction of motion is x_M and the relative displacement of the spring and the damper in the opposite direction is x_0 , as shown.

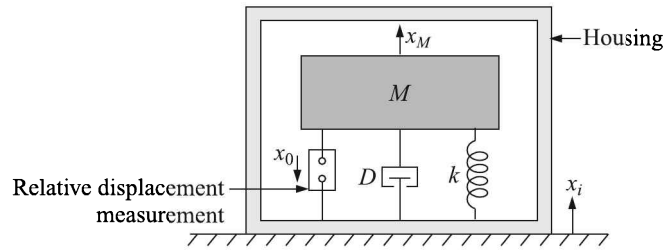


Fig. 9.1 The spring-mass system accelerometer.

By balancing the upward and downward forces acting on M , we get the equation

$$D \frac{dx_0}{dt} + kx_0 = M \frac{d^2x_M}{dt^2} \quad (9.1)$$

where $x_0 = x_i - x_M$. Substituting $x_M = x_i - x_0$ in Eq. (9.1) and rearranging, we get

$$M \frac{d^2x_0}{dt^2} + D \frac{dx_0}{dt} + kx_0 = M \frac{d^2x_i}{dt^2} \quad (9.2)$$

However, if the spring-mass system is exposed to a sinusoidal vibration of the form $x_i \sin \omega t$, then the resultant acceleration of the base is given by

$$a_i(t) = -\omega^2 x_i \sin \omega t \quad (9.3)$$

The Laplace transform of Eq. (9.2) gives

$$(Ms^2 + Ds + k)X_0(s) = Ms^2 X_i(s)$$

whence

$$\frac{X_0(s)}{X_i(s)} = \frac{Ms^2}{Ms^2 + Ds + k} \quad (9.4)$$

We are really interested in the frequency-domain solution of Eq. (9.2). The frequency-domain representation of Eq. (9.4) is given by

$$\frac{x_0}{x_i}(j\omega) = \frac{M(j\omega)^2}{M(j\omega)^2 + D(j\omega) + k} \quad (9.5)$$

Substituting

$$\left. \begin{aligned} \omega_n &= \sqrt{\frac{k}{M}} \\ \zeta &= \frac{D}{2\sqrt{kM}} \end{aligned} \right\} \quad (9.6)$$

where ω_n is the natural frequency of vibration of the system and ζ is the damping ratio, Eq. (9.5) can be written as

$$\frac{x_0}{x_i}(j\omega) = \frac{(j\omega)^2/\omega_n^2}{(j\omega/\omega_n)^2 + 2\zeta(j\omega)/\omega_n + 1} \quad (9.7)$$

By the usual procedure, we get the following expression for the amplitude ratio from the complex Eq. (9.7)

$$\left| \frac{x_0}{x_i} \right| = \frac{u^2}{\sqrt{(1-u^2)^2 + (2\zeta u)^2}} \quad (9.8)$$

where $u = \omega/\omega_n$. The graphical presentation of Eq. (9.8) is given in Fig. 9.2(a) for a few values of ζ .

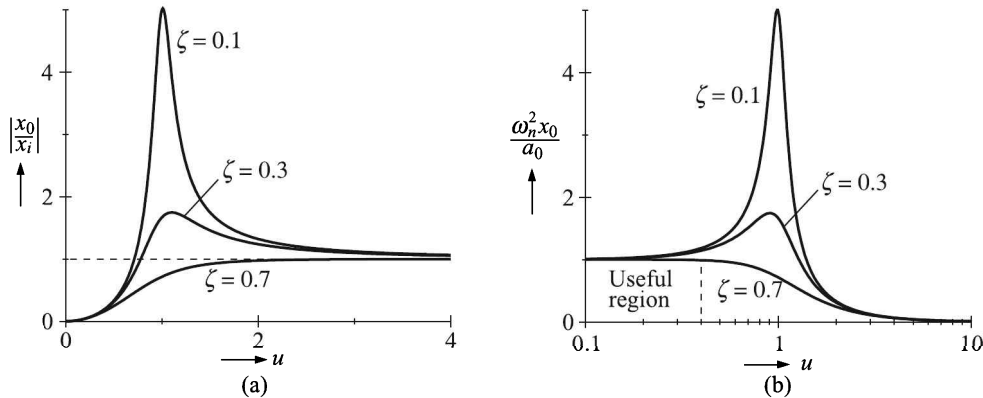


Fig. 9.2 Response of a spring-mass system to vibration: (a) displacement, and (b) acceleration.

Figure 9.2 shows the following two distinct cases:

For $u \ll 1$ i.e. $\omega \ll \omega_n$: The natural frequency has little effect on the basic spring-mass response. A rule of thumb is that a safe maximum applied frequency is $\omega < \omega_n/2.5$.

For $u \gg 1$ i.e. $\omega \gg \omega_n$: Here, $x_0/x_i \cong 1$. The mass-spring damper becomes a measure of vibration displacement x_i . It is interesting to note that the seismic mass is stationary in space in this case, and the housing, which is driven by the vibration, moves about the mass. A general rule is $\omega > 2.5\omega_n$ for this case.

Equation (9.8) can as well be written as

$$\left| \frac{\omega_n^2 x_0}{\omega^2 x_i} \right| = \frac{1}{\sqrt{(1-u^2)^2 + (2\zeta u)^2}}$$

or

$$\left| \frac{\omega_n^2 x_0}{a_0} \right| = \frac{1}{\sqrt{(1-u^2)^2 + (2\zeta u)^2}} \quad (9.9)$$

where $a_0 = \omega^2 x_i$ is the amplitude of acceleration of the object. Equation (9.9) is graphically presented in Fig. 9.2(b). We observe that the relative displacement x_0 is proportional to the acceleration of the object provided $u < 0.4$ and $\zeta = 0.7$.

The spring-mass principle applies to many common accelerometer designs. The mass that converts the acceleration to spring displacement is referred to as the *proof mass* or *seismic mass*. We see, then, that acceleration measurement reduces to linear displacement measurement. Most designs differ in how this displacement measurement is made.

Normally the proof mass of an accelerometer is mounted on a surface and the accelerometer records the acceleration of the surface on which it is mounted and produces an electrical output signal that varies with the acceleration. But consider the situation where the instrument slides on that surface. It still registers acceleration on the proof mass, but no longer is that the same as the surface upon which it rests.

Example 9.1

The seismic mass of a spring-mass accelerometer is 50 g and the spring constant is 5000 N/m. The amplitude of the mass displacement is ± 2 cm. Calculate

- the maximum measurable acceleration in g , and
- the natural frequency of oscillation of the system.

Solution

(a) The maximum measurable acceleration occurs when the LHS of Eq. (9.9) equals 1 [see Fig. 9.2(b)]. Using this result, we get for the maximum measurable acceleration as

$$a_0 = \frac{k}{m} x_0 = \frac{(5000)(0.02)}{0.05} = 2000 \text{ m/s}^2 = \frac{2000}{9.81} g = 204 g$$

(b) Using Eq. (9.6), we get for the natural frequency of oscillation as

$$f_n = \frac{1}{2\pi} \sqrt{\frac{k}{m}} = \frac{1}{2\pi} \sqrt{\frac{5000}{0.05}} = 50.3 \text{ Hz}$$

Example 9.2

A seismic accelerometer has a seismic mass $M = 0.005$ kg, stiffness $k = 8$ N/m and damping ratio $\zeta = 0.1$. If the input acceleration is $a_i(t) = 50 \sin 30t$ m/s², find the displacement $x_M(t)$ of the mass.

Solution

Given: $M = 0.005$ kg, $k = 8$ N/m, $\zeta = 0.1$, $\omega = 30$ rad/s and $a_0 = 50$ m. We have to find out x_M .

Now,

$$\omega_n = \sqrt{\frac{k}{M}} = \sqrt{\frac{8}{0.005}} = 40 \text{ rad/s}$$

$$\Rightarrow u = \frac{\omega}{\omega_n} = \frac{30}{40} = 0.75$$

Therefore, from Eq. (9.9) we have

$$\begin{aligned} x_0 &= \frac{a_0}{\omega_n^2 \sqrt{(1-u^2)^2 + (2\zeta u)^2}} \\ &= \frac{50}{(40)^2 \sqrt{(1-0.75^2)^2 + (2 \times 0.1 \times 0.75)^2}} \\ &= 0.0676 \text{ m} = 6.76 \text{ cm} \end{aligned} \quad \text{(i)}$$

From Eq. (9.8),

$$\begin{aligned} \left| \frac{x_0}{x_i} \right| &= \frac{0.75^2}{\sqrt{(1-0.75^2)^2 + (2 \times 0.1 \times 0.75)^2}} \\ &= 1.2162 \end{aligned}$$

This yields from Eq. (i), $x_i = 5.56 \text{ cm}$. Therefore,

$$|x_M| = |x_i - x_M| = |5.56 - 6.76| = 1.2 \text{ cm}$$

$$\Rightarrow x_M(t) = 0.012 \sin 30t \text{ m}$$

Transducers

From the previous discussion it is clear that the acceleration measurement boils down to the measurement of the relative displacement of the proof mass at any instant. This displacement can be measured with the help of transducers that can be divided into the following categories:

1. Resistive
2. Capacitive
3. Inductive
4. Piezoelectric
5. Piezoresistive
6. Hall effect
7. Magnetoresistive
8. Thermal

We have discussed the principles of operation of these transducers at length before. Here we will discuss only how these transducers are utilised to measure acceleration.

Resistive accelerometer

Resistive accelerometers detect the force imposed on a proof mass when acceleration occurs. The inertia of the mass resists the force of acceleration and thereby causes a physical displacement which can be measured by strain gauges as shown in Fig. 9.3. When the proof mass accelerates, two of the gauges get compressed and the other two stretched. The compressed and stretched strain gauges can be connected to the four arms of a Wheatstone bridge in the full-bridge configuration as shown in Fig. 7.11. It is often required to equip the gauges with damping devices, such as springs or magnets, to prevent oscillation.

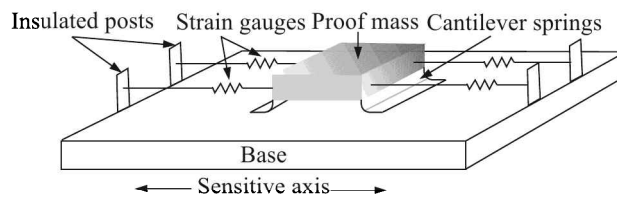


Fig. 9.3 Resistive accelerometer with strain gauge sensing of displacement.

Capacitive accelerometer

In capacitive accelerometers, micromachined capacitive plates (CMOS capacitor plates having depths of only 60 μm) form a mass of about 50 μg . As acceleration deforms the plates, a measurable change in capacitance results.

Inductive accelerometer

An inductive type of accelerometer utilises an LVDT to measure mass displacement. In these instruments, the LVDT core itself is the seismic mass. Displacements of the core are converted directly into a linearly proportional ac voltage. These accelerometers generally have a natural frequency of less than 80 Hz and are commonly used for steady-state and low-frequency vibration measurements. Figure 9.4 shows the basic structure of such an accelerometer.

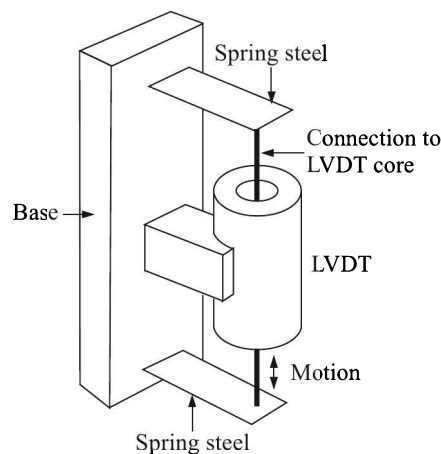


Fig. 9.4 Inductive accelerometer with LVDT sensing of displacement.

Piezoelectric accelerometer

The piezoelectric accelerometers are perhaps the most practical devices for measuring accelerations caused by shock and vibration. The device includes a mass that, when accelerated, exerts an inertial force on a piezoelectric crystal. The piezoelectric effect produces an accumulation of charge on the crystal. This charge is proportional to applied force which, in turn, is proportional to the acceleration. Piezoelectric crystals have both positive and negative outputs. In Fig. 9.5 the positive output electrode is connected to the mass and the negative electrode (ground) is connected to the housing. A signal conditioner converts the high impedance charge output into a usable voltage signal.

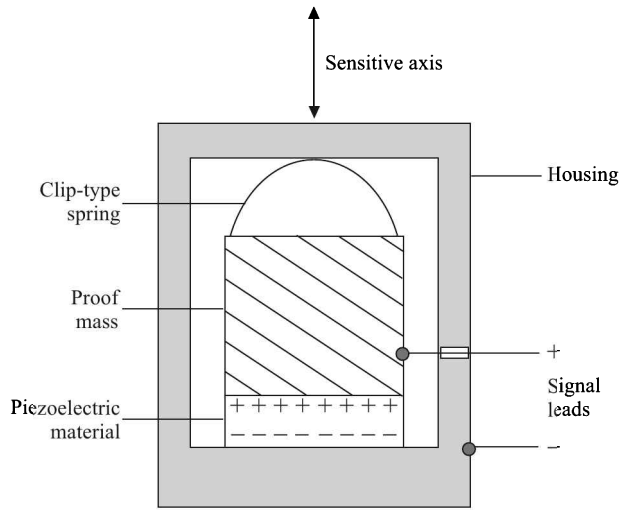


Fig. 9.5 Piezoelectric accelerometer.

There are several different piezo accelerometer design configurations including shear, compression and flexural i.e. bending (see Fig. 9.6).

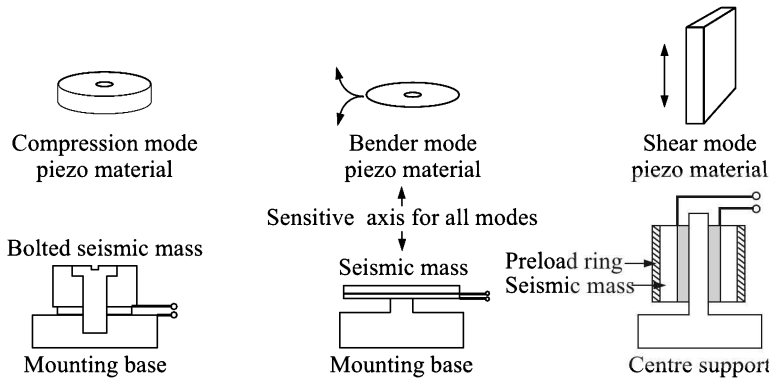


Fig. 9.6 Mechanical modes of the piezoelectric accelerometer. The piezo materials in the lower tier diagrams are indicated by the grey colour.

Although each design has its own advantages and disadvantages, the shear design is considered the most accurate since it is the least sensitive to temperature and base strain inputs. Shear is also the most widely used design.

For industrial machine vibration monitoring applications, shear structured accelerometers with integral electronics are packaged in robust hermetic sealed housings with durable electrical connectors to withstand tough factory environments. Sensors containing built-in signal conditioners are classified as *voltage mode*; *charge mode* sensors require external or remote signal conditioning.

Piezoresistive accelerometer

A piezoresistive accelerometer is similar to the resistive accelerometer shown in Fig. 9.3 having the strain gauges replaced with piezoresistive elements. Though piezoresistive and strain gauge sensors operate in a similar fashion, the strain gauge elements are temperature sensitive and require compensation. They are preferred for low frequency vibration, long-duration shock, and constant acceleration applications. Piezoresistive units are rugged, and can operate at frequencies up to 2 kHz.

Hall effect accelerometer

In a Hall effect accelerometer the proof mass is a Hall element that is placed in the field of a permanent magnet. Voltage variations stemming from a change in the magnetic field around the accelerometer are calibrated to the acceleration of the system.

Magnetoresistive accelerometer

The structure and function of a magnetoresistive accelerometer is similar to those of a Hall effect accelerometer except that instead of measuring voltage, the magnetoresistive accelerometer measures resistance.

Thermal accelerometer

Thermal accelerometers are among the newer mechanical accelerometer designs. In this sensor, a seismic mass is positioned above a heat source. If the mass moves because of acceleration, the proximity to the heat source changes and the temperature of the mass changes. Polysilicon thermopiles are used to detect changes in temperature which is proportional to the acceleration.

MEMS Accelerometers

Early accelerometers were analogue electronic devices that were later converted into digital electronic and microprocessor-based designs. However, nowadays hybrid micro-electro-mechanical systems (MEMS) are pretty common, especially in the automotive industry.

MEMS devices generally have components of size below 100 μm that are not machined using standard machining but using other techniques called micro-fabrication technology. Of course, this simple definition would also include microelectronics, but there is a characteristic that electronic circuits do not share with MEMS. While electronic circuits are inherently solid and compact structures, MEMS have holes, cavity, channels, cantilevers, membranes, etc., and,

in some way, imitate 'mechanical' parts. MEMS devices are often based on silicon because of the vast knowledge on silicon material and on silicon based microfabrication gained by decades of research in microelectronics. But then, many more MEMS are not based on silicon and can be manufactured in polymer, glass, quartz or even in metals.

Miniaturisation reduces cost by decreasing material consumption. It also increases applicability by reducing mass and size allowing to plant the MEMS in places where a traditional system does not fit. A typical example is that the accelerometer developed for airbag triggering sensor in motor cars is also used in digital cameras to help stabilise the image.

Another advantage of the MEMS is in the field of the system integration. Instead of having a series of external components (sensor, signal conditioner, etc.) connected by wire or soldered to a printed circuit board, the MEMS on silicon can be integrated directly with the electronics.

MEMS capacitive accelerometers are quite common. Also micromachined gyroscopes for acceleration measurement are available. This new technology is as vast as the microelectronics technology. We only made a cursory reference here.

Pendulous Integrating Gyro Accelerometer

The pendulous integrating gyro accelerometer (PIGA) is a type of accelerometer that measures acceleration and simultaneously integrates this acceleration against time to measure speed as well. It uses the gyroscopic action to convert an inertial force into a torque that precesses a gyro. The precession rate is proportional to the acceleration and the angle that the gyro turns through is the time integral of the acceleration, i.e. speed.

The diagram of a PIGA is shown in Fig. 9.7. The gyro is mounted, with its centre of mass offset along the spin axis, on a floated gimbal. This makes the gimbal pendulous. A suitable pick-up is mounted on the float. The gimbal fits across the diameter of a housing which is rotated by a torquer (not shown) about the sensitive axis. Power is fed to the gyro through slip rings (not shown).

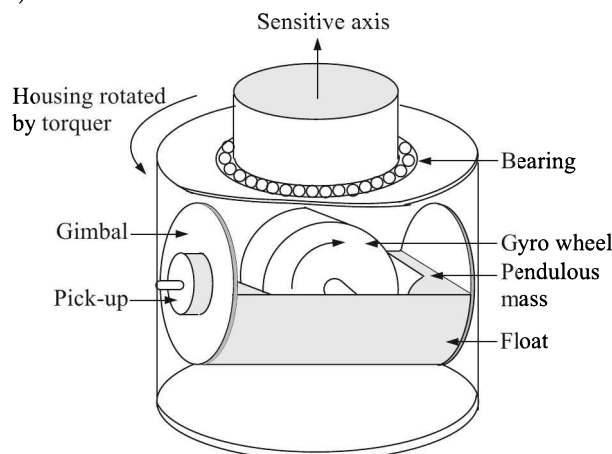


Fig. 9.7 Schematic diagram of the pendulous integrating gyro accelerometer.

If there is an acceleration along the sensitive axis, the gimbal pendulously builds up a torque about its axis. This motion is sensed by the pick-up which drives a servo to turn the

gimbal about the sensitive axis, generating a gyroscopic torque about the gimbal axis. This torque cancels the torque generated by the acceleration. Under a constant acceleration along the sensitive axis, the gimbal continuously rotates about the sensitive axis and this rotation rate is the measure of the acceleration.

PIGAs are very precise accelerometers. Their dynamic range is from 10^{-7} g to 50 g or more. Originally developed by Germany for guiding the V2 rockets during the WWII, the modified present form was arrived at by its inventor Mueller¹. Many of the missiles developed at the USA used PIGAs.

Applications of Accelerometers

1. Accelerometers are used in machinery vibration monitoring to diagnose, for example, out-of-balance conditions of rotating parts. An accelerometer-based vibration analyser can detect abnormal vibrations, analyse the vibration signature, and help identify its cause.
2. Accelerometers are used in structural testing, where the presence of a structural defect, such as a crack, bad weld, or corrosion can change the vibration signature of a structure. The structure may be the casing of a motor or turbine, a reactor vessel, or a tank. The test is performed by striking the structure with a hammer or by exciting the structure with a known forcing function. This generates a vibration pattern that can be recorded, analysed, and compared to a reference signature.
3. Accelerometers play a role in orientation and direction-finding. Miniature triaxial sensors detect changes in roll, pitch, and azimuth or X , Y , and Z axes. Such sensors are used to track drill bits in drilling operations, determine orientation for buoys and sonar systems, serve as compasses, and replace gyroscopes in inertial navigation systems.
4. In the computing world, a few manufacturers use accelerometers in their laptops to protect hard drives from damage. If the laptop is dropped accidentally, the accelerometer detects the sudden free fall, and switches the hard drive off so the heads do not crash on the platters.
5. In digital cameras, micro-machined accelerometers find use in stabilising the picture.
6. In automobiles, accelerometers are used for detecting car crashes and deploying airbags just at the right time.
7. The measurement of acceleration is used as an input into some types of control systems. The control systems use the measured acceleration to correct for changing dynamic conditions.

9.2 Force Measurement

Force, as we all know, is the product of mass and acceleration. So, once the acceleration is measured, it is easy to calculate the force by knowing the mass of the accelerating body.

But there are also many types of direct measuring force transducers, called *load cells*, and they are used with instrumentation of varying complexity. Different types of load cells are as follows.

¹Fritz K. Mueller (1907–2001) was a German engineer who emigrated to USA in 1945.

1. Proving ring
2. Strain gauge load cell
3. Hydraulic load cell
4. Pneumatic load cell
5. Inductive and reluctance-based load cell
6. Magnetoelastic load cell
7. Piezoelectric load cell
8. Fibre-optic load cell
9. Resonant wire load cell

Proving Ring

The proving ring consists of an elastic ring of known diameter with a measuring device located in the centre of the ring as illustrated in Fig. 9.8(a).

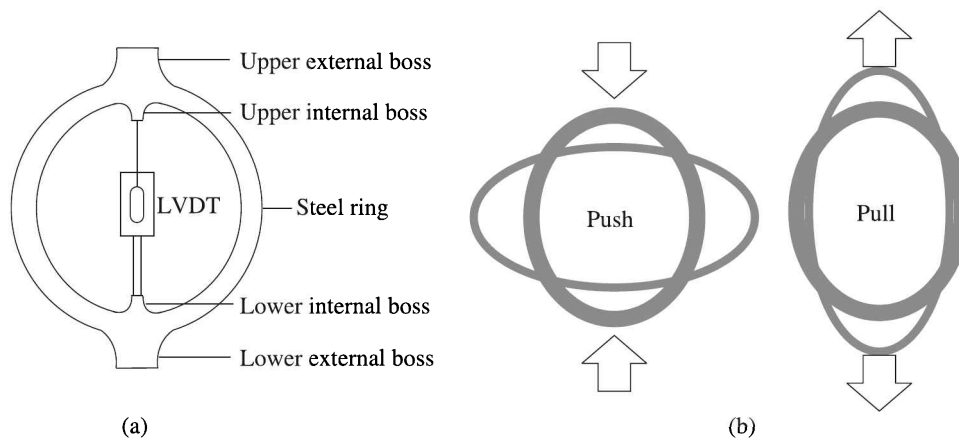


Fig. 9.8 Proving ring: (a) Schematic diagram and (b) changes, shown in an exaggerated way, in the ring diameter as compression (push) and tension (pull) forces are applied.

The proving ring consists of two main elements: the ring itself and the deflection-measuring system. Forces are applied to the ring through the external bosses (projection lugs). The resulting change in diameter, referred to as the *deflection of the ring*, is measured with an LVDT or a micrometer screw and the vibrating reed mounted diametrically within the ring. The deflection measuring arrangement is attached to the internal bosses of the ring. Nowadays the upper and lower internal and external bosses are machined as an integral part of the ring to avoid mechanical interferences during the application of the force.

Proving rings come in a variety of sizes. They are made of a steel alloy. Manufacturing consists of rough machining from annealed forgings, heat treatment, and precision grinding to final size and finish.

Proving rings can be designed to measure either compression or tension forces [see Fig. 9.8(b)]. Some are designed to measure both. The basic operation of the proving ring in tension is the same as in compression. However, tension rings are provided with pulling rods which are screwed onto the bosses.

Typically, proving rings are designed to have a deflection of about 0.84 mm to 4.24 mm. With LVDT as the deflection measuring device, forces in the range of 0.044 N to 0.44 MN (0.0045 kg to 45,000 kg) can be measured with the help of proving rings. The output may be 5 mV to 200 mV per volt of excitation. The relative measurement uncertainty can vary from 0.075% to about 0.125%.

Proving rings are cheap devices and they provide a means of measuring forces with a satisfactory degree of accuracy.

Strain Gauge Load Cell

The most common type of force transducer is the strain gauge load cell. Such a load cell for measuring compressive forces is shown schematically in Fig. 9.9.

The length of the load-sensing member is made short so that it does not buckle under the maximum allowable load and is designed to develop about 1500 μ -strains at the full-scale load.

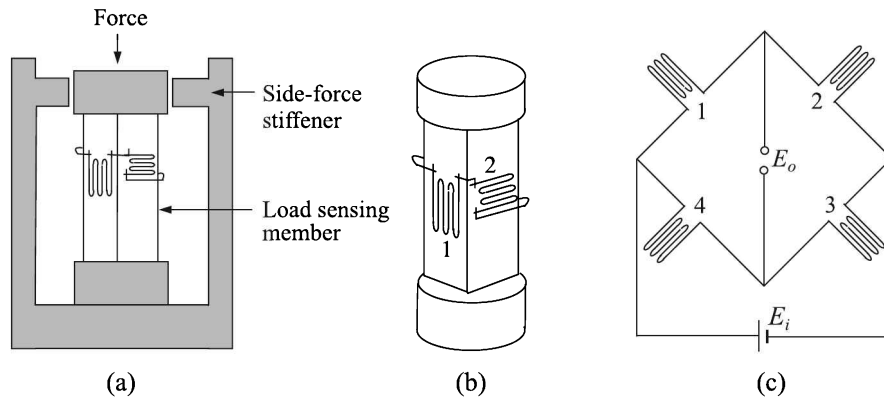


Fig. 9.9 (a) Load cell assembly. (b) Strain gauges on the load-sensing member. Gauge numbers 3 and 4 are located on opposite faces containing gauge numbers 1 and 2 respectively. (c) Wheatstone bridge arrangement.

In the arrangement shown in Fig. 9.9, gauges 1 and 3 are bonded to two opposite faces in such a way that they measure direct stress owing to the force, while gauges 2 and 4 are bonded to the other two faces at right angles to gauges 1 and 3 such that they measure the transverse stress which is related to the axial stress by the Poisson's ratio ν . This type of mounting of strain gauges is usually called the *Poisson arrangement*.

If all the strain-gauges are of equal resistance R and if the fractional change in resistance of gauge numbers 1 and 3 is $\Delta R/R$ and that corresponding to gauge numbers 2 and 4 is $\nu\Delta R/R$, it can be seen that the relation between the output voltage E_o and input voltage E_i of the bridge, shown in Fig. 9.9 (c), is given by

$$E_o = \frac{(1 + \nu)G_f \varepsilon}{2 + (1 - \nu)G_f \varepsilon} E_i = \frac{(1 + \nu)G_f \varepsilon}{2} E_i \quad [\text{since } 2 \gg (1 - \nu)G_f \varepsilon]$$

where G_f is the gauge factor and ε is the strain. Knowing Young's modulus of the load cell material, we can calculate the value of the corresponding stress, and therefore, the applied force.

Various shapes of the elastic elements (Fig. 9.10) are used in load cells. The choice of a shape depends on a number of factors including the range of force to be measured, dimensional limits, final performance and production costs.

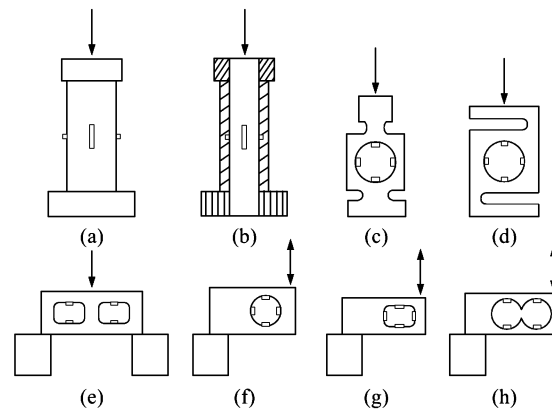


Fig. 9.10 Typical elastic elements of strain gauge load cells. Small grey rectangles indicate the positions of strain gauges and arrows indicate the directions of application of force on each element.

The capacities of strain gauge load cells range from 5 N to more than 50 MN. To give an idea of the capacity of each of the eight designs we have presented the relevant data in Table 9.1.

Table 9.1 Approximate capacities of different shapes of elastic elements

| Shape | Ref. in Fig. 9.10 | Typical capacity |
|-----------------------------|-------------------|------------------|
| Compression cylinder | (a) | 50 kN to 50 MN |
| Ditto (hollow) | (b) | 10 kN to 50 MN |
| Compression ring | (c) | 1 kN to 1 MN |
| Bending or shear S-beam | (d) | 200 N to 50 kN |
| Shear beam (double-ended) | (e) | 20 kN to 2 MN |
| Shear beam (double-bending) | (f) | 500 N to 50 kN |
| Shear beam (double-bending) | (g) | 1 kN to 500 kN |
| Shear beam (double-bending) | (h) | 100 N to 10 kN |

Usually tool steel, stainless steel, aluminium or beryllium copper is used for the fabrication of the elastic element. The aim is that the material should exhibit a linear relationship between the applied force (input) and the strain (output) with low hysteresis and low creep in the working range. There also has to be high level of repeatability. To achieve these ends, it is usual to apply a special heat treatment, like a sub-zero cycle, to the material.

The strain gauge load cell has the following advantages:

1. Since it is a full-bridge measurement, it is automatically temperature compensated.
2. Its sensitivity is $2(1 + \nu)$ times that can be achieved with a single active strain gauge in the bridge.
3. It is not sensitive to any off-centre force applied to it. Because, if the gauges are symmetrically placed, bending stresses due to off-centre force on gauges 1 and 3 will be of opposite sign and, therefore, will cancel each other.

Hydraulic Load Cell

Hydraulic load cells are force-balance devices, measuring weight as a change in pressure of the internal filling fluid. The liquid (usually oil) has a pre-load pressure. Application of the force to the loading member increases the fluid pressure which is measured by a pressure transducer or displayed on a pressure gauge dial via a Bourdon tube. A schematic diagram is shown in Fig. 9.11.

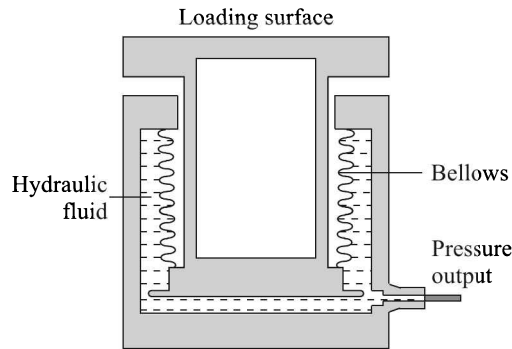


Fig. 9.11 Schematic diagram of a hydraulic load cell.

Although capacities of up to 5 MN are available, most devices are in the range of 500 N to 200 kN. A special fluid-filled hose can be used to locate the pressure gauge, that indicates the force, several metres away from the device.

Basically self-contained devices needing no external power, hydraulic load cells are inherently suitable for use in potentially hazardous areas. They can be tension or compression type devices. Measurement accuracies of around 0.25% can be achieved. The cells are sensitive to temperature changes and therefore, usually have facilities to adjust the zero output reading.

Pneumatic Load Cell

Pneumatic load cells also operate on the force-balance principle. The force is applied to one side of a diaphragm of flexible material and balanced by pneumatic pressure on the other side. This counteracting pressure is proportional to the force and is displayed on a pressure dial. Figure 9.12 shows the arrangement.

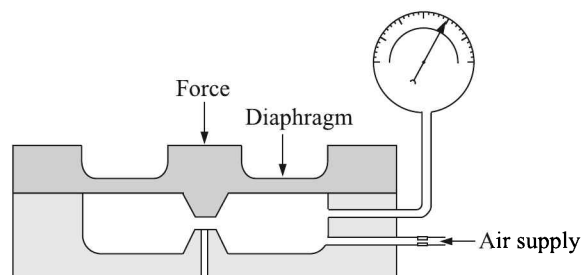


Fig. 9.12 Schematic diagram of a pneumatic load cell. The construction of the central portion limits the application of an excess force.

The sensing device consists of a chamber with a close-fitting cap. The air pressure that is applied to the chamber builds up until it is equal to the force on the cap. Any further increase in pressure will lift up the cap allowing the air to bleed around the edge until pressure equilibrium is achieved. At this equilibrium position, the pressure in the chamber is an indication of the force on the cap and can be read by the pneumatic dial-type pressure gauge.

The advantages of this type of load cell are that they are inherently explosion proof and insensitive to temperature variations. Also, they contain no fluids that might contaminate the process if the diaphragm ruptures. Disadvantages include their relatively slow speed of response and the need for clean, dry, regulated air or nitrogen.

Inductive and Reluctance-based Load Cell

These cells are based on the measurement of displacement of a ferromagnetic core caused to a force-summing device, like a diaphragm or bellows, by the applied force. The former changes the inductance of a solenoid coil due to the movement of its iron core while the latter changes the reluctance of a very small air gap.

Magnetoelastic Load Cell

We know that the magnetic permeability of a ferromagnetic material changes when subjected to a mechanical stress². The operation of the magnetoelastic load cell, a special type of which is commonly known as *pressductor*, is based on this phenomenon.

The load cell is built from a stack of ferromagnetic laminations forming a load-bearing column. A set of primary and secondary transformer coils, oriented at right angles to each other, are wound through holes in the column as shown in Fig. 9.13 (a). Coil set A is excited with an ac voltage while the output voltage is sensed by coil set B. With no load, the permeability of the material will be uniform throughout the structure. So, there will be little coupling between the coil sets and there will be no output. But when a load is applied, the corresponding force causes distortions in the flux pattern to establish a coupling between the coil sets and thus generates an output signal proportional to the applied force as shown in Fig. 9.13(b).

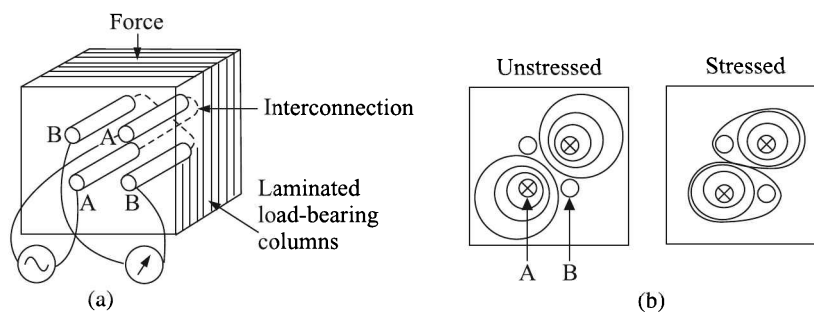


Fig. 9.13 Magnetoelastic load cell: (a) arrangement of coil sets A and B, and (b) magnetic flux patterns in no-stress and stressed conditions.

²Villari effect, see Section 5.2 at page 118.

The construction together with its housing of a special magnetoelastic load cell, known as pressductor, is shown in Fig. 9.14. Due to its sturdy construction, high signal level and small internal resistance, the pressductor can be used in rough and electrically disturbed environments such as in rolling mills.

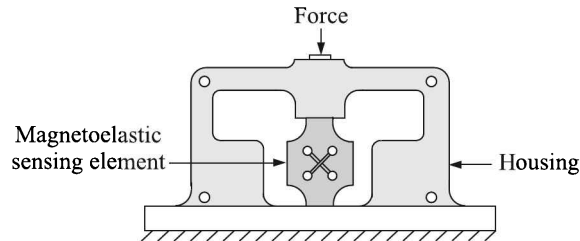


Fig. 9.14 Schematic diagram of a pressductor and its housing.

The rated capacities of these devices are in the range of 2 kN to 5 MN.

Piezoelectric Load Cell

A schematic diagram of a piezoelectric load cell is shown in Fig. 9.15(a). Usually, a pre-tensioned bolt, which allows the measurement of forces in both tension and compression, is used. Mounting of a load washer in this way is illustrated in the figure. The pre-loading ensures an optimum linearity and facilitates calibration after mounting.

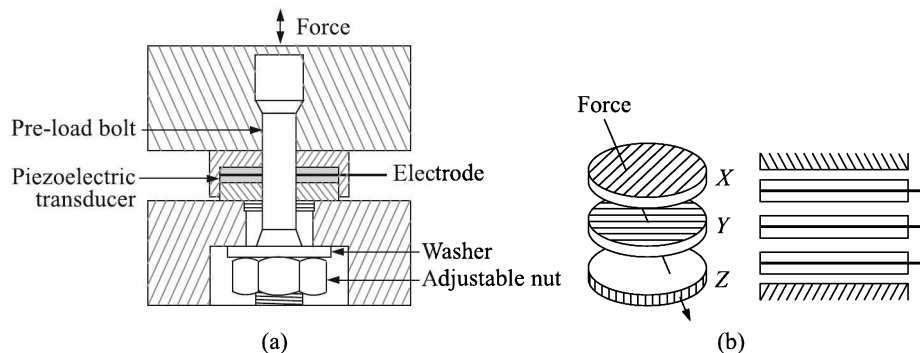


Fig. 9.15 Piezoelectric sensor used for force measurement: (a) Use of a pre-tensioned bolt, and (b) multi-component force sensing.

A small leakage of charge inherent in the charge amplifier causes a drift of the signal. So, though piezoelectric force transducers are ideal for dynamic measurements, they are not suitable for static measurements. For measurements to be made over a period, *quasi-static* measurements are resorted to.

For multi-component force measurement, a stack of transducers measures the forces along the three orthogonal axes. Figure 9.15(b) shows the operating principle of such an arrangement. The force that acts upon the stack, is transmitted to each of the three discs with the same magnitude and direction. These discs—shown ‘exploded’ in Fig. 9.15(b)—were cut along specific axes. Each produces a charge proportional to the force component specific to it. The charge is collected via the electrodes inserted into the stack.

Piezoelectric force sensors are suitable for measurements in laboratories as well as in industrial settings. The measuring range is very wide and the transducers survive high overload (typically $\sim 100\%$ of the full-scale output).

The piezoelectric crystal sensors being active sensing elements, no power supply is needed. The deformation to generate a signal is very small which has the advantage of a high frequency response of the measuring system with introducing little geometric changes to the force measuring path. When compressed under a force of 10 kN, a typical piezoelectric transducer deflects only 0.001 mm.

The high frequency response (up to 100 kHz) enabled by this stiffness makes these sensors very suitable for dynamic measurements. Extremely fast events such as shock waves in solids, or impact printer and punch press forces can be measured with these devices when otherwise such measurements might not be achievable. The sensors' small dimensions, large measuring range and rugged packaging make them very easy to use. They can operate in temperatures of up to 350°C .

Fibre-optic Load Cells

Like a wire strain gauge, a fibre-optic strain gauge can be fabricated using optical fibres. If this fibre-optic strain gauge is bonded to the elastic element of a load cell, an applied force will cause length changes in the optical fibres.

Suppose, we have two fibre-optic strain gauges bonded to two different members one of which is strained. Now, if a monochromatic light is used to feed the two gauges experiencing different strain levels then the phase difference between the two beams emerging from the gauges, in number of half wavelengths, is a measure of the applied force.

These systems have a limited temperature range from 5°C to 40°C for an overall performance better than 0.01%. The hysteresis and creep are small.

Resonant Wire Load Cell

The resonant-wire load cell consists of a taut ferromagnetic wire that is excited into resonant transverse vibrations by a drive coil. A pick-up coil detects these vibrations. The resonant frequency is a measure of the tension of the wire and hence, applied force at that instant. The arrangement is similar to that shown in Fig. 8.22 at page 304.

The advantage of the vibrating wire transducer is its direct frequency output which can be handled directly by digital electronics.

9.3 Industrial Weighing Systems

It will be in order here to have a brief discussion on the practice of weighing adopted in industries. We may divide the methods into two classes, namely

1. Manual mechanical weighing
2. Conveyor belt weighing

Manual Mechanical Weighing

Mechanical industrial weighing is based on the principle of levers. We know, there are three classes of levers as shown in Fig. 9.16. They are termed *Class I* [Fig. 9.16(a)], *Class II* [Fig. 9.16(b)] and *Class III* [Fig. 9.16(c)] levers.

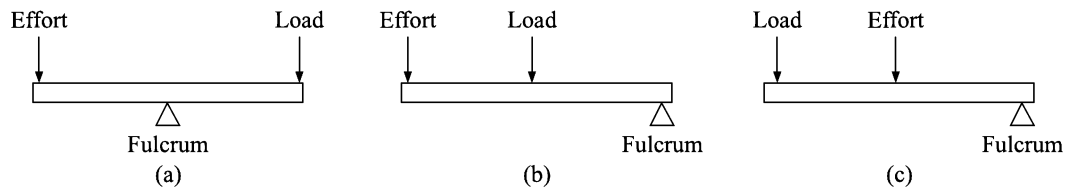


Fig. 9.16 Three classes of levers: (a) Class I, (b) class II, and (c) class III.

From the calculation of moments around the fulcrums, it is easy to figure out that Classes I and II offer mechanical advantages if the load is placed very near the fulcrum. Which means that in these classes a small *effort* can balance a large load. But, Class III does not offer any mechanical advantage. Nevertheless, it finds use because it may be *convenient* for some purpose.

Common (or Roman) steelyard

The common steelyard is a kind of weighing balance which utilises the principle of Class I lever. It consists of

1. A graduated beam on which a fixed weight E can be slid
2. A fulcrum F , and
3. A scale pan where the object L to be weighed can be placed

A schematic diagram of a steelyard is shown in Fig. 9.17.

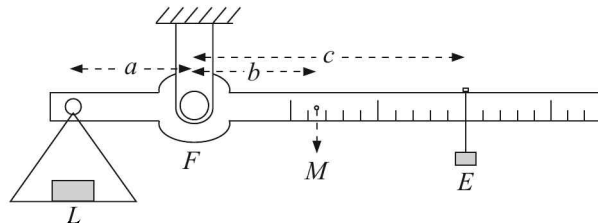


Fig. 9.17 Diagram of a common steelyard.

The weighing of the object is done by sliding the fixed weight on the graduated scale so as to make the beam horizontal. Under this condition,

$$aL = bM + cE$$

where M is the mass of the beam acting through its centre of gravity. The factor bM being constant, the load and effort relation is linear.

Platform scale

The platform scale (Fig. 9.18) utilises the link-lever mechanism.

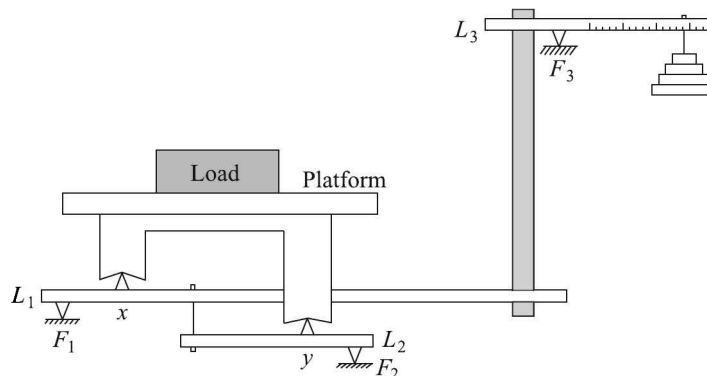


Fig. 9.18 Platform scale with link lever mechanism.

The scale basically uses three levers L_1 , L_2 and L_3 , the first two of which belong to Class II and the third to Class I. The levers move around three fulcrums F_1 , F_2 and F_3 . The weighing platform rests on knife-edges x and y on L_1 and L_2 respectively. Two links connect L_1 to L_2 and L_3 . The fulcrums of all levers are placed near the load to extract mechanical advantage at each lever.

In this case, the standard weights suspended from the right side of L_3 constitute the effort that balances the load. Small fraction of weights are measured by sliding a small weight (not shown) along the graduated beam. Instead of having weights to balance the load, the vertical rod can have rack and pinion arrangement to rotate a pointer on a dial to show the weight of the load.

Modern platform scales use load cells to generate electrical signals corresponding to the weight of the measurand and display the same by electronic means. Now, let us discuss conveyor belt weighing which finds extensive use in industries.

Conveyor Belt Weighing

In today's competitive market, reduction of material costs and improvement of product quality are of growing concern. Using conveyor belt weighing to accurately control the rate of material during packaging or blending operations can help industries improve quality and lower costs.

Conveyor belt weighing methods generally fall into two broad groups:

1. Continuous belt weighing
2. Batch weighing or weigh-feeding

Continuous belt weighing

A continuous belt weighing system consists of three elements:

1. The weigh frame, which measures the instantaneous mass
2. A tachometer which measures the belt speed
3. The belt weigher electronics which integrates both these inputs to determine flow rate and totalised weight.

The belt weigher system thus integrates *conveyor belt loading* with *conveyor belt travel* to calculate the total amount of material that has been carried past the weighing system and which also calculates the flow rate of material (almost) instantaneously using the equation

$$\text{Weight} \times \text{Speed} = \text{Rate}$$

To measure *conveyor belt loading* a weight sensitive frame is inserted into the conveyor structure which supports a section of the loaded conveyor. This weigh frame weighs a specific length of the loaded conveyor which is called the *weigh length* whose length might be measured in metres and the weigh frame is also calibrated to read weight in real units such as kilograms. As a result, the weigh frame is able to take a measurement of the kilograms per metre belt loading that happens to be the case at this instant on the loaded conveyor.

By *conveyor belt travel* is meant how many metres of belt are travelling by. To measure belt travel, a wheel or pulley in contact with the belt is used which is basically a tachometer³. As a result, at any instant or distance interval, we can measure both the distance the belt has travelled, probably calibrated in metres, and how much the material on that belt weighed, in kilograms per metre. When multiplied together, these inputs yield just the weight that has passed in the interval. The final steps are to sum this weight to a totaliser (counter) and to calculate a flow rate of the material, in some appropriate units such as tonnes per hour.

Continuous belt weighers are not entirely a separate piece of equipment but become part of an existing system. Their performance depends on a number of factors:

- Belt Tension:* Variation in belt tension can have a detrimental effect on weighing accuracy.
- Belt Stiffness:* The belt troughing angle and idler spacing both affect belt stiffness as does the obvious variable of belt composition. Variations in belt temperature have quite a significant influence on belt stiffness.
- Alignment:* Good idler alignment, in both horizontal and vertical planes, is essential to achieve accurate weighing with continuous belt weighers.
- Consistent Load:* Feeders need to be designed to provide consistent feed rate to the belt weigher if accurate weighing is required.

When a weigh frame is designed adequately, taking into account and compensating for a variety of factors which work against accurate weighing of the material passing over the weigh frame, the belt scale is a true weighing system, not merely an *indicator*. Accuracies of 0.5% FS are not uncommon for correctly matched and installed systems. There is a growing acceptance of these weighing systems as systems accuracy improves further.

Many limitations of continuous belt weighing are however overcome in weigh feeding systems, though a weigh-feeder is entirely a separate piece of equipment, not part of the ongoing process.

Weigh-feeding systems

Weigh-feeders are designed to *deliver* a designated rate of material in a process. They are used to convey, weigh, and control the flow rate of bulk materials by varying the speed of a belt conveyor. Basic components of a weigh-feeder are shown in Fig. 9.19.

³See Section 9.5 at page 359.

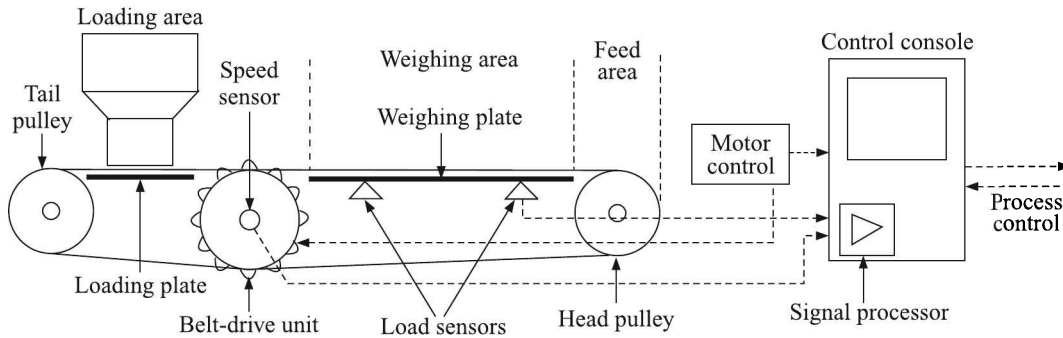


Fig. 9.19 Basic components of a weigh-feeder.

Material weight on the belt is measured by load cells, which produce a voltage signal that is sent to the integrator/controller. The integrator also receives input in the form of electronic pulses per revolution from a belt speed sensor connected to the belt-drive unit. Using these two points of data, the integrator/controller calculates the rate of material transferred along the belt (usually in pounds/kgs/tons per hour). The motor speed control is derived from a PID⁴ controller analogue signal sent to drive. This signal is calculated using actual and desired material flow rate, or load values, modified by PID parameters that are entered by the user and stored in the control console.

Weigh-feeders are often enclosed in a housing that protects or contains the material. Clean-out devices such as a dust collection port, scavenger screw, or drag chain may be included with the housing to remove any material that falls from the belt. Belt scrapers and return belt ploughs are commonly used to keep the belt clean and free from material buildup.

Batch weighing or weigh feeding is generally used for flow control purposes. The systems are generally designed for a smaller throughput than continuous weighers and are usually installed as a separate unit for feeding and weighing from a storage bin. Batch weighers or weigh-feeders generally provide accuracy in the range 1.0% to 0.1% and are accepted weighing devices by statutory authorities in many countries. Through their construction, misalignment of idlers is negligible and belt speed control is precise and responsive.

9.4 Torque Measurement

Before we describe methods of torque measurement, let us refresh our knowledge on the definition of torque and the implications of its measurement.

Definition

Torque is the time-derivative of angular momentum \mathbf{L} , just as force is the time derivative of linear momentum. Written mathematically,

$$\tau = \frac{d\mathbf{L}}{dt}$$

⁴Proportional plus Integral plus Derivative.

We know, the angular momentum of a rigid body can be written in terms of its moment of inertia J and its angular velocity ω as

$$\mathbf{L} = J\boldsymbol{\omega}$$

So, if J is constant for a body, its torque is given by

$$\boldsymbol{\tau} = J\frac{d\boldsymbol{\omega}}{dt} = J\boldsymbol{\alpha}$$

where $\boldsymbol{\alpha}$ is angular acceleration, a quantity usually measured in rad/s^2 .

Now, the work done W by a rotating system is given by

$$W = \boldsymbol{\tau} \cdot \boldsymbol{\theta}$$

where $\boldsymbol{\theta}$ is the angle (in radians) moved. Power P is the work done per unit time. So,

$$P = \frac{dW}{dt} = \boldsymbol{\tau} \cdot \frac{d\boldsymbol{\theta}}{dt} = \boldsymbol{\tau} \cdot \boldsymbol{\omega} \quad (9.10)$$

We may note that Eq. (9.10) implies that the power injected by the torque depends only on the instantaneous angular speed—not on the change of the angular speed while the torque is being applied. This also implies that the output of a power generating system, the rpm of which is kept fixed, is determined by the torque applied to the generator's axis of rotation. Therefore,

$$P \text{ (kW)} = \frac{\tau \text{ (N-m)} \times 2\pi \times \text{rpm}}{60 \times 1000} \quad (9.11)$$

Why Measure Torque

An engine is often specified by its torque. We know from Eq. (9.10) that its power output is expressed as the product of its torque and angular velocity. Internal combustion engines produce useful torque only over a limited rpm which typically varies between 800 and 6000 for a small car. The varying torque output over that range, measured with the help of a dynamometer, is shown as a torque curve. The peak of that torque curve usually corresponds to a little lower than the overall power peak. By definition, the torque peak cannot occur at higher rpm than the power peak.

The knowledge of the relationship between torque, power and engine speed is vital in automotive engineering. because it is concerned with the transmitting power from the engine through the drive train to the wheels, a proper selection of the gearing of the drive train is essential to make the most of the torque characteristics of the motor.

The electric motors mostly produce maximum torque near zero rpm. Their torque decreases as rotational speed rises because of increasing friction and other constraints. Therefore, the drive trains of these types of engines usually differ from those of internal combustion engines.

Dynamometers

As is evident from the foregoing discussion, rotational power measurement rather than simple torque measurement is more useful. The devices which measure such power are called *dynamometers*⁵. Dynamometers belong to either of the following three types:

⁵Or simply *dynos*. It should not be confused with *dino* which is an abbreviation for dinosaur.

1. Driving type
2. Absorption type
3. Transmission type

Driving type dynamometer

Driving type dynamometers are a specialised type of adjustable-speed drives. The driver (source) unit can be either an ac motor or a dc motor. The sink unit can also be an ac or dc motor which can operate as a generator. The sink is driven by the unit under test [Fig. 9.20(a)]. The control unit for an ac motor is a variable frequency drive and that for a dc motor is a dc drive. In both cases, regenerative control units can transfer power from the unit under test to the electric utility.

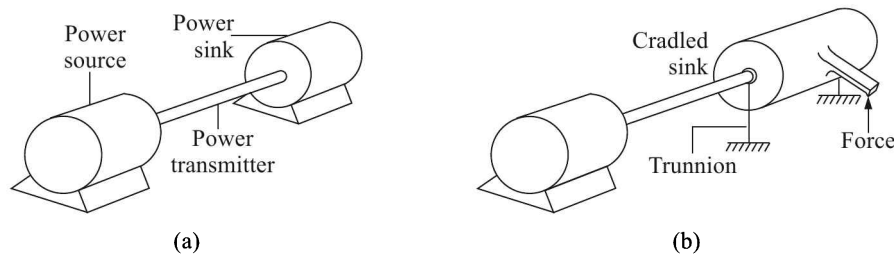


Fig. 9.20 Dynamometers: (a) Driving type, and (b) frictional absorption type.

If V is the voltage, I is the current generated in the sink and η is the efficiency of the sink motor, then the power supplied to the sink is

$$P = \eta VI$$

Driving type dynamometers are generally more costly and complex than other types of dynamometers.

Absorption type dynamometer

This type of dynamometer consists of an absorption unit, and usually includes a means for measuring torque and rotational speed. An absorption unit consists of some type of rotor in a housing. The rotor is coupled to the engine or other equipment under test and is free to rotate at whatever speed is required for the test. Some means is provided to develop a braking torque between dynamometer's rotor and housing. The means for developing braking torque can be frictional, hydraulic or electromagnetic according to the type of absorption unit.

Frictional absorption. In frictional absorption type dynamometer, the sink housing is mounted in such a manner that it is free to turn except that it is restrained by a torque arm [Fig. 9.20(b)]. This way of restricting the sink is called a *cradled sink* arrangement.

The housing can be made free to rotate by using trunnions connected to each end of the housing to support the sink in pedestal mounted trunnion bearings. The torque arm, connected to the sink housing, rests on a force balancing device, such as a load cell, so that it measures the force exerted by the sink while attempting to rotate. The torque is the force indicated by the balanced force multiplied by the length of the torque arm measured from the centre of the sink. The electrical output signal of the load cell can be calibrated in units of the torque.

Hydraulic absorption. Figure 9.21 shows the most common variable level type hydraulic absorption. Water is added until the engine is held at a steady rpm against the load. Water is then kept at that level and replaced by constant draining and refilling. That is necessary to carry away the heat generated owing to the absorption of power. The rotor has a special

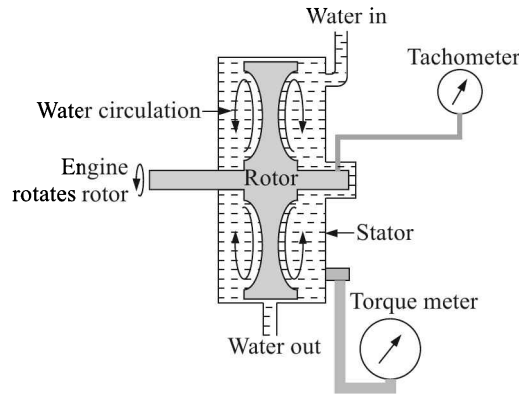


Fig. 9.21 Hydraulic absorption dynamometer.

cup shape in order to create vortices in the water and thus generate enough braking power. The housing attempts to rotate in response to the torque produced but is restrained by the scale or torque metering cell which measures the torque. The tachometer measures the rpm.

Water brake absorbers are quite common. They are noted for their high power capability, small package, light weight, and relatively low manufacturing cost. Their drawbacks are that they can take a relatively long period of time to stabilise their load amount and that they require a constant supply of water which is needed for cooling.

Electromagnetic absorption. Electromagnetic absorption is produced by eddy current generation in the toothed rotor which is attached to the shaft. The stator contains field windings excited by dc. As the rotor rotates, the eddy currents generated in it by the magnetic field of the stator windings concentrate at the teeth. This eddy current generated magnetic field in the rotor, we know from Lenz's law, opposes the magnetic field of the stator and tries to rotate the stator. The torque is measured by an arm of the stator, which is mounted on trunnions.

Eddy current dynamometers are currently the most common absorbers used in modern chassis dynos. They provide the quickest load change rate for rapid load settling. Some are air cooled, but many require external water cooling arrangements. In a properly designed system, as little as 5 amps at 220 V ac will provide approximately 150 horsepower worth of load.

Prony brake. The Prony brake is an absorption type simple dynamometer used to measure the amount of torque produced by a motor or engine in order to determine its brake power rating. The device was invented in 1821 by de Prony⁶.

In its simplest form, a Prony brake consists of a pulley wheel attached to the shaft of the motor or engine and a leather belt is held firmly against the pulley. Spring balances are attached to the two ends of the belt [Fig. 9.22(a)].

⁶Gaspard Clair François Marie Riche de Prony (1755–1839) was a French mathematician and engineer.

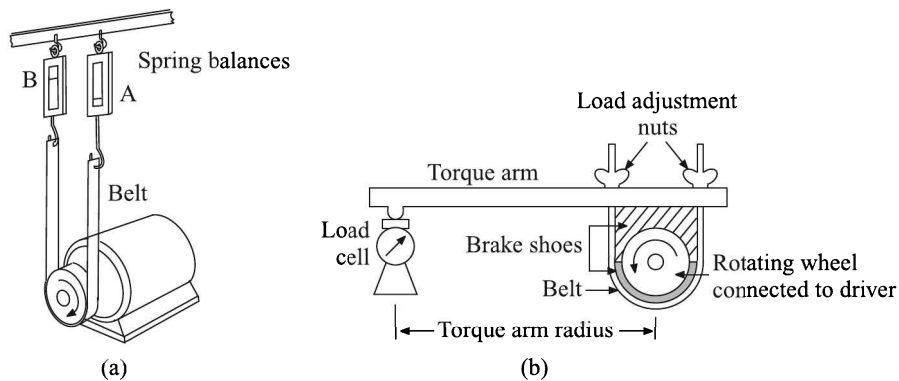


Fig. 9.22 Prony brake: (a) Simplest form and (b) another form.

When the motor is standing still, both spring balances read the same. When the pulley turns, say, in a clockwise direction, the friction between the belt and the pulley makes the belt try to move with the pulley. Therefore, the pull on balance A will be greater than the pull on balance B. The difference between the two readings is a measure of the drag force and it can be used to calculate the torque, the radius of the driven pulley being known. If the engine speed (rpm) is measured with a tachometer, the brake power is easily calculated (see Example 9.3).

An alternative set-up is to clamp a torque arm to the shaft and measure the force using a load cell [Fig. 9.22(a)]. The torque is then related to the torque arm radius and measured force. The power output may be calculated as follows:

$$P = \tau \cdot \omega = (Fr)(2\pi n) = 2\pi rFn \quad (9.12)$$

where P is the power output (N-m/s or ft-lb/s)

F is the measured force (N or lb)

r is the torque arm radius (m or ft)

n is the revolutions per second

The device can be used over a range of engine speeds to obtain power and torque curves for the engine, since for most engine types the relationship between torque and engine speed is nonlinear.

The friction caused between the rotating wheel and the static brake generates heat, which needs removal by sufficient cooling. Else the brake shoes/belt will be damaged by heat. For this limitation, Prony brakes are best suited for small speed engines.

Their other limitation is that the load is fixed by the tightening level of belt/brake shoes. Varying loads is therefore somewhat difficult. This dynamometer is also incapable of adjusting itself to the fluctuating power outputs from the engine. When the load is excessive the engine becomes prone to stall, as the brakes seem to clamp on to the wheel thus jamming its rotation.

These difficulties notwithstanding, this cheap and easy dynamometer continues to be widely used by tractor and heavy equipment manufacturers.

Example 9.3

In a simple Prony brake, when the motor is standing still, both the spring balances read 12 lb. The diameter of the pulley is 1.5 ft. When the motor starts rotating at 1800 rpm, one balance reads 22 lb and the other 2 lb. Calculate the horse power (HP) of the motor.

Solution

From the given data, we observe that the drag force F on the motor when it is rotating, is $(22 - 2) = 20$ lb. Revolutions per second n is $(1800 \div 60) = 30$. Therefore, from Eq. (9.12) we get

$$\begin{aligned} P &= 2\pi \left(\frac{1.5}{2} \right) (20)(30) \text{ ft-lb/s} \\ &= \frac{2\pi(0.75)(20)(30)}{550} \text{ HP} \\ &= 5.14 \text{ HP} \end{aligned}$$

Transmission type dynamometer

We know that the relation between the torque and the parameters of a solid cylindrical shaft under shear stress is given by⁷

$$\tau = \frac{G\pi r^4 \phi}{2l} \quad (9.13)$$

where G is the modulus of rigidity of the shaft material

r is the radius of the shaft

ϕ is the angle of deflection

l is the length of the shaft

In case of a hollow cylinder of outer radius r_1 and inner radius r_2 , the relation is

$$\tau = \frac{G\pi\phi}{2l} (r_1^4 - r_2^4) \quad (9.14)$$

In the transmission type dynamometer, the torque is measured either by sensing the shaft deflection angle ϕ caused by a twisting force, or by detecting the effects of this deflection on transducers like strain gauges. If strain gauges are utilised to measure tensile strain caused by the shear, then the relations for a solid shaft are

$$\varepsilon_{45} = \frac{\theta}{2} \quad (9.15)$$

$$\sigma_s = G\theta$$

$$\therefore \tau = \frac{\pi}{2} \sigma_s r^3 = \pi G \varepsilon_{45} r^3 \quad (9.16)$$

where ε_{45} is the tensile strain at an angle 45°

θ is the shear strain

σ_s is the shear stress

⁷See, for example, *The General Properties of Matter*, FH Newman and VHL Searle, Edward Arnold, p. 109.

In the case of a hollow cylindrical shaft, the equation for the torque is

$$\tau = \pi G \varepsilon_{45} \left(\frac{r_1^4 - r_2^4}{r_1} \right) \quad (9.17)$$

The deflection measuring system is called the *torsion meter* and the strain monitoring system the *torque meter*.

Torsion meter. As discussed earlier, either of Eqs. (9.13) and (9.14), as the case may be, is utilised to measure the torque by measuring the deflection angle ϕ between two points of the shaft situated at a specific distance.

The deflection angle can be measured by a number of methods. One of them is to attach a pair of toothed wheels to the shaft at a certain distance apart. Two proximity sensors fixed on top of the toothed wheels (Fig. 9.23) will produce output voltages with phase difference proportional to the torque.

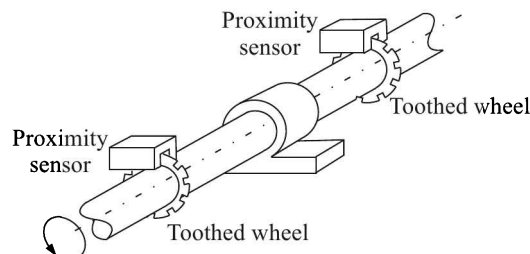


Fig. 9.23 Torsion measurement with proximity sensors.

Alternatively, an IR beam may be transmitted through an optical fibre to specially designed optical forks each of which can be positioned at each of the wheels such that the transmitter and receiver stay on two sides of the wheel. The IR beams passing through the air gaps get pulse modulated by the coding wheels mounted on the rotating shaft. The generated pulse pattern will depend on the shaft torque and speed.

The resulting IR light signal in the optical fibre contains the information on the torsional angle between the two toothed wheels on the rotating shaft. In addition, the IR light signal contains the information on the shaft rpm.

Example 9.4

A torsion meter is used to measure the torque and the power required to drive a compressor. The diameter of the shaft transmitting power is 15 cm. Modulus of rigidity of the shaft material is $8 \times 10^{10} \text{ N/m}^2$. The angle of twist of the shaft is measured by measuring the time interval between the pulses obtained from two appropriately shaped discs placed at a distance of 1.5 m on the shaft by photoelectric sensors. When there is no torque transmitted, the time interval between the pulses from the two discs is zero. If the time interval measured between the pulses from the two photoelectric sensors is 0.4 milliseconds and the shaft speed is 800 rpm, determine

- The torque transmitted to the compressor
- The power transmitted to the compressor.

Solution

Given: $d = 15 \text{ cm} = 0.15 \text{ m}$, $G = 8 \times 10^{10} \text{ N/m}^2$, shaft speed $N = 800 \text{ rpm}$, length of the shaft between two monitoring points $l = 1.5 \text{ m}$ and the time interval $\Delta t = 0.4 \text{ ms}$. The shaft describes $\frac{2\pi(800)}{60}$ rad in 1 s. So, the angle of deflection corresponding to a time delay of 4 ms is

$$\phi = \frac{2\pi(800)(4 \times 10^{-3})}{60} = 0.335 \text{ rad}$$

(a) Therefore, we get from Eq. (9.13)

$$\tau = \frac{(8 \times 10^{10})(0.335)\pi(0.15)^4}{32(1.5)} \cong 888 \text{ kN-m}$$

(b) From Eq. (9.10), we get

$$P = \frac{2\pi(800)}{60} \tau = \frac{2\pi(800)(888 \times 10^3)}{60} \cong 93 \text{ kW}$$

Torque meter. Torque meters can be of various types. We consider three of them, namely

1. Strain gauge type
2. Magnetostrictive type
3. Magnetoelastic type

Strain gauge type. In strain gauge type torque meters, strain gauge elements are usually mounted in pairs, each subtending an angle of 45° to the shaft axis⁸, on one side of the shaft. One gauge measures the increase in length (in the direction in which the surface is under tension) and the other measures the decrease in length in the other direction. Another similar pair is mounted on the opposite side of the shaft. The arrangement is shown in Fig. 9.24. In Fig. 9.24(a), which shows the view from a side, grey coloured strain gauges indicate that they are mounted on the opposite side of the shaft. The cross-sectional view of the arrangement is shown in Fig. 9.24(b).

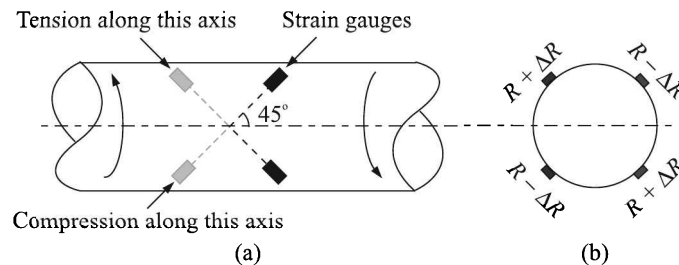


Fig. 9.24 Torsion measurement with strain gauge sensors: (a) side view where grey coloured gauges indicate that they are mounted on the opposite side of the shaft, and (b) the cross-sectional view of the arrangement.

Because the shaft is rotating, the torque sensor can be connected to its power source and signal conditioning electronics via a transformer. The excitation voltage for the strain gauge is inductively coupled, and the strain gauge output is converted to a modulated pulse frequency (Fig. 9.25).

⁸A 45° angle is chosen because at that angle the shear strain bears a simple relation with the tensile strain [see Eq. (9.15)].

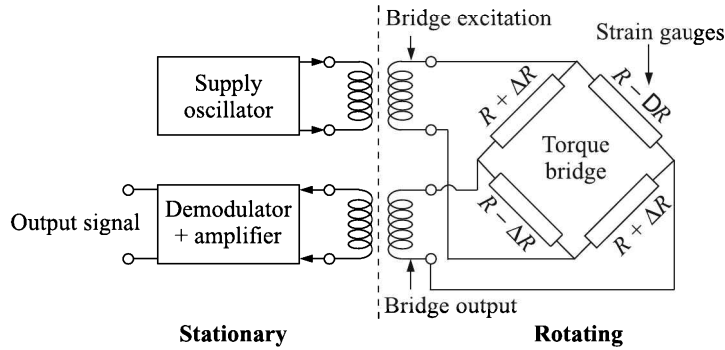


Fig. 9.25 Inductive coupling of the rotating Wheatstone bridge of strain gauges to power source and output electronics.

Maximum speed that such an arrangement can handle is 15,000 rpm. The system is susceptible to noise and errors induced by the alignment of the transformer primary-to-secondary coils. Because of the special requirements imposed by the rotary transformers, specialised signal conditioning is also required in order to produce a signal acceptable for most data acquisition systems.

Instead of the rotary transformer, the infrared torque sensor is often utilised as a contactless method of getting the torque signal from a rotating sensor back to the stationary world. The IR beam is used to power a circuit on the rotating sensor. The circuit provides excitation voltage to the strain gauge bridge, and digitises the output signal. This digital output signal is then transmitted, via infrared light, to stationary receiver diodes, where another circuit checks the digital signal for errors and converts it back to an analogue voltage.

Example 9.5

A circular steel shaft is to transmit power up to 40 kW at a constant speed of 20 rps. It is proposed that the torque be sensed by a pair of strain gauges bonded to the shaft as shown in Fig. 9.24. Assume that the maximum strain value of the gauges is 0.0015 and the maximum allowable stress on the shaft is 350 MPa. Calculate the diameter of the shaft if the modulus of elasticity of the shaft is $200 \times 10^9 \text{ N/m}^2$.

Solution

Given: power $P = 40 \times 10^3 \text{ W}$, speed of rotation $N = 20 \text{ rps}$, and $\sigma_s = 350 \times 10^6 \text{ Pa}$. So, from Eq. (9.10)

$$\tau = \frac{P}{2\pi N} = \frac{40 \times 10^3}{2\pi(20)} = \frac{10^3}{\pi} \text{ N-m}$$

Now, from Eq. (9.16)

$$\tau = \sigma_s \left(\frac{\pi}{16} d^3 \right)$$

which gives

$$d = \sqrt[3]{\frac{16\tau}{\pi\sigma_s}} = \sqrt[3]{\frac{16 \times 10^3}{\pi^2 \times 350 \times 10^6}} = 17 \text{ mm}$$

Magnetostrictive type. The magnetic permeability of the shaft varies with torque. This phenomenon can be utilised to measure torque using a magnetostrictive sensor. With no loading, the permeability of the shaft is uniform. Under torsion, permeability and the number of flux lines increase in proportion to torque. This type of sensor can be mounted to the side of the shaft using two primary and two secondary windings. Alternatively, it can be arranged with many primary and secondary windings on a ring around the shaft.

Magnetoelastic type. In a magnetoelastic type torque sensor, the changes in the magnetic field produced by the sensor owing to changes in its permeability, are measured.

Such a sensor is constructed as a thin ring of steel tightly coupled to a stainless steel shaft. This assembly (Fig. 9.26) acts as a *permanent magnet* whose magnetic field is proportional to the torque applied to the shaft. The shaft is connected between a drive motor and the driven device. A magnetometer senses the generated magnetic field and converts it into an electrical output signal that is proportional to the torque being applied.

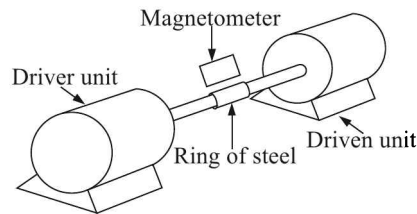


Fig. 9.26 Magnetoelastic torque meter. The shaft and the steel ring assembly is a permanent magnet.

9.5 Tachometers

We have seen before [see Eq. (9.10) at page 351] that to measure rotational power, we need to know the rotational speed or rpm of the rotating system. Tachometers⁹ are used to measure rpm or rotational speed of say, a rotating shaft. They can also be used to measure the flow rate of liquid or gas by attaching them to wheels with inclined vanes.

We describe a few tachometers categorising them according to their method of sensing the angular speed as follows:

1. Fly-ball tachometer
2. Tachogenerator
3. Eddy current tachometer
4. Inductive pulse tachometer
5. Hall effect pulse tachometer
6. Optical pulse tachometer
7. Stroboscope
8. Digital tachometer

We describe them in that order.

⁹The word is derived from Greek *tachos* meaning *speed* and *metron* meaning *to measure*.

Fly-ball Tachometer

A mechanical fly-ball consists of two metal balls at the end of two arms hinged at the top of a *vertical* shaft as shown in Fig. 9.27.

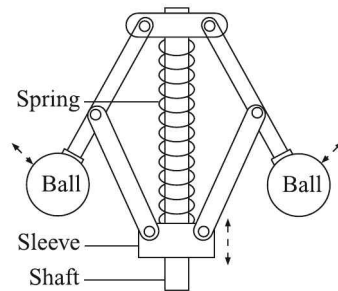


Fig. 9.27 Mechanical fly-ball.

According to the rotational speed of the shaft, the balls fly apart owing to a centrifugal force acting on them. The two arms holding the balls are connected by two links to a sleeve which slides up and down as the shaft rotates with different speeds. The movement of the sleeve is restricted by a spring as shown in the figure. At equilibrium, the centrifugal force and the force of restitution offered by the spring balance. Therefore, the position of the sleeve on the shaft can give a visual information of the rotational speed positioning a pointer on a scale. Position sensors can be added to convert the pointer position to an electrical signal if this is needed.

The force of restitution exercised by the spring is given by

$$F_r = kx$$

where k is the spring constant and x is the displacement of the spring end.

Centrifugal force acting on constant masses is given by

$$F_c = m\omega^2 r$$

where m is the mass of balls

ω is the rotational speed of balls

r is the radius of rotation of balls at equilibrium

Equating F_r and F_c , we get

$$x = \left(\frac{mr}{k} \right) \omega^2 \quad (9.18)$$

Equation (9.18) shows that the fly-ball tachometer has a nonlinear speed scale. In some designs the nonlinearity is compensated for by using a spring of nonlinear response.

Apart from the nonlinearity, the threshold of the fly-ball tachometer is considerable. Basically, it is used as a switch to control rotational speed rather than measuring the same.

Tachogenerator

Tachometer generators or *tachogenerators* are small ac or dc generators that output a voltage in proportion to the rotational speed of a shaft. They are capable of measuring the speed

and direction of rotation, but not position. They convert a rotational speed into an isolated analogue voltage signal that is suitable for remote indication and control applications.

The ac tachogenerator

The tachogenerator shown in Fig. 9.28 comprises two stator coils at right angles to each other, and an aluminium or copper cup rotor which rotates around a stationary, soft-iron, magnetic core.

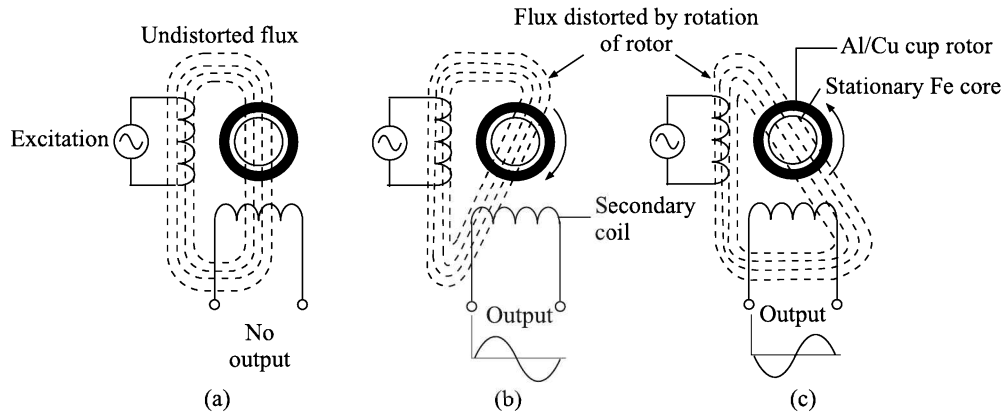


Fig. 9.28 The ac tachogenerator showing coils, rotor and core. Magnetic flux and output: (a) when the rotor is stationary, (b) when the rotor turns clockwise and (c) when the rotor turns counter-clockwise.

The primary stator coil is excited by an ac source. The secondary coil gives the generator output.

Voltage output. The voltage applied to the primary coil produces a magnetic flux which cuts the secondary coil at right angles when the rotor is stationary, as shown in Fig. 9.28(a). As a result, the output is zero.

When the rotor turns by a mechanical linkage from the load, the magnetic field gets distorted so that the flux is no longer 90 electrical degrees from the secondary. Flux lines cut the secondary coil at some other angle, and a voltage is induced in the output coil as shown in Figs. 9.28(b) and (c). The amount of magnetic flux that will be distorted is determined by the speed of the rotor. Therefore, the magnitude of the voltage induced in the secondary coil is proportional to the rotor's speed.

Output phase and frequency. The direction of the distortion of the magnetic field is determined by the direction of the motion of the rotor. If the rotor turns, say clockwise, the lines of flux will cut the secondary coil like that shown by dotted lines in Fig. 9.28(b). If the rotor turns counter-clockwise, the flux will cut the secondary coil as shown in Fig. 9.28(c). As a result, the phases of the voltage induced in the secondary coil, measured with respect to the phase of the excitation voltage, will differ. While the phase will be in sync with the excitation for the clockwise rotation, it will differ by 180° for the counter-clockwise rotation. The phases are indicated at the output coil in Figs. 9.28(b) and (c). So, by determining the phases, the directions of motion of the rotor can be ascertained.

The frequency of the tachometer generator output voltage will be the same as that of the excitation voltage.

The other type of ac tachogenerator has a squirrel-cage rotor. Otherwise its construction and principle of operation are identical to the one described above.

The dc tachogenerator

In the dc tachogenerator a permanent magnet is used to provide the magnetic flux wherein the coil attached to the shaft rotates (Fig. 9.29).

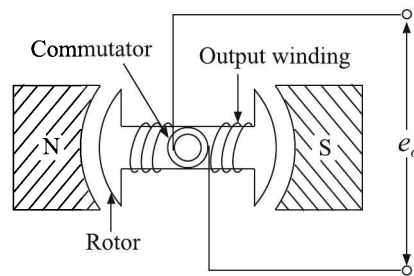


Fig. 9.29 Schematic diagram of the dc tachogenerator.

The commutator converts the ac output to a dc voltage which is measured with a voltmeter having a high input impedance. Although one pair of poles of magnets and one winding of conductor is shown, multiple poles and windings can be used to increase the output. The generated output e_o is given by

$$e_o = \frac{pc\Phi\omega}{60N_{||}} \times 10^{-8} \text{ V} \quad (9.19)$$

where

- p is the number of poles of magnets
- c is the number of conductors in the armature
- Φ is the flux per pole
- ω is the number of rotations per minute
- $N_{||}$ is the number of parallel paths between positive and negative brushes in the commutator

Other factors remaining constant for a construction, Eq. (9.19) shows that

$$e_o \propto \omega$$

Alternatively, a permanent magnet may be attached to the rotor and a stationary coil may be used to sense the rpm as shown in Fig. 9.30.

Eddy-current Tachometer

The eddy-current tachometer aka *drag-cup tachometer* is not a tachogenerator, i.e. it does not produce a voltage proportional to the rpm of the rotor. Instead, it provides a visual indication of the speed of rotation by means of a pointer and scale. The drag-cup tachometer is a common device and it finds use as a speed and rpm indicator in automobiles, airplanes and other vehicles.

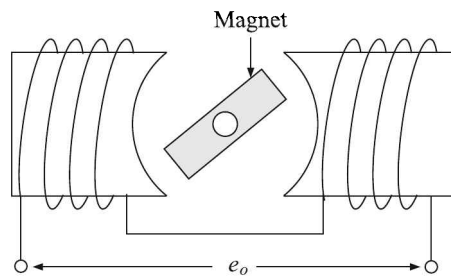


Fig. 9.30 Alternative form of dc tachogenerator.

An exploded view of the drag-cup tachometer is shown in Fig. 9.31.

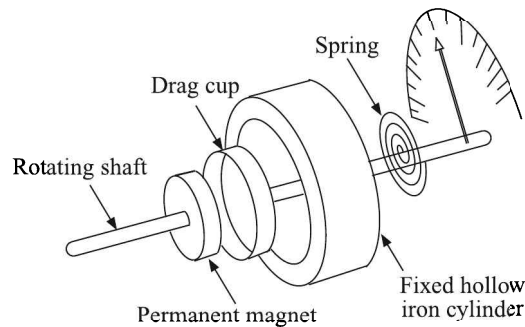


Fig. 9.31 Drag-cup tachometer (exploded view).

It consists of a rotating cylindrical permanent magnet driven by the rotating shaft. The magnet rotates within a hollow conductive sleeve in the form of a cup. The cup, which is usually made of aluminium, can also rotate, but its rotation is restrained by a spring. Surrounding the conductive cup, a fixed hollow iron cylinder completes the magnetic circuit.

As the shaft with the magnet rotates, its revolving magnetic field induces eddy-currents in the conductive cup. The amplitudes of these currents are proportional to the time-derivative of inductive flux, i.e. proportional to the speed of rotation. These currents, in turn, will produce magnetic fields that will interact with the rotating field of the permanent magnet. This interaction will generate an electrodynamic torque that is proportional to

1. The amplitude of the field of the permanent magnet, and
2. The amplitude of eddy-currents

The former being constant, the torque, which is proportional to the speed of rotation, will be proportional to the amplitude of eddy-currents.

The electrodynamic torque causes a displacement of the cup to a position where the resisting torque of the spring balances it. It can be shown that the relation between the rpm and the cup displacement is nearly linear.

The permanent magnet in a drag-cup tachometer is usually made of Al-Ni alloy and can have up to five pairs of poles. These tachometers are used for measuring speed up to 10,000 rpm and the accuracy is on the order of 1%.

Inductive Pulse Tachometer

The inductive pulse tachometer consists of a toothed wheel driven by the rotating shaft. The rotation speed of the wheel is sensed by a magnetic pick-up which consists of a permanent magnet around which a coil is wound. The pick-up is placed in close proximity—about 1 mm away—of the wheel. A schematic diagram is shown in Fig. 9.32(a).

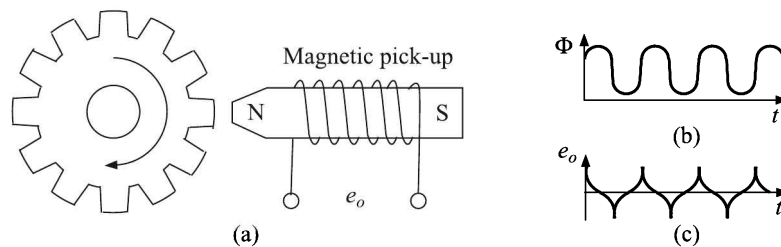


Fig. 9.32 Inductive pulse tachometer: (a) Schematic diagram, (b) flux change at the pick-up coil, and (c) output voltage.

As the wheel rotates, the teeth and gaps pass successively by the pole of the magnet causing a periodic variation of the flux linkage of the coil as shown in Fig. 9.32(b). This variation of flux linkage generates a voltage in the coil according to Faraday's law

$$e_o = -\frac{d\Phi}{dt} \quad (9.20)$$

The voltage variation is shown in Fig. 9.32(c). Both the amplitude and frequency of the voltage are proportional to the number of teeth passing per second, which, in turn, is proportional to the rpm of the wheel.

The rotation speed can be obtained from the measured voltage—its peak or time-average or rms value. But this method is fraught with the problems of variation of the voltage amplitude owing to temperature variation or position of pick-up relative to the teeth. Also, as Eq. (9.20) reveals, the voltage will be low for slow rotations.

The alternative method, which is used generally, is measuring the frequency f of the generated ac voltage which is given by

$$f = \frac{nN}{60}$$

where n is the number of teeth and N is the rpm of the wheel. By placing a second pick-up in quadrature, the direction of rotation can also be ascertained from the phase of pulses.

The pulses may be made sharper by using a Schmidt trigger and may be measured analogically or digitally. The problem of low amplitude of pulses for low speeds persists, however. By increasing the number of teeth, though the time derivative factor of Eq. (9.20) will increase and hence the voltage amplitude, but then smaller teeth width will decrease the flux linkage Φ .

Being a non-contact way of measuring rpm, the tachometer can be used in low torque drives. The torque requirement can be further lowered by using a light plastic wheel with embedded ferromagnetic material in place of teeth.

Inductive pulse tachometers are low cost, almost maintenance-free and require no power supply and therefore no sparks. So, they can be used in hazardous areas as well.

Hall Effect Pulse Tachometer

The Hall effect pulse tachometer construction is similar to that of an induction pulse tachometer except that instead of a magnetic pick-up, here a Hall effect sensor is used to sense the variation in the magnetic field.

The magnetic field can be provided by a stationary permanent magnet within which the toothed ferromagnetic wheel attached to the rotating shaft rotates [Fig. 9.33(a)]. The airgap variation caused by the successive passage of the teeth and gaps of the wheel, causes a variation of the intensity of the \mathbf{B} -field over the Hall probe. The output voltage of the Hall element, being proportional to the intensity of the \mathbf{B} -field, follows this variation.

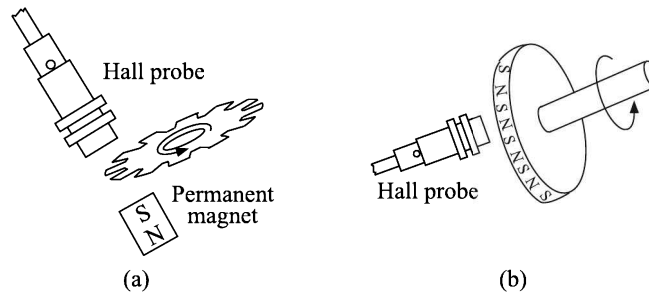


Fig. 9.33 Hall effect pulse tachometer: (a) stationary permanent magnet, and (b) magnets embedded in the rotating wheel.

Alternatively, a wheel with magnetic poles in its periphery [see Fig. 9.33(b)] can be used and the stationary magnet can be avoided. The Hall probe will sense the magnetic field variation from encountering the wheel's poles.

We note here that in inductive pulse sensors, it is the time derivative of the magnetic flux and hence $d\mathbf{B}/dt$, that is sensed while with Hall sensors, it is \mathbf{B} itself that is sensed. As a result, Hall effect pulse tachometers can sense all speeds equally. This is an additional advantage of Hall effect pulse tachometers over their inductive pulse counterparts. However, they have the disadvantage of needing a constant current source.

When compared with optical pulse tachometers, Hall effect pulse tachometers have the advantage of being less sensitive to environmental conditions like humidity, dust, vibration, and having a characteristic that suffers less drift with time.

Typical commercial Hall probes consume current on the order of 10 mA and have an output sensitivity of 10 volt/tesla. To know the direction of rotation, another probe may be used in quadrature like in inductive pulse tachometers.

Optical Pulse Tachometer

Like the Hall effect pulse tachometer, an optical pulse tachometer generates a train of constant amplitude pulses whose frequency is proportional to the speed of rotation that is to be measured. A schematic diagram of the optical pulse tachometer is shown in Fig. 9.34.

In Fig. 9.34(a) we have shown that pulses are generated by a light sensor when an incident light falls on it after passing through the transparent windows on the periphery of an opaque disc. The disc is mounted on the rotating shaft.

An alternative way [Fig. 9.34(b)] of generating pulses is by way of reflection from a disc that consists of alternate light reflecting and light absorbing zones on its periphery and the light

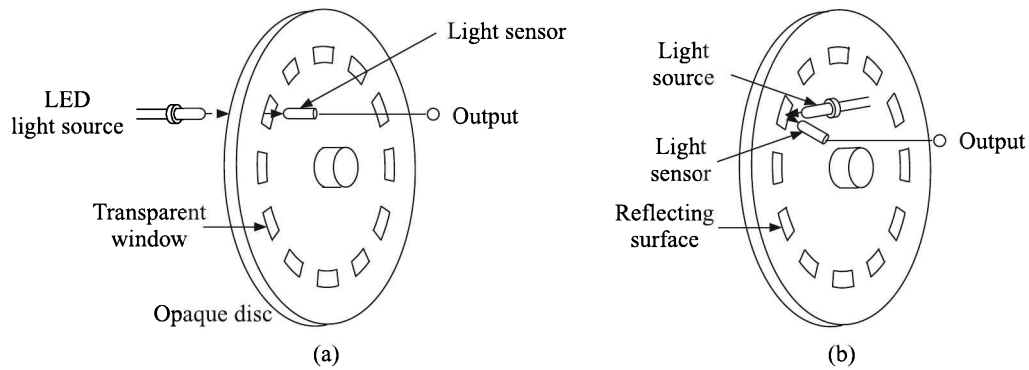


Fig. 9.34 Schematic diagram of optical pulse tachometer: (a) Transmitted-light type, and (b) reflected-light type.

source and detector are placed on the same side of the disc. Although the latter method is easier to install, it is more prone to errors caused by vibration, misalignment, dust accumulation on the reflecting surface, etc.

Like other pulse tachometers, optical pulse tachometers do not indicate the direction of rotation per se. If that is needed, a second optical pulse tachometer needs to be installed in quadrature for phase detection and direction determination.

In general, optical pulse tachometers are cheap, but they are susceptible to environmental pollution, dust and ageing.

Stroboscope

The stroboscope method of determination of the speed of rotation is based on the persistence of vision of a human eye. Therefore, a human observer is necessary in this method.

If a rotating or vibrating object is illuminated by a flashing light whose flashing rhythm is the same as the speed of rotation or vibration of the object, the object appears stationary to a human observer. This happens because a human eye cannot distinguish between two successive flashes if they are too rapid and so it generates a sensation that the object is standing still.

A stroboscope consists of a gas discharge lamp in which a high voltage is applied between its cathode and anode from a capacitor that is periodically charged (Fig. 9.35). Voltage pulses from a pulse generator is applied to the grid. Each pulse generates an electrical discharge

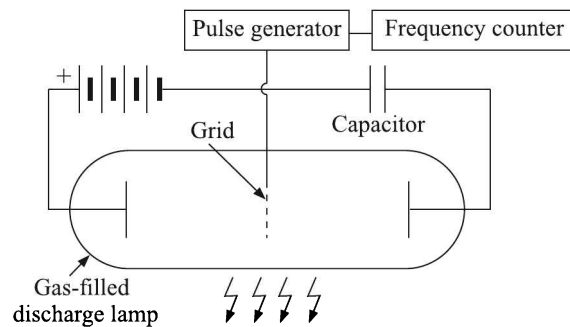


Fig. 9.35 Schematic diagram of stroboscope.

from the capacitor through the gas, resulting an emission of an intense flash of a short duration of about $1 \mu\text{s}$. The capacitor gets recharged before the next pulse is applied to the grid. The frequency of pulses is counted by the frequency counter associated with the pulse generator.

Suppose, we observe that the object appears frozen at a frequency f_1 . It does not necessarily mean that the speed of rotation object is the same, because it may appear frozen at a higher harmonic of the fundamental speed of rotation. If N is the rotations per second, let us assume

$$N = kf_1$$

or

$$\frac{N}{f_1} = k \quad (9.21)$$

where k is a constant.

Now, by adjusting the flash rate of the stroboscope we find that the next lower frequency is f_2 when the rotating object is again frozen. Then,

$$N = (k - 1)f_2$$

or

$$\frac{N}{f_2} = k - 1 \quad (9.22)$$

Subtracting Eq. (9.22) from Eq. (9.21), we get

$$\frac{N}{f_1} - \frac{N}{f_2} = k - (k - 1) = 1$$

$$\Rightarrow \frac{1}{f_1} - \frac{1}{f_2} = \frac{1}{N}$$

$$\Rightarrow N = \frac{f_1 f_2}{f_1 - f_2} \quad (9.23)$$

Equation (9.23) helps us determine the speed of rotation of the object unambiguously.

Stroboscope is an easy, straightforward and mobile device of determining the speed of rotation without the hassle of its mechanical setting. However, it has the disadvantage of not providing an automated output. It is available in ranges from 5 Hz (300 rpm) to 417 Hz (25,020 rpm). The accuracy is on the order of 1%.

Digital Tachometer

All tachometers so far described generate analogue output voltages. These voltages can be converted to produce digital values with the help of analogue-to-digital converters. But as discussed in Section 6.6 at page 216, they are not digital tachometers per se. Digital tachometers utilise rotary encoders aka *shaft encoders* to sense the speed of rotation.

Rotary encoders can be of three types as discussed in Section 6.6. Of them, the output of incremental encoders provides information about the motion of the shaft which is typically further processed elsewhere into information such as direction of rotation, speed and rpm. Absolute encoders are more sophisticated. Their output indicates the current position of the shaft, making them angle measurement transducers. So, we will discuss only incremental encoders, which are mostly used as tachometers, here.

An incremental rotary encoder, or quadrature encoder or relative rotary encoder, has two outputs called *quadrature outputs*. They can be either mechanical or optical (see Fig. 9.36). In the optical type, there are two Gray coded tracks, while the mechanical type has two contacts [one contact shown in Fig. 9.36(a)] that are actuated by cams on the rotating shaft. The mechanical type requires debouncing¹⁰ and is typically used as digital potentiometers on equipment including consumer devices.

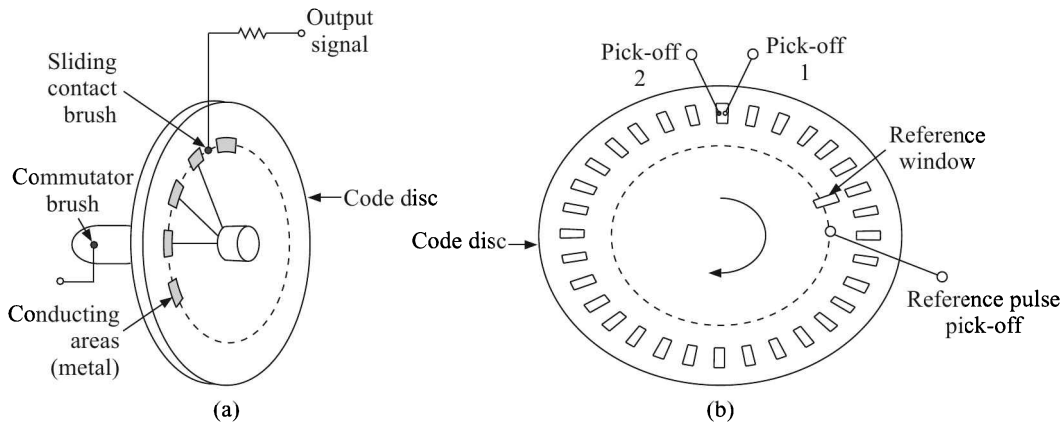


Fig. 9.36 Rotary incremental encoder: (a) Mechanical type, and (b) optical type.

The use of only two sensors in incremental encoders does not compromise their accuracy. Incremental encoders with up to 10,000 counts per revolution, or more are available commercially. The direction can be determined and very accurate measurements can be made. There can be an optional third *reference* output, which is produced once per revolution. This is necessary, say, in positioning systems where an absolute reference is required.

The mechanical type can handle limited rotational speeds. The optical type is used when higher rpm is to be measured or a higher degree of precision is required.

The two outputs from incremental encoders are usually termed *A* and *B*. These are called *quadrature outputs*, as they are 90° out of phase. The state tables for clockwise and counter-clockwise rotations are shown as follows.

| Gray coding for CW rotation | | | Gray coding for CCW rotation | | |
|-----------------------------|---|---|------------------------------|---|---|
| State | A | B | State | A | B |
| 1 | 0 | 0 | 1 | 1 | 0 |
| 2 | 0 | 1 | 2 | 1 | 1 |
| 3 | 1 | 1 | 3 | 0 | 1 |
| 4 | 1 | 0 | 4 | 0 | 0 |

¹⁰Contact bounce (aka *chatter*) is a common problem with mechanical switches. Switch contacts are usually made of springy metals that are forced into contact by an actuator. When the contacts strike together, their momentum and elasticity act together to cause bounce. The result is a rapidly pulsed electric current instead of a clean transition from zero to full current. Contact circuits can be filtered to reduce or eliminate multiple pulses. This is called *debouncing*.

The quadrature term means that the two output wave forms are 90° out of phase (see Fig. 9.37). These signals are decoded to produce a count up pulse or a count down pulse. The A and B outputs are read and the above tables are used to decode the direction. For example, if the last value was 00 and the current value is 01, the device has moved one half step in the clockwise direction.

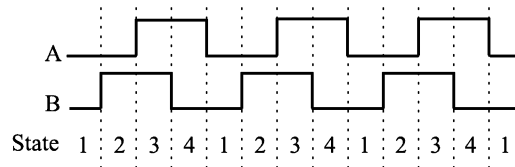


Fig. 9.37 Two square waves in quadrature for clockwise rotation.

As told in Section 6.6 at page 216, rotary encoders with a single output cannot sense direction, but can sense the rpm. They are called *tachometer encoders*.

Review Questions

- 9.1 A transducer that measures force has a normal resistance of $300\ \Omega$, forms a four arm strain gauge bridge and is excited by 7.5 V dc. When the force of 0.1 N is applied, all the four strain gauge resistances are changed by $5.2\ \Omega$. Find the output voltage and determine its sensitivity.
- 9.2 Four strain gauges are used to measure the torque of a cylindrical shaft.
 - (a) Draw a labelled diagram showing the arrangement of gauges on the shaft and the bridge configuration.
 - (b) Calculate the maximum bridge output for a strain of 500×10^{-6} . The gauges have resistance of $120\ \Omega$ each, and a gauge factor of 2.1. The maximum permissible gauge current is 50 mA.
- 9.3 Find the displacement error in mm of a seismic displacement pickup having a damping ratio of 0.8 and natural frequency of 8 Hz. The relative displacement of the mass is measured as 1.5 mm.
- 9.4 A hollow circular steel shaft (shear modulus $G = 8 \times 10^{10}\ \text{N/m}^2$) with outer and inner radii of 33 mm and 25 mm respectively, has a length of 150 mm. It is transmitting a torque of T N-m. The strain indicated by a strain gauge fixed on the outer periphery at an angle of 45° to the axis of the shaft is 5.5 microstrain (micrometre per metre). Estimate the value of T and the angular deflection of the shaft.
- 9.5 A seismic accelerometer is used to measure linear acceleration over a range of $30\ \text{m/s}^2$ to $300\ \text{m/s}^2$. The natural frequency of the instrument is 300 Hz and the value may vary by 3 Hz owing to temperature fluctuations. Calculate the allowable uncertainty in the relative displacement measurement in order to ensure an uncertainty of not more than 5% in the acceleration measurement.
- 9.6 An accelerometer consisting of an elastic element and a potentiometric displacement sensor has to meet the following specifications:

| | | | |
|------------------|-------------|---------------|-------------------------|
| Input range | = 0 to 5g | Damping ratio | = 0.8 |
| Output range | = 0 to 10 V | Seismic mass | = 0.005 kg |
| Spring stiffness | = 20 N/m | g | = 9.81 m/s ² |

- (a) Calculate the input displacement range of the potentiometer and the damping constant.
- (b) If a 1 k Ω potentiometer is used along with a recorder of 10 k Ω , find the percentage error in the measured signal for an acceleration of $2g$.

9.7 Indicate the correct choice:

- (a) In a seismic pick-up for getting an output proportional to acceleration it is desirable to have
- a natural frequency very small in comparison to frequency of input and a damping ratio around 0.7
 - a natural frequency very large in comparison to frequency of input and a damping ratio around 0.7
 - a natural frequency equal to frequency of input and a damping ratio around 0.7
 - a natural frequency very small in comparison to frequency of input and a damping ratio higher than unity
- (b) The natural frequency of a seismic mass type pick-up is f_n . The pick-up will give the correct information, if the exciting frequency f_e is
- f_n
 - $\frac{1}{\sqrt{2}}f_n$
 - $\gg f_n$
 - $\ll f_n$
- (c) In a seismic acceleration sensor with a displacement transducer as the secondary element, decreasing the mass while maintaining all other parameters of the sensor unchanged will
- reduce both natural frequency and steady-state sensitivity
 - increase both natural frequency and steady-state sensitivity
 - increase natural frequency but reduce steady-state sensitivity
 - increase natural frequency without affecting steady-state sensitivity
- (d) A viscous damper consists of a sliding piston and a cylinder filled with oil of kinematic viscosity 5×10^{-5} m²/s. A damping force of 20 N is applied on the piston and the steady state velocity reached is 10 mm/s. The damping coefficient of the damper is
- 2 Ns/mm
 - 4 Ns/mm
 - 10 Ns/mm
 - 20 Ns/mm

-
- (e) A cantilever-type micro-accelerometer is designed by scaling down each dimension of a macro-accelerometer by a factor of 100. If the natural frequency of the macro-accelerometer is ω , then the natural frequency of the micro-accelerometer will be
- (i) 100ω
 - (ii) 10ω
 - (iii) 0.1ω
 - (iv) 0.01ω
- (f) A piezoelectric type accelerometer has a sensitivity of 100 mV/g. The transducer is subjected to a constant acceleration of 5 g . The steady state output of the transducer will be
- (i) 0 V
 - (ii) 100 mV
 - (iii) 0.5 V
 - (iv) 5 V
- (g) The torque in a rotating shaft is measured using strain gauges. The strain gauges must be positioned on the shaft such that axes of the strain gauges are at
- (i) 0° with respect to the axis of the shaft
 - (ii) 30° with respect to the axis of the shaft
 - (iii) 45° with respect to the axis of the shaft
 - (iv) 90° with respect to the axis of the shaft

Temperature Measurement

The change in temperature of a body produces various primary effects such as

1. Change in its physical or chemical state, e.g. phase transition
2. Change in its physical dimensions
3. Variation in its electrical properties
4. Generation of thermoelectricity
5. Variation in its optical properties
6. Change in the frequency of vibration of piezoelectric crystals
7. Change in the velocity of sound
8. Change in the intensity of the emitted radiation

Any of these effects can be exploited to measure the temperature of a body though the first one is generally used for standardisation of temperature sensors rather than for direct measurement of temperature.

But before we discuss temperature transducers, we must note that temperature measurement differs from all other measurements because here a *scale* rather than a unit has to be defined.

10.1 Temperature Scale

Consider, for example, length measurement where a certain length is defined as 1 metre. If two sticks of 1 metre length are placed end to end, the resulting length is 2 metres. This concept is not valid in the case of temperature measurement where two bodies, each of temperature T , when placed in thermal contact, will result in producing temperature T' !

This situation baffled scientists until Lord Kelvin¹ showed in 1848 that the basis of defining an absolute scale for temperature lay in the second law of thermodynamics and the concept of an ideal reversible Carnot² cycle where a perfectly reversible heat engine takes an amount of heat Q_2 from a reservoir of infinite capacity at temperature T_2 and supplies an amount of heat Q_1 to another such reservoir at temperature T_1 according to the relation

$$\frac{Q_2}{Q_1} = \frac{T_2}{T_1}$$

¹William Thomson, 1st Baron Kelvin (1824–1907) was a British mathematical physicist and engineer.

²Nicolas Léonard Sadi Carnot (1796–1832), a French scientist, physicist and military engineer. He was mostly known as the founder of the science of Thermodynamics.

Kelvin's thermodynamic scale is absolute in the sense that it is independent of any material properties. But it is not realisable as such because of its dependence on an ideal cycle.

Fortunately, it can be shown that a temperature scale defined by a constant-volume or a constant-pressure gas thermometer using an ideal gas is identical to the thermodynamic scale. The concerned relations are

$$\text{Boyle's law}^3 \quad \frac{T_2}{T_1} = \frac{p_2}{p_1} \quad \text{when } V \text{ is constant} \quad (10.1)$$

$$\text{Charles' law}^4 \quad \frac{T_2}{T_1} = \frac{V_2}{V_1} \quad \text{when } p \text{ is constant}$$

If the same fixed point, such as the triple point of water, is selected for the reference point, the two scales are numerically equal.

But, here there is a problem. Suppose, we take a constant volume thermometer with a certain volume of gas and measure pressures at the ice- and steam-points. Let these pressures be p_i and p_s respectively. Then the pressure ratio is

$$R = \frac{p_s}{p_i}$$

Next we take a smaller volume of the same gas in the same thermometer and repeat the experiment. Let the corresponding quantities be p'_i , p'_s and

$$R' = \frac{p'_s}{p'_i}$$

According to the ideal gas equation R should equal R' which it does not! This happens because an ideal gas is a mathematical concept which assumes that

1. Gas molecules are point masses, occupying no space
2. Collisions between them are elastic
3. No intermolecular force acts between them

Unfortunately, no real gas satisfies this description and herein lies the problem.

But there is a way out. If all R 's corresponding to different volumes of a particular gas A are plotted against the corresponding p_i 's, the points lie on a straight line and if this line is extrapolated to $p_i = 0$, the corresponding R equals 1.36609 (Fig. 10.1). If now the experiment is repeated with another gas B, the corresponding p_i vs. R plot will be a different straight line, but the value of the intercept will be 1.36609. Therefore, the answer to the problem is that gas thermometers will obey thermodynamic relations provided readings are extrapolated to correspond to zero pressure.

Therefore, for a constant volume thermometer, if the extrapolated pressures at the ice point and any temperature T be $p_{273.16}$ and p_T respectively, then the required temperature is calculated from the relation

$$T = 273.16 \frac{p_T}{p_{273.16}}$$

³Robert Boyle (1627–1691) was an English natural philosopher, chemist, physicist, and inventor.

⁴Jacques Alexandre César Charles (1746–1823) was a French inventor, scientist, mathematician, and balloonist. Charles and the Robert brothers launched the world's first (unmanned) hydrogen-filled balloon in August 1783.

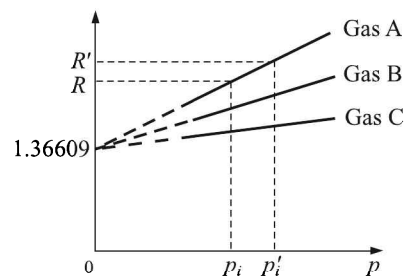


Fig. 10.1 p_i vs. R plot for real gases.

But for all practical purposes, temperature measurement by gas thermometers is cumbersome and therefore, a few fixed points, known as the *International Temperature Scale of 1990* (ITS-90) have been defined at which all thermal transducers are calibrated.

International Temperature Scale of 1990

ITS-90 is so designed that it represents the thermodynamic (absolute) temperature scale (referencing absolute zero) as closely as possible throughout its range.

Salient features

The salient features of ITS-90 are as follows:

1. It defines eighteen points, all of which are based on various thermodynamic equilibrium states of fourteen pure chemical elements and one compound (water).
2. Most of the defined points are based on a phase transition—the melting/freezing point of a pure chemical element, to be specific. Example, the freezing point of aluminium is defined as 660.323°C .
3. The low cryogenic points are based exclusively on the vapour pressure/temperature relationship of helium and its isotopes.
4. The rest of the cold points (those lower than the room temperature) are based on triple points such as the triple point of hydrogen ($-259.3467^{\circ}\text{C}$).
5. It also distinguishes between *freezing* and *melting* points. The distinction depends on whether heat is going into (melting) or out of (freezing) the sample when the measurement is made. Only gallium is measured while melting, all the other metals are measured while the samples are freezing.
6. The triple point of Vienna Standard Mean Ocean Water (VSMOW⁵) is assumed to be known with absolute precision—whatever calibration standard is employed—because the very definitions of both the Kelvin and Celsius scales are fixed by international agreement based, in part, on this point.

⁵Vienna Standard Mean Ocean Water (VSMOW) is a water standard defining the isotopic composition of water. It was promulgated by the International Atomic Energy Agency in 1968. For more details see <http://www.iaea.org/>.

Fixed points

The fixed points defined by ITS-90 are listed in Table 10.1.

Table 10.1 Fixed points of ITS-90

| <i>Substance</i> | <i>Its state</i> | <i>Defined temperature (°C)</i> | <i>Defined temperature (K)</i> |
|------------------|---|-------------------------------------|------------------------------------|
| Helium-3 | Vapour pressure vs. temperature relation (by equation) | −272.50 to −269.95 | 0.65 to 3.2 |
| Helium-4 | Vapour pressure vs. temperature relation below its λ -point (by equation) | −271.90 to −270.9732 | 1.25 to 2.1768 |
| Helium-4 | Vapour pressure vs. temperature relation above its λ -point (by equation) | −270.9732 to −268.15 | 2.1768 to 5.0 |
| Helium | Vapour pressure vs. temperature relation | −270.15 to −268.15 | 3 to 5 |
| Hydrogen | Triple point | −259.3467 | 13.8033 |
| Neon | Ditto | −248.5939 | 24.5561 |
| Oxygen | Ditto | −218.7916 | 54.3584 |
| Argon | Ditto | −189.3442 | 83.8058 |
| Mercury | Ditto | −38.8344 | 234.3156 |
| Water | Ditto at STP ^a | 0.01 | 273.16 |
| Gallium | Melting point | 29.7646 | 302.9146 |
| Indium | Freezing point ^b | 156.5985 | 429.7485 |
| Tin | Freezing point | 231.928 | 505.078 |
| Zinc | Freezing point | 419.527 | 692.677 |
| Aluminium | Freezing point | 660.323 | 933.473 |
| Silver | Freezing point | 961.78 | 1234.93 |
| Gold | Freezing point | 1064.18 | 1337.33 |
| Copper | Freezing point | 1084.62 | 1357.77 |

^a Standard Temperature and Pressure

^b Freezing point is distinguished from melting point. The distinction depends on whether heat is going into (melting) or out of (freezing) the sample when the measurement is made.

Recommended devices

To cover the entire range, different ways of measuring temperatures have been recommended. These are listed in Table 10.2.

With this background on the temperature scale, we discuss thermal transducers based on effects stated before.

Table 10.2 Ranges and methods of measuring temperatures

| Range | Method |
|---|--|
| Between 0.65 K and 5.0 K | Vapour-pressure temperature relationship of ^3He and ^4He . |
| Between 3.0 K and 24.5561 K (triple point of neon) | A helium gas thermometer calibrated at three fixed points in this range. |
| Between 13.8033 K (triple point of equilibrium hydrogen) and 1234.93 K (freezing point of silver) | A standard platinum resistance thermometer (RTD) calibrated at the defining fixed points and using specified interpolation procedures. |
| Above 1234.93 K (freezing point of silver) | In terms of a defining fixed point and the Planck radiation law as used by radiation pyrometers ^a . |

^a See Section 10.8 at page 421.

10.2 Change in Dimensions

Bimetals

Bimetals are formed by firmly bonding two strips of metals A and B having different thermal expansion coefficients α_A and α_B . If this bimetal is a straight line at temperature T_1 , then at an elevated temperature T_2 the strip will form a uniform circular arc of radius of curvature r (Fig. 10.2) such that

$$T_2 \approx T_1 + \frac{2t}{3r(\alpha_A - \alpha_B)}$$

where t is the thickness of the strip.

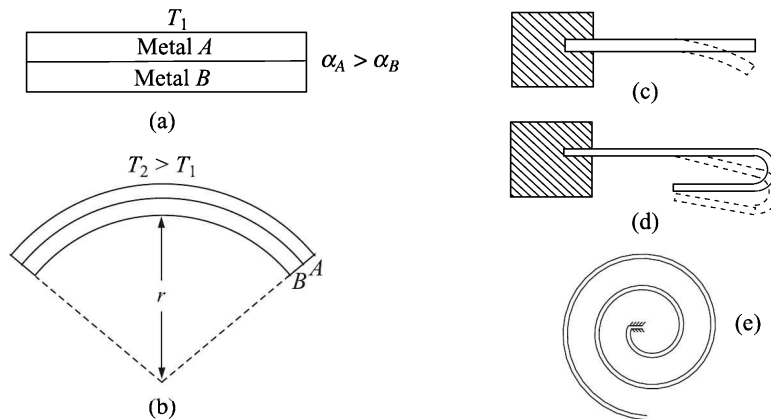


Fig. 10.2 Bimetallic sensors: (a) two straight metal strips are bonded at temperature T_1 , (b) the same bimetallic strip forms a circular arc when raised to a higher temperature T_2 , (c) cantilever type bimetal, (d) U-shape bimetal, and (e) spiral type bimetal.

Different forms—such as cantilever, U-shape, spiral—of bimetals are used. The element B is generally invar⁶ because its thermal expansion coefficient is almost zero. Temperatures ranging from -75°C to 550°C can be measured with the help of bimetals with an accuracy of 0.5 to 1%.

Liquid-in-glass Thermometers

A liquid-in-glass thermometer is a first order instrument⁷. The factors governing its sensitivity, etc. of a first order instrument have been discussed in Section 4.5 at page 87. The point which needs to be stressed here is that these thermometers are basically of two types, namely

1. Full immersion type
2. Partial immersion type

If a full immersion type thermometer is partially dipped in the measured medium, the reading is bound to be incorrect.

Full immersion type thermometers are more accurate because ambient temperature does not influence their readings as it does with the partial immersion types.

The accuracy of a liquid-in-glass thermometer is generally around 0.2°C .

Filled System Thermometers

The filled system, or just filled thermometers are those that work on pressure or volume change of a gas or changes in vapour pressure of a liquid.

A filled system thermometer consists of four parts:

1. Bulb
2. Capillary tube
3. Pressure- or volume-sensitive element
4. Indicating device

The capillary tube connects the bulb containing a fluid that is sensitive to temperature changes to the element that is sensitive to pressure or volume changes. The pressure-sensitive or volume-sensitive element may be a Bourdon tube (Fig. 10.3), a helix, a diaphragm, or bellows. The motion of the temperature- or volume-sensitive element couples mechanically to the indicating, recording, or controlling device.

The gas-filled systems are sometimes called *gas thermometers*, or *pressure thermometers*.

Operating principle

The volume expansion of a liquid is given by the equation

$$V_T \cong V_0(1 + \gamma T) \quad (10.2)$$

where V_T and V_0 are the volumes of the liquid at temperatures T and 0°C respectively, and γ is the coefficient of volume expansion. But the bulbs, which contain the liquids also expand or

⁶An alloy of nickel and iron.

⁷See Section 4.5 at page 87.

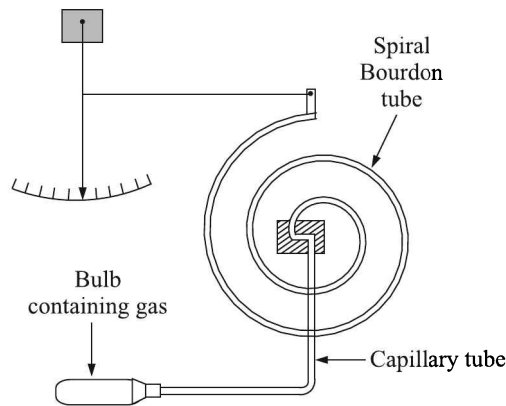


Fig. 10.3 Filled system thermometer using a spiral Bourdon tube.

contract with the variation of temperature. Though these expansions/contractions are much smaller than those of the liquids, they introduce some degree of nonlinearity in the relation given by Eq. (10.2).

For gas-filled systems, the guiding equation is the well-known Boyle's law given by Eq. (10.1). Here, the absolute temperature rather than the one in the Celsius scale maintains a linear relationship with the pressure. But here the nonlinearity introduced by the volume change of the container with temperature is negligibly small.

Classification of filled system thermometers

Mainly based on the classification made by the Scientific Apparatus Makers' Association (SAMA) of the USA, filled system thermometers are usually divided into four classes as shown in Table 10.3.

Table 10.3 Classification of filled system thermometers

| Principle | Class | Filling material |
|----------------------|----------------|---|
| Volumetric expansion | I | Liquid other than mercury |
| | V ^a | Mercury- or mercury-thallium eutectic amalgam |
| Pressure generation | II | Vapour |
| | III | Gas |

^a It is interesting to note that there is no Class IV. Actually, the erstwhile Class IV of the manufacturers of instruments of the USA, which used only mercury filling, has been re-classified by the SAMA as Class V by making its scope a little wider with the inclusion of the amalgam.

Ranges

The liquids generally used in the Class I thermometers are alcohol, pentane and toluene. The minimum temperature that can be measured by such systems depends on the freezing point of the organic liquid used. This lies usually between -200°C and -75°C depending on the liquid

used. The maximum measurable temperatures, however, depend not only on the boiling points of the liquids but also on the linearity of their volume expansions at higher temperatures. For organic liquids, this maximum temperature is generally 300°C . For mercury-filled Class V thermometers, the range is from -38°C to 650°C . For low spans, organic liquid-filled systems are preferred because they offer better sensitivity owing to their higher thermal expansions.

Gas-filled systems cover the widest temperature ranges. On the low side, their ranges are limited by their critical temperatures. On the high side, the ranges are limited by the melting point of the bulb material as well as the by the point at which the Bourdon tube is overstressed. Generally, the range of gas-filled thermometers is from 5 K to 925 K.

Sources of error

Especially filled system thermometers are vulnerable to two sources of error arising out of the differential expansions of the (i) fluid contained in the temperature-sensing bulb and (ii) the fluid contained in the Bourdon tube and the capillary which will remain at the room temperature.

These errors are minimised through what are called *case compensation* and *full compensation*. In the case compensation, only the error arising out of the differential expansion of the fluid in the Bourdon tube is compensated for (Fig. 10.4).

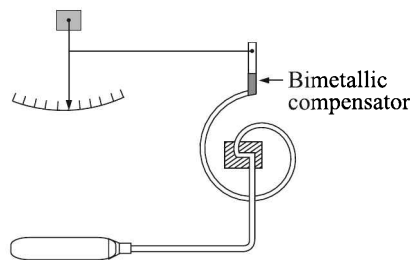


Fig. 10.4 Case compensation for the filled system thermometer.

In full compensation (Fig. 10.5), the compensation covers both the Bourdon tube and the capillary. The full compensation is necessary if the capillary length is greater than 3 m.

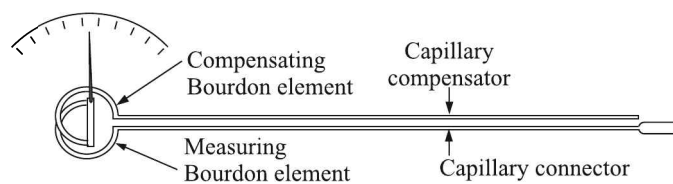


Fig. 10.5 Full compensation for the filled system thermometer.

Despite compensations, to obtain sufficiently accurate data from such arrangements, the volume of the temperature-sensing bulb has to be sufficiently larger than the volume of the Bourdon tube.

Advantages and disadvantages

The advantages and disadvantages of the filled system thermometers are listed in Table 10.4

Table 10.4 Advantages and disadvantages of filled system thermometers

| <i>Advantages</i> | <i>Disadvantages</i> |
|---|---|
| 1. Self-contained. Do not require power unless they are combined with an electronic transmission system. | 1. In case of system failure, usually the entire unit needs to be replaced. |
| 2. Simple, rugged and have minimum possibility of being damaged during shipment or installation. Last almost indefinitely, as long as the metal tubing is not broken. | 2. Many applications may not allow insertion of a large bulb volume within the measured medium. |
| 3. Simplicity of the design makes them inexpensive. | 3. Performance characteristics may vary considerably from one filling fluid to another. The user needs to be careful about choosing a system for a particular application. |
| 4. Sensitivity, speed of response and accuracy satisfactory for most of the process industries. | 4. Maximum temperature is more limited than most of the electrical measuring systems. |
| 5. Capillary allows considerable separation between the point of measurement and the point of indication. | 5. Separation between the sensing and indicating elements is rather limited in comparison to electrical systems. |
| 6. Can be designed to deliver significant power to drive indicating or controlling mechanisms, including valves. | 6. Low cost of electronic devices to read the output of thermocouples and RTDs and to indicate or control, together with the ability to locate the sensor independently of the receiving device, has made electronic means more attractive. |

10.3 Change in Electrical Properties

Of all the electrical properties of substances, only the variation of resistance with temperature has been found to be useful in measuring temperatures. From this variation of resistance in metals and semiconductors, the platinum resistance thermometer and the thermistor have been fabricated.

Platinum Resistance Thermometer

First constructed by William Siemens⁸ (1871) and later perfected by Callendar and Griffiths (1887), the platinum resistance thermometer is widely used to measure temperature. At present, resistance thermometers which utilise the variation of their resistance to detect temperatures are often referred to as resistance temperature detectors (RTD).

The relation between electrical resistance of a metal R_T and the corresponding temperature T is generally given as

$$R_T = R_0(1 + C_1T + C_2T^2 + \dots + C_nT^n)$$

where C 's are constants and R_0 is the resistance at temperature $T = 0^\circ\text{C}$.

⁸Sir William Siemens (1823–1883) was a German-born British engineer.

Although this is a nonlinear relationship, it can be seen from Fig. 10.6 that the curve is nearly linear for copper and platinum over a fairly long range. However, copper being easily susceptible to chemical reactions such as oxidation, sulphate formation etc., platinum is chosen for the RTDs. The platinum resistance thermometers are also referred to as PRTs.

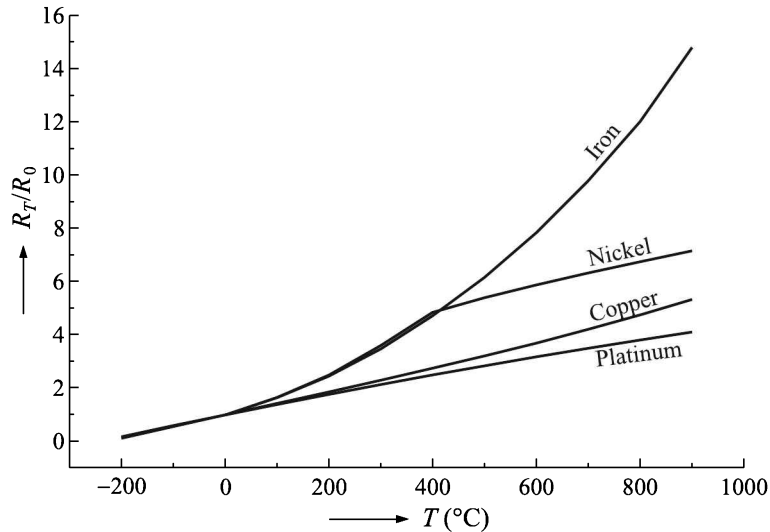


Fig. 10.6 Resistance-temperature characteristics of metals.

Two types

PRTs are of two types—thin film and wire-wound.

Thin film PRT. The film PRT is produced by depositing a thin film of platinum onto a substrate usually of ceramic material through cathodic atomisation or *sputtering*⁹. The thickness of the layer of sputtered layer of platinum on the substrate can be controlled by limiting the period of sputtering process. The layer may be as thin as 1 μm . After the deposition is made, a laser is used to trim the platinum layer to a precise resistance.

This new technology has helped produce PRTs at a relatively low cost and having versatile shapes and designs. They can also be made much smaller than their wire-wound counterparts. In fact, it is now possible to manufacture a PRT element of the size of a pencil tip! By making the element as small as possible, the PRT assembly can be made to have faster response and be *tip sensitive*. The tip sensitivity of a thermocouple has always been an advantage over the PRT. Now this advantage has been all but eliminated.

However, the differential expansion between the substrate and platinum film gives rise to *strain gauge effects* and stability problems.

Wire-wound PRT. Wire-wound PRTs can have a greater accuracy, especially for wide temperature ranges.

⁹See, for example, *Solid State Electronic Devices*, 6th Ed by BG Streetman and SK Bannerjee, PHI Learning (New Delhi) 2010, pp 167–68 for details.

Construction. Pure platinum wire, free from silicon, carbon, tin and other impurities, is doubled to avoid induction effects and wound on a thin insulating mica former. The ends of this wire are joined to terminals A and B on the top of the instrument (Fig. 10.7). Another exactly similar lead, with its lower end shorted to B, is connected to terminal C, to compensate for the resistance of the leads.

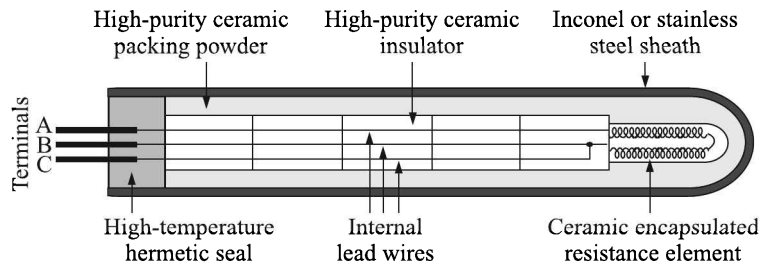


Fig. 10.7 Diagram of a three-wire platinum resistance thermometer.

The thermometer assembly is enclosed in a tube with proper spacers to prevent leads from short-circuiting and the tube is sealed at the top to prevent moisture, etc. depositing on the mica former.

Pt-100. The IEC 751:1983 is the current international standard which specifies tolerance and the temperature to electrical resistance relationship for platinum resistance thermometers. Usually, the devices used in industry have a nominal resistance of $100\ \Omega$ at 0°C , and are called Pt-100 RTDs. The Pt-100 RTD has a sensitivity of $0.385\ \Omega/^\circ\text{C}$ according to the European standard¹⁰.

Measurements with RTDs

For measuring the resistance, precautions against two interferences need to be taken. They are

1. Self-heating
2. Lead wire resistance

Self-heating. A small current requires to be passed through the RTDs in order to avoid the resistive heating. For example, 1 mA through a $100\ \Omega$ RTD generates $100\ \mu\text{W}$. This may seem insignificant, but it can raise the temperature of some RTDs a significant fraction of a degree.

A typical value for self-heating error is $1^\circ\text{C}/\text{mW}$ in free air. The same RTD rises $0.1^\circ\text{C}/\text{mW}$ in air flowing at 1 m/s. Using

1. the minimum excitation current that provides the desired resolution, and
2. the largest physically practical RTD

will help reduce self-heating errors.

¹⁰The American standard, which uses a purer grade of platinum, has a sensitivity of $0.392\ \Omega/^\circ\text{C}$

Lead wire resistance. Because of the rather low resistance of the RTD, the lead wire resistance may also interfere in the measurement. For example, lead wires with a resistance of $1\ \Omega$ connected to a $100\ \Omega$ platinum RTD cause a 1% measurement error.

Of the two possible sources of error, as mentioned, the self-heating can be avoided by measuring the resistance of the PRT by balancing a bridge circuit when a small current may be allowed to flow through the resistances. The general methods of measuring the RTD resistance by using bridges of different configurations are described below.

Two-wire connection. The simplest resistance measurement configuration uses two wires to connect the thermometer to a Wheatstone bridge [Fig. 10.8(a)].

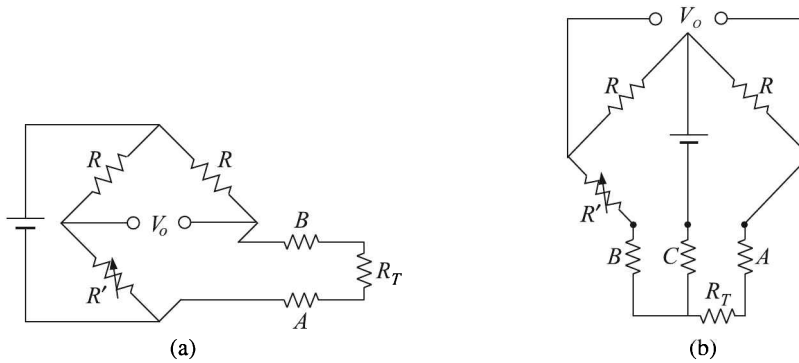


Fig. 10.8 Bridge configurations for the PRT: (a) two-wire, and (b) three-wire connections.

In the balanced condition of the bridge, we have

$$R + R' = R + A + R_T + B$$

or

$$R' = R_T + (A + B)$$

where R_T is the resistance of the RTD

A , B are resistances of the lead wires

R' is the adjustable resistance

Therefore, in this configuration, the resistance of the connecting wires is always included with that of the sensor leading to errors in the measurement. It is mainly used when high accuracy is not required. Using this configuration, about 100 m of cable can be used. This applies equally to null measurement in a balanced bridge or deflection measurement in a fixed bridge system.

Three-wire connection. In order to minimise the effects of the lead resistances a three-wire configuration can be used. The Callendar and Griffiths' bridge is normally used to measure the resistance of the thermometer in a three-wire configuration [Fig. 10.8(b)].

Here, the two leads to the sensor are on the adjoining arms. There is a lead resistance in each arm of the bridge. They cancel out as can be seen from the following analysis.

If, R_T is the resistance of the PRT at temperature T

R' is the resistance used to balance the bridge

A , B , C are resistances of the lead wires

then for the balanced bridge,

$$R + R' + B + C = R + R_T + A + C$$

Therefore,

$$R' = R_T$$

provided $A = B$.

High quality connection cables should be used for this type of configuration because we have assumed that the two lead resistances are equal. This configuration allows for up to 600 m of cable.

Four-wire connection. The four-wire resistance thermometer configuration even further increases the accuracy and reliability of the measurement of resistance. In the four-wire connection, shown in Fig. 10.9, it can be shown that the effects of lead wire resistance are eliminated.

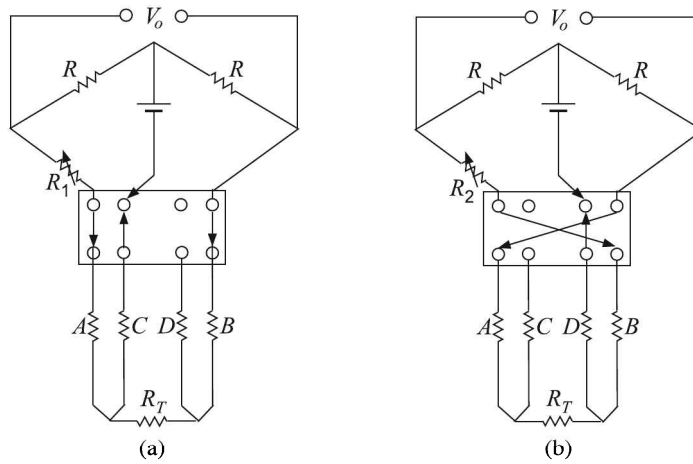


Fig. 10.9 Two connections of the four-wire RTD to eliminate the lead wire effects.

In the balanced bridge with connections as shown in Fig. 10.9(a), we have

$$R + R_1 + A + C = R + B + R_T + C$$

or

$$R_1 + A = R_T + B \quad (10.3)$$

where R_1 is the value of the resistance required to balance the bridge.

Next, the connections are changed to that shown in Fig. 10.9(b). Now, when the bridge is balanced with a changed value R_2 of the adjustable resistance, we have

$$R + R_2 + B + D = R + A + R_T + D$$

or

$$R_2 + B = R_T + A \quad (10.4)$$

Adding Eqs. (10.3) and (10.4), we get on simplification

$$R_T = \frac{R_1 + R_2}{2}$$

Thus, the lead wire resistance will have no interference in the measurement.

Cable resistance of up to $15\ \Omega$ can be handled, though in principle, the resistance error due to lead wire resistance is zero in four-wire measurements. It also provides full cancellation of spurious effects.

From the above explanation, we can conclude that two-wire RTD is the worst type, while four-wire RTD is the best. Three-wire RTD has a medium or good performance compared with two-wire or four-wire RTD. In industries, three-wire RTD is mostly used. Four-wire RTD is used only for a very special application that needs a very accurate temperature measurement.

However, in industries, since bridge measurement is time-consuming, arrangements like that shown in Fig. 10.10 are used along with automatic data processing. A constant current source supplies the required current.

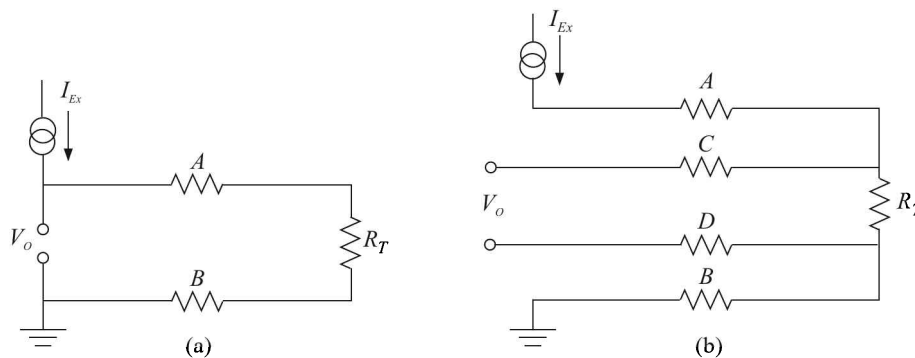


Fig. 10.10 Typical circuits for measurement of temperature in industries: (a) two-wire connection and (b) four-wire connection. A , B , C , D indicate resistance of lead wires, R_T that of the RTD, V_o is the output voltage and I_{Ex} is the constant current from an external source.

The four-wire method has the advantage of not being affected by the lead resistances because the voltage measurement is done by a high impedance device that makes the current through the path minimal. Therefore, this method ensures much more accurate measurement of the voltage across the RTD.

Temperature computation

From Callendar-Van Deusen relation. With the advent of computers, the temperature corresponding to a measured resistance can be found by the method of iteration from the Callendar-Van Deusen relations

$$R_T = R_0[1 + C_1T + C_2T^2 + C_3T^3(T - 100)] \quad [-200^\circ\text{C} < T < 0^\circ\text{C}]$$

$$R_T = R_0(1 + C_1T + C_2T^2) \quad [0^\circ\text{C} < T < 850^\circ\text{C}]$$

where

$$C_1 = 3.9083 \times 10^{-3} \text{ }^\circ\text{C}^{-1}$$

$$C_2 = -5.775 \times 10^{-7} \text{ }^\circ\text{C}^{-2}$$

$$C_3 = -4.183 \times 10^{-12} \text{ }^\circ\text{C}^{-4}$$

However, Callendar and Griffiths suggested a simpler way to figure it out as follows.

Callendar and Griffiths' method. Callendar and Griffiths observed that the following simple relation gives true readings up to 630°C.

$$R_T = R_0(1 + cT_{Pt})$$

where c is the mean temperature coefficient of resistance between 0°C and 100°C, and T_{Pt} is given by

$$T_{Pt} = \frac{R_T - R_0}{R_{100} - R_0} \times 100 \quad (10.5)$$

They further showed that the difference between the true temperature and that obtained from the Eq. (10.5) is given by

$$T - T_{Pt} = \delta(0.0001T^2 - 0.01T) \quad (10.6)$$

where, δ is a constant for a particular specimen of wire and its value varies between 1.488 and 1.498. So, the procedure is as follows:

- Step 1. Find platinum temperature from Eq. (10.5), by measuring R_T , R_{100} and R_0 . The last two quantities are to be determined once and for all.
- Step 2. Substitute this value for T_{Pt} on the left- and for T on the right-hand side of Eq. (10.6) to obtain a revised value of T .
- Step 3. Substitute the value of T obtained from Step 2 in the right-hand side of Eq. (10.6) to obtain a more accurate value for T .
- Step 4. Repeat Step 3 until the value of T converges. This iterative procedure is also called the *successive approximation method*.

However, the manufacturers usually supply a look-up table of R_T versus T with their thermometers. Values from this table, if necessary, may be linearly interpolated to obtain fairly accurate temperatures.

Range

The range of a platinum resistance thermometer is normally -40°C to 1200°C . Its accuracy varies from 0.2% to 1.2% at different ranges. However, standard platinum resistance thermometers (SPRTs) are the instruments specified in the International Temperature Scale of 1990 (ITS-90) for performing measurements within the range of 13.8033 K (the triple point of hydrogen) to 1234.93 K (the freezing point of silver). The ITS adds

However, it must be said immediately that no single thermometer or design of thermometer can be expected to cover this entire 1221 K span, but it can be covered by three designs of thermometer, with substantial overlap of range.

Before moving on to the next method, let us consider a few examples related to the measurement of temperature by RTDs.

Example 10.1

A platinum resistance thermometer is to be used to measure temperature between 0° and 200°C. Given that the resistance R_T at $T^\circ\text{C}$ is

$$R_T = R_0(1 + \alpha T + \beta T^2)$$

and $R_0 = 100.0 \Omega$, $R_{100} = 138.50 \Omega$ and $R_{200} = 175.83 \Omega$, calculate the nonlinearity at 100°C as a per cent of full-scale deflection.

Solution

As discussed in Section 2.1 at page 4, the nonlinearity may be worked out as follows. Here,

$$\begin{aligned}(R_T)_{\text{linear}} &= R_0 + \left(\frac{R_{200} - R_0}{200} \right) T \\ &= 100.0 + \left(\frac{175.83 - 100.0}{200} T \right) \\ &= 100.0 + 0.37915 T\end{aligned}$$

Therefore $(R_{100})_{\text{linear}} = 100 + 37.915 = 137.915 \Omega$

But $(R_{100})_{\text{actual}} = 138.50 \Omega$

Nonlinearity,

$$N = 138.50 - 137.915 = 0.585 \Omega$$

As % of FSD $= \frac{0.585}{175.83} \times 100 = 0.33$

Alternatively

The given relation can be rearranged as

$$\alpha + \beta T = \frac{R_T - R_0}{R_0 T}$$

For 100°C , $\alpha + 100\beta = \frac{138.50 - 100.0}{100 \times 100} = 38.5 \times 10^{-4}$ (i)

For 200°C , $\alpha + 200\beta = \frac{175.83 - 100.0}{100 \times 200} = 37.965 \times 10^{-4}$ (ii)

From Eqs. (i) and (ii), we get

$$\alpha = 39.035 \times 10^{-4} / ^\circ\text{C}$$

$$\beta = -5.35 \times 10^{-7} / ^\circ\text{C}^2$$

The nonlinear term in the resistance-temperature relation is $R_0\beta T^2$. Hence, the absolute value of nonlinearity at 100°C is

$$|R_0\beta t^2| = 100 \times 5.35 \times 10^{-7} \times 10^4 = 0.535 \Omega$$

Since the thermometer is to measure a maximum temperature of 200°C , the full-scale deflection should occur at that temperature. The corresponding value of the resistance is 175.83Ω . Therefore, the required nonlinearity at 100°C is

$$\frac{0.535}{175.83} \times 100 \cong 0.33\% \text{ FSD}$$

Example 10.2

The following table gives the variation of resistance with temperature for an RTD:

| Temperature(°C) | 15 | 18 | 21 | 24 | 26.5 | 29.5 | 33 |
|------------------------|-------|--------|--------|-------|--------|--------|--------|
| Resistance(Ω) | 106.6 | 107.14 | 108.22 | 109.3 | 110.38 | 111.46 | 112.75 |

Find the linear and quadratic approximation of the above resistance-temperature curve for temperature variation between 15°C and 33°C.

Solution

Let the linear relation be

$$R_T = R_0(1 + \alpha t)$$

Then, the normal equations (see Section 3.6 at page 54) are

$$\Sigma R_T = nR_0 + (R_0\alpha)\Sigma T$$

$$\Sigma R_T T = R_0\Sigma T + (R_0\alpha)\Sigma T^2$$

Therefore,

$$R_0 = \frac{\Sigma T^2 \Sigma R_T - \Sigma T \Sigma R_T T}{n \Sigma T^2 - (\Sigma T)^2}$$

$$R_0\alpha = \frac{n \Sigma R_T T - \Sigma T \Sigma R_T}{n \Sigma T^2 - (\Sigma T)^2}$$

The relevant calculation is given below in a tabular form, the bold-face numbers in the last row indicating sums of corresponding columns.

| R_T | T | T^2 | $R_T T$ |
|---------------|--------------|---------------|-----------------|
| 106.6 | 15.0 | 225.0 | 1599.00 |
| 107.14 | 18.0 | 324.0 | 1928.52 |
| 108.22 | 21.0 | 441.0 | 2272.62 |
| 109.3 | 24.0 | 576.0 | 2623.32 |
| 110.38 | 26.5 | 702.25 | 2825.07 |
| 111.46 | 29.5 | 870.25 | 3288.07 |
| 112.75 | 33.0 | 1089.0 | 3720.75 |
| 765.85 | 167.0 | 4227.5 | 18357.35 |

$$R_0 = \frac{(4227.5)(765.85) - (167)(18357.35)}{(7)(4227.5) - (167)^2} = 100.94$$

$$R_0\alpha = \frac{(7)(18357.35) - (167)(765.85)}{(7)(4227.5) - (167)^2} = 0.35486$$

$$\alpha = 0.003516$$

Hence,

$$R_T = 100.94(1 + 0.0035167T)$$

Let the quadratic relation be

$$R_T = R_0(1 + \alpha T + \beta T^2)$$

Then the normal equations are

$$\Sigma R_T = nR_0 + (R_0\alpha)\Sigma T + (R_0\beta)\Sigma T^2$$

$$\Sigma R_T T = R_0\Sigma T + (R_0\alpha)\Sigma T^2 + (R_0\beta)\Sigma T^3$$

$$\Sigma R_T T^2 = R_0\Sigma T^2 + (R_0\alpha)\Sigma T^3 + (R_0\beta)\Sigma T^4$$

The relevant calculation is given below in a tabular form, the bold-face numbers in the last row indicating sums of corresponding columns.

| R_T | T | T^2 | T^3 | T^4 | R_T | $R_T T^2$ |
|---------------|--------------|---------------|-----------------|--------------------|-----------------|------------------|
| 106.6 | 15.0 | 225.0 | 3375.0 | 50625.0 | 1599.0 | 23985.0 |
| 107.14 | 18.0 | 324.0 | 5632.0 | 104976.0 | 1928.52 | 34713.36 |
| 108.22 | 21.0 | 441.0 | 9261.0 | 194481.0 | 2272.62 | 47725.02 |
| 109.3 | 24.0 | 576.0 | 13824.0 | 331776.0 | 2623.32 | 62956.8 |
| 110.38 | 26.5 | 702.25 | 18609.625 | 493155.0625 | 2925.07 | 77514.355 |
| 111.46 | 29.5 | 870.25 | 25672.375 | 757335.0625 | 3288.07 | 96998.065 |
| 112.75 | 33.0 | 1089.0 | 35937.0 | 1185921.0 | 3720.75 | 122789.75 |
| 765.85 | 167.0 | 4227.5 | 112511.0 | 3118269.125 | 18357.35 | 466677.35 |

Substituting these values in the normal equations and solving them by applying Cramer's rule, we get

$$R_0 = 102.7017$$

$$R_0\alpha = 0.1983 \quad \alpha = 0.00193$$

$$R_0\beta = 3.2684 \quad \beta = 0.03182$$

Hence

$$R_T = 102.7017(1 + 0.00193T + 0.03182T^2)$$

Example 10.3

Design an electronic circuit using RTD which may provide 0–200 mV output corresponding to 0–2000°C. Assume that $R_0 = 100 \Omega$ and $R_{200} = 180 \Omega$.

Solution

We assume a linear relation between the temperature and resistance as

$$R_T = R_0(1 + \alpha T)$$

So, from the given data we get

$$\alpha = \frac{R_T - R_0}{T} = \frac{180 - 100}{100} = 0.004$$

and

$$R_{2000} = 100(1 + 0.004 \times 2000) = 900 \Omega$$

We consider a Wheatstone bridge where all the arms contain 100Ω resistances at 0°C . At 2000°C , the resistance of one of the arms, containing the RTD, turns out to be 900Ω . The problem to solve is that what input voltage to the bridge should yield a variation of 0 to 200 mV for this variation of resistance. From Fig. 10.11, we get

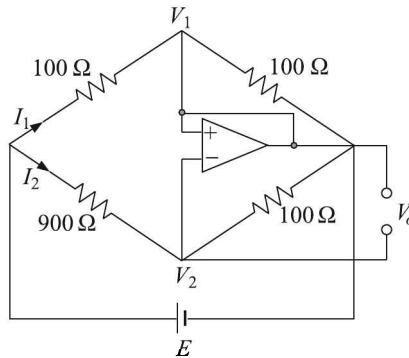


Fig. 10.11 Wheatstone bridge (Example 10.3).

$$I_1 = \frac{E}{200}$$

$$I_2 = \frac{E}{1000}$$

$$V_1 = I_1 \times 100$$

$$V_2 = I_2 \times 900$$

We need to figure out the supply voltage E .

We see,

$$V_1 - V_2 = 0.2 \text{ (given)}$$

$$= E \left(\frac{900}{1000} - \frac{100}{200} \right)$$

\Rightarrow

$$E = \frac{0.2}{0.9 - 0.5} = 0.5 \text{ V}$$

Thermistors

Thermistors¹¹ are thermally sensitive resistors. As a matter of fact the resistance of all resistors vary with temperature, but thermistors are constructed of materials with a resistivity that is especially sensitive to temperature.

If ΔR is the change in resistance corresponding to a temperature change of ΔT and to a first approximation they are related as

$$\Delta R = \alpha_T \Delta T \quad (10.7)$$

¹¹A contraction of [therm]al res[istor].

where α_T is called the temperature coefficient of resistivity. Thermistors can be divided into two categories depending on whether α_T is positive or negative. Those with a positive temperature coefficient are called PTC and those with a negative temperature coefficient are called NTC thermistors.

Thermistors are small, constructed from ceramics or polymers, and have a small temperature range with a high response.

Positive temperature coefficient (PTC) thermistors

Commercial PTC thermistors fall into the following two major categories (see Fig. 10.12):

1. Thermally sensitive silicon resistors, sometimes referred to as *silistors*. These devices exhibit a fairly uniform positive temperature coefficient (about $+0.77\%/^{\circ}\text{C}$) through most of their operational range, but can also exhibit a negative temperature coefficient region at temperatures in excess of 150°C . These devices are most often used for temperature compensation of silicon semiconducting devices in the range of -60°C to $+150^{\circ}\text{C}$.

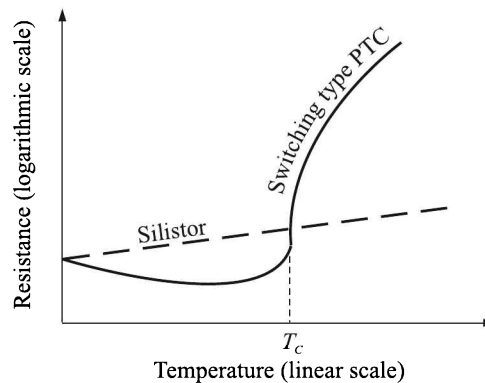


Fig. 10.12 Resistance vs. temperature curves for PTC thermistors: - - -, silistors and continuous curve, switching type.

2. Thermistors made of a doped polycrystalline ceramic containing barium, lead and strontium titanates with additives such as yttrium, manganese, tantalum and silica. These devices have a resistance-temperature characteristic that exhibits a very small negative temperature coefficient until the device reaches a critical temperature T_C , that is referred to as its *Curie, switch or transition* temperature.

The dielectric constant of this latter ferroelectric material varies with temperature. Below the Curie point, it shows a low resistance and a small negative temperature coefficient. At the Curie point the resistance increases sharply.

The reason for this peculiar behaviour is that below the Curie point, its high dielectric constant prevents the formation of potential barriers between the crystal grains, leading to a low resistance. The dielectric constant drops sufficiently at the Curie point to allow the formation of potential barriers at the grain boundaries with a consequent increase in its resistance.

Another group of PTC thermistors are made of polymers. They are sold under brand names like *Polyswitch*, *Semifuse* and *Multifuse*. They are made of slices of plastics with carbon grains embedded in them. At lower temperatures, carbon grains are in contact with each other allowing a conductive path through the device. At higher temperatures, the thermal expansion of the plastic slice forces the carbon grains apart. This causes the resistance to rise sharply.

PTC thermistors are mostly used for *switching* in temperature controllers rather than proportional temperature measurement.

Negative temperature coefficient (NTC) thermistors

Owing to their larger resistance change with temperature, NTC devices are usually more suitable for precision temperature measurement even though the relevant characteristic curve is nonlinear.

Construction. Manganese, nickel or cobalt oxides are milled, mixed in proper proportion with binders, pressed into desired shapes and then sintered to form thermistors in the form of rods, discs, flakes or beads (Fig. 10.13).

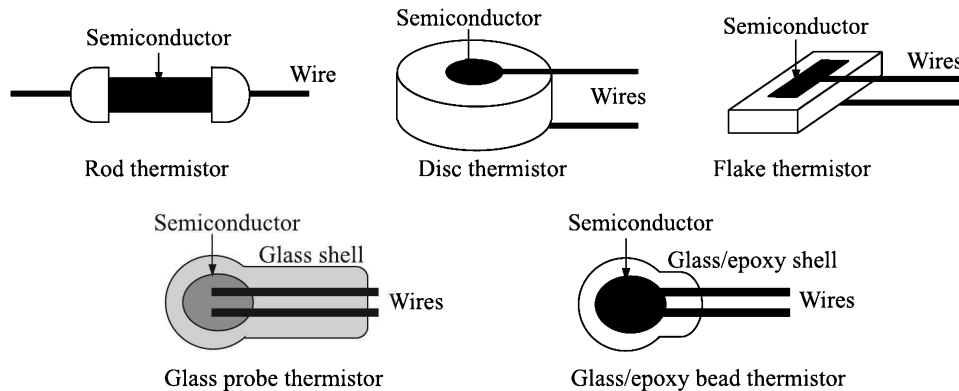


Fig. 10.13 Different forms of thermistors.

Lastly, the wire leads are attached, and the combination is coated with glass or epoxy. This coating provides mechanical strength and electrical resistance. Because the electrical resistance of the epoxy/glass shell is high, the only noise pickup occurs through capacitive coupling. The thermistor element and the connecting wires along with their insulation are put inside a metal sheath to form a thermistor probe as shown in Fig. 10.14.

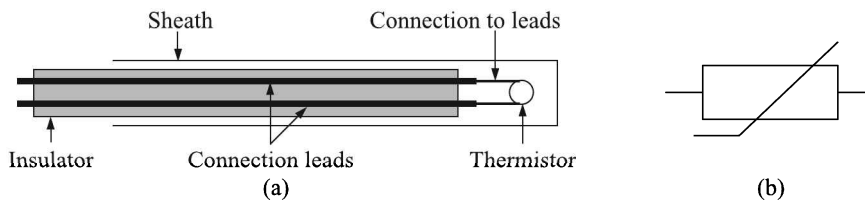


Fig. 10.14 Thermistor: (a) cutaway view, and (b) symbol.

By varying the mixture of oxides, it is possible to produce thermistors having a range of resistance values from $30\ \Omega$ to $20\ \text{M}\Omega$ (at 25°C). Spherical thermistor beads come in diameters ranging from 0.3 to 2 mm. The speed of response of a temperature transducer is related to its surface area. So, a system with a fast response time can be designed by using a small thermistor.

Thermistor resistances are usually specified at room temperature (i.e. 25°C). The resistances of commercially available thermistors are usually 2252, 5000 and 10000 ohm. The standard temperature ranges are -55°C to 150°C although some devices have been shown to remain stable for long terms from -100°C to 400°C .

No special leads are required to connect thermistors to instruments because the device operates at very high resistance compared to the leads.

Power dissipation. When electrical power is delivered to a thermistor, its temperature will rise. So, when using the thermistor to measure temperature, the temperature rise caused by self-heating is a source of measurement error. The heat dissipation constant D is defined as

$$D = \frac{Q}{\Delta T} \quad (10.8)$$

where Q is the power delivered and ΔT is the corresponding rise in temperature.

Equation (10.8) can be utilised to determine the maximum allowable power that can be applied to the thermistor. For example, if the desired temperature resolution is δT , then the interface has to be so designed that

$$Q \leq (\delta T)D$$

It is important to take into account the thermal environment around the thermistor when considering errors caused by self-heating. The dissipation constant for the typical thermistor is $2.5\ \text{mW}/^\circ\text{C}$ for still air and $5\ \text{mW}/^\circ\text{C}$ for still water.

Thermistors are ideally used in Wheatstone bridge circuits which allow a very low current through it.

Resistance-temperature relation. The resistance R_T of a thermistor at temperature T can be represented by the Steinhart-Hart equation¹² as

$$\frac{1}{T} = A + B \ln R_T + C (\ln R_T)^3 \quad (10.9)$$

where A , B and C are called the Steinhart-Hart parameters. They require to be specified for each device. The inverse relation of resistance as a function of temperature can be obtained from Eq. (10.9) to yield

$$R_T = \exp \left[\left(\beta - \frac{\alpha}{2} \right)^{1/3} - \left(\beta + \frac{\alpha}{2} \right)^{1/3} \right] \quad (10.10)$$

where

$$\alpha = \frac{A - (1/T)}{C} \quad \beta = \sqrt{\left(\frac{B}{3C} \right)^3 + \frac{\alpha^2}{4}}$$

¹²Developed by John S Steinhart and Stanley R Hart at the Carnegie Institution of Washington in 1968.

In Eqs. (10.9) and (10.10), R_T and T are in ohm and kelvin respectively. The error in the Steinhart-Hart equation is generally less than 0.02 K in the measurement of temperature. The constants, A , B and C can be determined from experimental measurements of resistance (see Example 10.4), or they can be calculated from the R_T vs. T data supplied by the manufacturer.

NTC thermistors can also be characterised with what is called the B - parameter equation. The B -parameter equation, written as

$$\frac{1}{T} = \frac{1}{T_0} + \frac{1}{B} \ln \left(\frac{R_T}{R_0} \right) \quad (10.11)$$

is essentially the Steinhart-Hart equation with $C = 0$. Since, C -value is generally on the order of 10^{-8} , the error will not be high if it is neglected. The inverse form of B -parameter equation is

$$R_T = R_0 \exp \left[B \left(\frac{1}{T} - \frac{1}{T_0} \right) \right] \quad (10.12)$$

where the reference point T_0 is generally 298 K (25°C) at which temperature the constant B is nearly 4000 K. From Eq. (10.12), we get for $T = 298$ K

$$\frac{1}{R_T} \frac{dR_T}{dT} = -\frac{B}{T^2} = -\frac{4000}{298^2} = -0.045/^\circ\text{C}$$

This is evidently a rather high temperature coefficient because for a platinum resistance thermometer the corresponding figure is $0.00385/^\circ\text{C}$. The plot of resistance ratio (R_T/R_{25}) vs. temperature (Fig. 10.15) will also demonstrate this comparison.

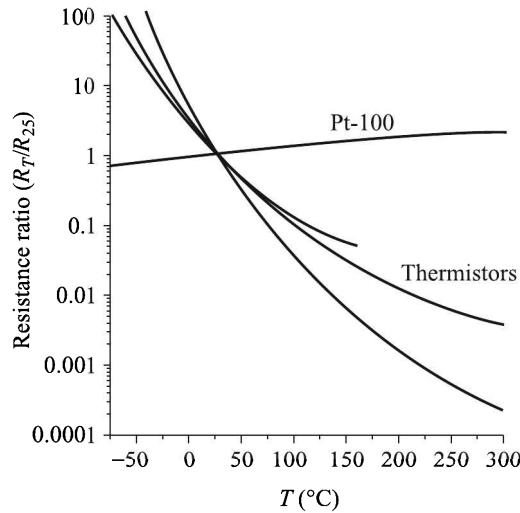


Fig. 10.15 Comparison of resistance ratio (R_T/R_{25}) of platinum with those of three representative NTC thermistors.

Equation (10.11) may alternatively be used to find temperatures by evaluating A and B from a pair of known R_T and T .

Why negative temperature coefficient. A question may be raised, why NTC thermistors have a negative temperature coefficient. The answer can be found from the band structure of solids which tells us that for the semiconductors, the gap between the top of the valence band¹³ and the bottom of the conduction band is small, and the conduction band is empty at $T = 0$ K. The carrier concentration of intrinsic semiconductors is given by the well-known equation¹⁴

$$n_i(T) = 2 \left(\frac{2\pi kT}{h^2} \right)^{\frac{3}{2}} (m_e^* m_h^*)^{\frac{3}{4}} \exp \left(\frac{-E_g}{2kT} \right) \quad (10.13)$$

where k is the Boltzmann's constant
 h is the Planck's constant
 T is the temperature in kelvin
 m_e^* is the effective mass of electron
 m_h^* is the effective mass of hole
 E_g is the band gap.

It may be seen from Eq. (10.13) that the higher the temperature, the higher is the carrier concentration. At higher temperatures, more and more electrons gain energy from the heat supplied and are lifted to the conduction band. This causes the resistivity of semiconductors to go down at higher temperatures. In fact, Eq. (10.12) is an empirical relation based on Eq. (10.13).

Linearisation. With the help of computers, thermistors can easily be used for accurate temperature measurements, the nonlinearity of their temperature-resistance characteristic notwithstanding. However, sometimes it is convenient to use a linear transducer.

A fixed resistor of appropriate value may be placed in parallel with the thermistor to create a more linear resistance versus temperature response for small temperature ranges. The parallel combination of a fixed resistor and a thermistor will flatten and straighten resistance versus temperature response.

Because R_T and R_p are connected in parallel,

$$R_{\text{eff}} = \frac{R_T R_p}{R_T + R_p} \quad (10.14)$$

where R_T is the thermistor resistance
 R_p is the fixed parallel shunt resistor
 R_{eff} is the effective network resistance.

Now, the question is how to determine the value of R_p so that the linearity is maximised. The following procedure will ensure that.

If T_m is the midpoint temperature in kelvin
 R_m is the thermistor resistance at that temperature

¹³Meaning maximum energy level occupied by electrons.

¹⁴See, for example, *Solid State Electronic Devices*, BG Streetman, Prentice-Hall of India, New Delhi (1993), Section 3.3.3.

then, to maximise the linearity of Eq. (10.12), R_p will be chosen such that

$$R_p = \frac{R_m(B - 2T_m)}{(B + 2T_m)} \quad (10.15)$$

where B is expressed in kelvin.

The procedure has been made clear in Example 10.7 by considering an actual situation. It may be seen from the graph (Fig. 10.16) that while the parallel shunt improves the linearity, it decreases the sensitivity. The decrease in sensitivity necessitates a larger gain in the analogue amplifier. But that makes the system more susceptible to noise.

On the other hand, without linearisation, the system is more sensitive (higher slope in the R_T vs. T curve) at lower temperatures. This means that the system will work better for lower temperatures than for higher temperatures. For small temperature ranges, this disparity is not a problem. But for large temperature ranges, a linearised system is desirable.

Example 10.4

The following data were obtained by measuring the resistance of a 5 k Ω (at room temperature) thermistor at three different temperatures.

| | | | |
|----------------------------|-------|------|------|
| T ($^{\circ}\text{C}$) | 0 | 25 | 50 |
| R_T (Ω) | 16330 | 5000 | 1801 |

Evaluate the Steinhart-Hart parameters for the thermistor.

Solution

From the given data, the following equations can be set:

$$\frac{1}{273} = A + B \ln(16330) + C [\ln(16330)]^3$$

$$\frac{1}{298} = A + B \ln(5000) + C [\ln(5000)]^3$$

$$\frac{1}{323} = A + B \ln(1801) + C [\ln(1801)]^3$$

or,

$$A + 9.70076B + 912.88731C = 3.66300 \times 10^{-3}$$

$$A + 8.51719B + 617.85917C = 3.35570 \times 10^{-3}$$

$$A + 7.49610B + 421.21677C = 3.09598 \times 10^{-3}$$

By applying Cramer's rule, we can solve the three simultaneous equations as

$$A = \frac{\begin{vmatrix} 3.66300 \times 10^{-3} & 9.70076 & 912.88731 \\ 3.35570 \times 10^{-3} & 8.51719 & 617.85917 \\ 3.09598 \times 10^{-3} & 7.49610 & 421.21677 \end{vmatrix}}{\Delta} = \frac{-0.0880}{\Delta}$$

$$B = \frac{\begin{vmatrix} 1 & 3.66300 \times 10^{-3} & 912.88731 \\ 1 & 3.35570 \times 10^{-3} & 617.85917 \\ 1 & 3.09598 \times 10^{-3} & 421.21677 \end{vmatrix}}{\Delta} = \frac{-0.0162}{\Delta}$$

$$C = \frac{\begin{vmatrix} 1 & 9.70076 & 3.66300 \times 10^{-3} \\ 1 & 8.51719 & 3.35570 \times 10^{-3} \\ 1 & 7.49610 & 3.09598 \times 10^{-3} \end{vmatrix}}{\Delta} = \frac{-6.3842 \times 10^{-6}}{\Delta}$$

$$\Delta = \begin{vmatrix} 1 & 9.70076 & 912.88731 \\ 1 & 8.51719 & 617.85917 \\ 1 & 7.49610 & 421.21677 \end{vmatrix} = -68.5102$$

Therefore,

$$A = 0.0013 \quad B = 2.3641 \times 10^{-4} \quad C = 9.3185 \times 10^{-8}$$

Example 10.5

For a certain thermistor $\beta = 3100$ K and its resistance at 20°C is known to be 1050Ω . The thermistor is used for temperature measurement and the resistance measured is 2300Ω . Find the measured temperature if the temperature-resistance characteristics of the thermistor is given by

$$R = R_0 \exp \left[\beta \left(\frac{1}{T} - \frac{1}{T_0} \right) \right]$$

where T is in kelvin.

Solution

Here, $R_0 = 1050 \Omega$, $T_0 = 293$ K. Now, the given relation can be written as

$$\frac{1}{T} = \frac{1}{T_0} - \frac{1}{\beta} \ln R_0 + \frac{1}{\beta} \ln R_T$$

Hence from the given data, we get

$$\frac{1}{T} = \frac{1}{293} - \frac{\ln 1050}{3100} + \frac{\ln 2300}{3100} = 3.6659 \times 10^{-3}$$

which gives

$$T = 272.8 \text{ K} \cong 0^\circ\text{C}$$

Example 10.6

The resistance of a thermistor is 800Ω at 50°C and $4 \text{ k}\Omega$ at the ice-point. Calculate the characteristic constants (A , B) for the thermistor and the variation in resistance between 30°C and 100°C .

Solution

Putting $C = 0$ in Eq. (10.9), we get the following conditions

$$A + B \ln 800 = \frac{1}{323}$$

$$A + B \ln 4000 = \frac{1}{273}$$

which give

$$A = 7.4084 \times 10^{-4}$$

$$B = 3.5232 \times 10^{-4}$$

The variation in resistance can be calculated from the relation

$$R_T = \exp \left[\left(\frac{1}{T + 273} - A \right) \div B \right]$$

as

| T (°C) | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|-----------|--------|--------|-------|-------|-------|-------|-------|-------|
| R_T (Ω) | 1428.9 | 1059.3 | 800.0 | 614.5 | 479.3 | 379.1 | 303.8 | 246.3 |

Example 10.7

The following data were obtained for a thermistor:

| T (°C) | 25.0 | 30.1 | 35.3 | 40.1 | 45.2 | 50.2 |
|-----------|--------|-------|-------|-------|-------|-------|
| R_T (Ω) | 1101.0 | 911.3 | 754.8 | 636.0 | 533.7 | 451.1 |

Find the shunt resistance to linearise the temperature-resistance characteristic at this range and show the linearity by plotting a suitable graph.

Solution

We take $T_0 = 25.0^\circ\text{C} = 298.2\text{ K}$. Therefore, from the given data, $R_0 = 1101.0\ \Omega$. The average value of B , i.e. B_{av} can be calculated using Eq. (10.11) as follows, the last row indicating mean (average) values of corresponding columns:

$$B = \frac{\ln(R_T/R_0)}{(1/T) - (1/T_0)}$$

| T | $(1/T - 1/T_0)$ | $\ln(R_T/R_0)$ | B |
|--------|-------------------------|----------------|--------|
| 303.3 | -6.14×10^{-5} | -0.1891 | 3079.8 |
| 308.5 | -1.17×10^{-4} | -0.3775 | 3226.5 |
| 313.3 | -1.667×10^{-4} | -0.5488 | 3292.1 |
| 318.4 | -2.178×10^{-4} | -0.7241 | 3324.6 |
| 323.4 | -2.664×10^{-4} | -0.8923 | 3349.5 |
| 313.38 | | | 3254.5 |

Therefore, from Eq. (10.12)

$$\begin{aligned} R_m &= R_0 \exp \left[B_{av} \left(\frac{1}{T} - \frac{1}{T_0} \right) \right] \\ &= 1101.0 \exp \left[3254.5 \left(\frac{1}{313.38} - \frac{1}{298.2} \right) \right] = 648.9 \Omega \end{aligned}$$

From Eq. (10.15)

$$R_p = \frac{R_m(B_{av} - 2T_m)}{B_{av} + 2T_m} = \frac{648.9(3254.5 - 2 \times 313.38)}{3254.5 + 2 \times 313.38} = 439.3 \Omega$$

Using this value of the shunt resistor, we find the following values of R_{eff} from Eq. (10.14):

| | | | | | | |
|-------------------------------|-------|-------|-------|-------|-------|-------|
| T (K) | 298.2 | 303.3 | 308.5 | 313.3 | 318.4 | 323.4 |
| R_{eff} (Ω) | 314.0 | 294.4 | 277.7 | 259.8 | 241.0 | 222.6 |

The values of T , R_T and R_{eff} are plotted in Fig. 10.16 to show the effect of linearisation.

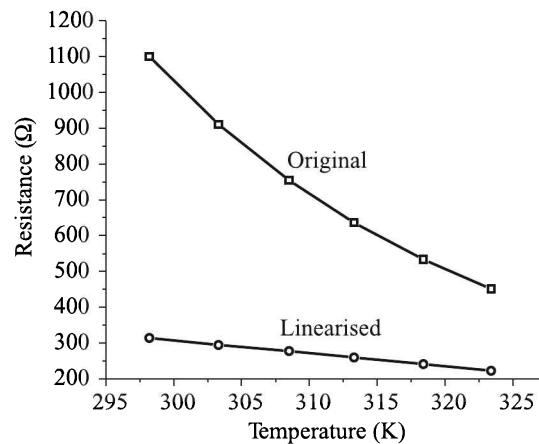


Fig. 10.16 Temperature vs. resistance plot: original and linearised curves.

10.4 Thermoelectricity

Maintaining a thermal gradient in some metals and alloys generate electricity and vice versa. This generation of electricity is called *thermoelectricity*.

Thermoelectricity Generation

The generation of thermoelectricity occurs through three processes known as Seebeck, Peltier and Thomson effects.

Seebeck effect

If two wires or strips of dissimilar metals are welded together at both ends to form a complete circuit and if the two junctions are maintained at different temperatures, an electric current flows through the circuit. The device thus formed is called a *thermocouple* and the phenomenon is called the *Seebeck*¹⁵ effect [Fig. 10.17(a)]. A microvoltmeter of very high input impedance may be included in the circuit to measure the resulting emf which is generally called the *thermo-emf*.

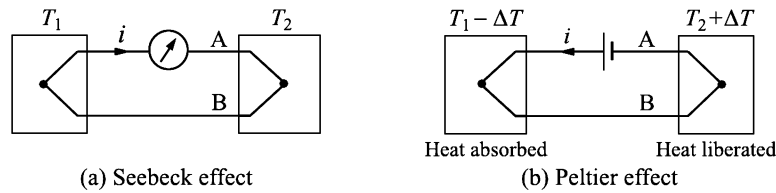


Fig. 10.17 Seebeck and Peltier effects.

Alternatively, the resulting current may be measured by including a milliammeter in the circuit. But measurement of current has an associated problem which we will discuss presently.

Peltier effect

The Seebeck effect is reversible. That means, if a current flows through a thermocouple, heat is absorbed at one junction and liberated at the other. As distinct from the Joule heating which liberates heat irrespective of the direction of current flow, a current flow in the opposite direction in a thermocouple will reverse heat absorption and liberation ends. This inverse effect is called the *Peltier*¹⁶ effect [Fig. 10.17(b)].

Thomson effect

The reversibility of the Peltier effect prompted Lord Kelvin¹⁷ to consider a thermocouple as a heat engine with source at one temperature T_2 and a sink at another temperature T_1 . Then the ratio of the heat absorbed at T_2 to that liberated at T_1 should be T_2/T_1 where these are absolute temperatures. This follows from the derivation of Carnot.

If a charge Q is carried round the circuit, heat absorbed at the hot junction is $\Pi_2 Q$ and that given up at the other is $\Pi_1 Q$. Then,

$$\frac{\Pi_2 Q}{\Pi_1 Q} = \frac{T_2}{T_1}$$

or

$$\frac{\Pi_2 - \Pi_1}{\Pi_1} = \frac{T_2 - T_1}{T_1}$$

where Π_1 and Π_2 , which have the dimension of voltage, are called *Peltier coefficient* or *Peltier emf* and the emf produced by the thermocouple is

$$E = \Pi_2 - \Pi_1 = \Pi_1 \left(\frac{T_2 - T_1}{T_1} \right) \quad (10.16)$$

¹⁵Named after Thomas Johann Seebeck (1770–1831), an Estonian-German physicist.

¹⁶Discovered by Jean Charles Athanase Peltier (1785–1845), a French physicist.

¹⁷Then Sir William Thomson.

It follows from Eq. (10.16) that if the cold junction is maintained at constant temperature, T_1 and Π_1 are constants, and therefore, the emf generated is a linear function of the temperature difference between the junctions. But experiments show that the relation is parabolic with neutral and inversion temperatures as shown in Fig. 10.18.

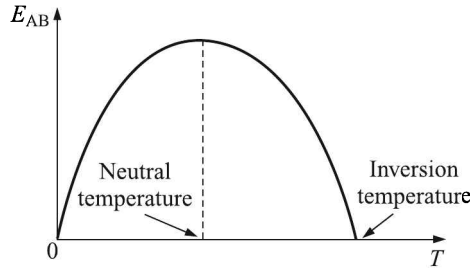


Fig. 10.18 Variation of thermo-emf with temperature.

For a copper-iron thermocouple, the neutral temperature is 275°C and the inversion temperature is 550°C if the cold junction is maintained at the ice-point.

Thomson, therefore, concluded that Seebeck emf was not the sole source of emf in a thermocouple and proposed that there must be another emf in a homogeneous metal wire whenever there is a temperature gradient in it.

Indeed, it may be experimentally observed that if a current is allowed to flow through a copper wire maintained in a temperature gradient as shown in Fig. 10.19(a), the left half becomes cooler than the right-half. Cadmium, antimony, silver and zinc behave in the opposite way [Fig. 10.19(b)]. The former behaviour is called the *positive Thomson effect* and the latter, the *negative Thomson effect*.

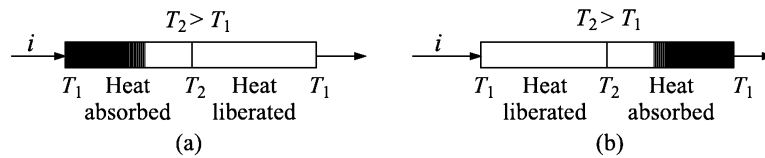


Fig. 10.19 Thomson effect: (a) positive, and (b) negative.

Thus, the total thermo-emf is the sum of Seebeck and Thomson emfs. While the former varies as the temperature difference between the junctions, the latter varies as the difference between squares of junction temperatures. That is,

$$E = C_1(T_2 - T_1) + C_2(T_2^2 - T_1^2) \tag{10.17}$$

However, Eq. (10.17) is only an approximate expression for the thermo-emf generation by temperature difference between the two junctions of a thermocouple.

Why thermoelectric effects occur can be understood from the band structure of solids as discussed in the following section.

Explanation of Thermoelectric Effects

The energy band models of two dissimilar metals are shown in Fig. 10.20(a). Here, Φ_A and Φ_B are work functions¹⁸ corresponding to metals A and B respectively. When these two metals are brought in contact, their Fermi levels equalise. As a result some electrons from A flow to B, leaving A and B positively and negatively charged respectively as shown in Fig. 10.20(b).

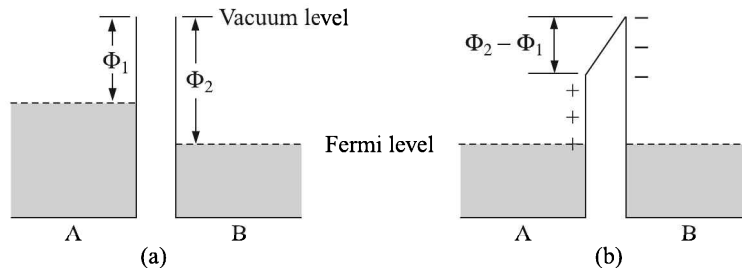


Fig. 10.20 (a) Metals A and B and their energy bands, and (b) metals A and B and their energy bands when A and B are joined together.

This produces a potential difference between the two metals called the *contact potential* V defined by

$$eV = \Phi_A - \Phi_B$$

Because of the high conductivity of metals, a local field cannot exist, but a change in the free electron distribution takes place at the contact surface. Now, if an external voltage is applied across the metal-metal junction, a current I flows and heat is generated (or absorbed, depending on the direction of current flow) because electrons release some energy when moving down the potential hill (in the reverse case, they have to acquire energy from the surroundings to climb the hill). This heating is over and above the normal I^2R Joule heat production which is direction-independent. This phenomenon is called the Peltier effect.

Now suppose, two junctions are kept at different temperatures. The electrons at the hot junction gain energy so that the Fermi level at the hot junction is higher than that at the cold junction. This causes an electron flow from the hot junction to the cold junction which is the Seebeck effect.

In the same way, there will be an electron flow from the hotter part of the same metal to its colder part. That is the Thomson effect.

Thermo-emf Measurement

From this brief discussion on the thermo-emf, the following points emerge:

1. The magnitude of the emf depends on the materials and the temperature difference between the junctions.
2. The process converts heat energy to electrical energy. The conversion is reversible.
3. Measurement of thermo-emf cannot be linked to temperature measurement unless the current flow within the thermocouple is inhibited. Because, if current is allowed to flow, it will change the temperatures of the junctions (Peltier effect).

¹⁸Work function is the work that must be done to eject an electron from the solid. At 0 K, the work function equals the energy difference between the Fermi level and the so called vacuum (or continuum) level.

From the third point, one may notice that potentiometers are ideal for measuring thermo-emfs. An op-amp voltage follower¹⁹ coupled with a microvoltmeter is almost equally good.

Thermocouple Laws

The behaviour of thermocouples can be summarised into two laws:

1. Law of intermediate temperatures
2. Law of intermediate metals

Law of intermediate temperatures

If a thermocouple is made of two metals A and B (for convenience, henceforward we will mention it as AB thermocouple), then this implies:

1. If the junctions of the thermocouple are maintained at temperatures T_1 and T_2 , then it will produce the same emf whatever be the value of the temperature at other parts of the thermocouple [Fig. 10.21(a)].
2. If E_{12} is the emf generated for junctions at T_1 and T_2 , E_{23} is that for junctions at T_2 and T_3 , then the emf E_{13} for junctions at T_1 and T_3 is $E_{12} + E_{23}$ [Fig. 10.21(b)].

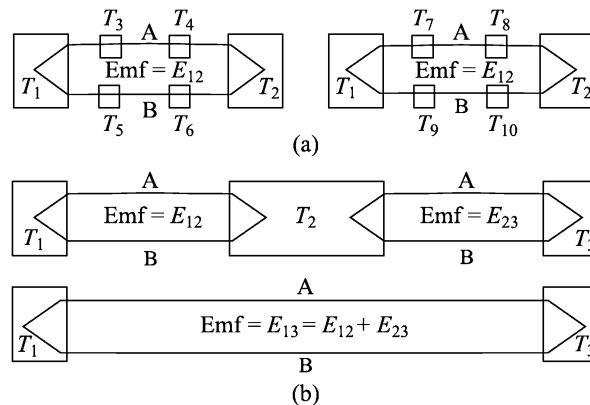


Fig. 10.21 Implications of the law of intermediate temperatures.

Law of intermediate metals

Suppose the junctions are maintained at T_1 and T_2 . Then this law has the following implications:

1. If we have an AB thermocouple, and another with AB junctions, but another metal C incorporated somewhere in between, then the net emf will be the same in both the cases [Fig. 10.22(a)].
2. If E_{AB} is the emf generated by the AB thermocouple and E_{BC} is that by the BC thermocouple, then the emf generated by the AC thermocouple will be $E_{AB} + E_{BC}$ [Fig. 10.22(b)].

¹⁹See Section 16.2 at page 762.

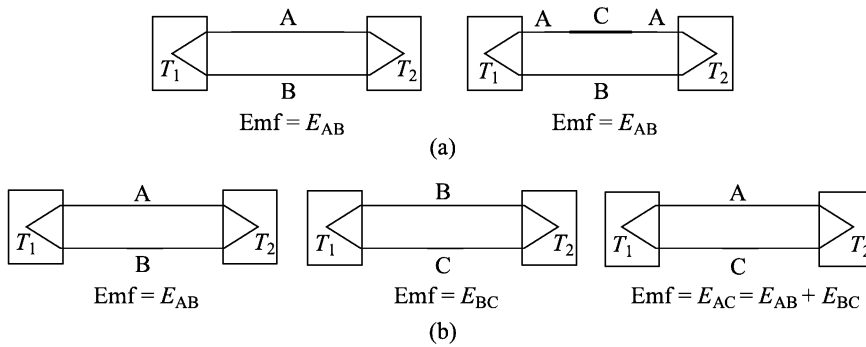


Fig. 10.22 Implications of the law of intermediate metals.

Example 10.8

The extension wires of an iron-constantan thermocouple were improperly wired as shown in Fig. 10.23. The voltmeter calibrated in °C will then read

- (a) 240
- (b) 180
- (c) 140
- (d) 120

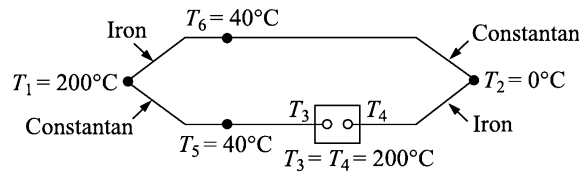
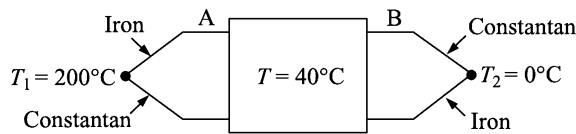


Fig. 10.23 Thermocouple (Example 10.8).

Solution

The arrangement is equivalent to that presented below.



The thermocouple A measures a temperature of $(200 - 40) = 160^\circ\text{C}$ and the thermocouple B measures a temperature of $(0 - 40) = -40^\circ\text{C}$. The rest of the intermediate temperatures can be ignored following the law of intermediate temperatures. So, the thermocouple AB will indicate a temperature of $(160 - 40) = 120^\circ\text{C}$. Ans. (d).

Example 10.9

The table provides the thermo-emf sensitivity of five materials with reference to platinum around 273 K.

| Material | Constantan | Nickel | Copper | Iron | Nichrome |
|-------------------------------------|------------|--------|--------|-------|----------|
| Sensitivity (μVK^{-1}) | -35 | -25 | +6 | +18.5 | +25 |

- (i) The thermocouple pair that gives the maximum sensitivity around 273 K is
- (a) Platinum-constantan (b) Nichrome-constantan
(c) Nickel-constantan (d) Copper-nickel
- (ii) Two copper constantan thermocouples are connected such that the two constantan wires are joined together. The two copper wires are connected to the input of a low noise chopper stabilised differential amplifier having a gain of 1000. One of thermocouple junctions is immersed in a flask containing ice and water in equal proportions. The other thermocouple is at a temperature T . If the output of the amplifier is 2.050 V, the temperature T
- (a) 205 °C (b) 102.5 °C
(c) 51.25 °C (d) 50 °C

Solution

The calculation of thermo-emf sensitivities of the given pairs of materials is presented in the following table. The material 'A' is platinum. E_{AB} and E_{AC} are obtained from the given data. And, from the law of intermediate metals, $E_{BC} = E_{BA} + E_{AC} = -E_{AB} + E_{AC}$.

| B | C | E_{AB} | E_{AC} | $E_{AC} - E_{AB} = E_{BC}$ | $ E_{BC} $ |
|----------|------------|----------|----------|----------------------------|------------|
| Nichrome | Constantan | 25 | -35 | $-35 - 25 = -60$ | 60 |
| Nickel | Constantan | -25 | -35 | $-35 + 25 = -10$ | 10 |
| Copper | Nickel | +6 | -25 | $+6 + 25 = 31$ | 31 |
| Copper | Constantan | +6 | -35 | $+6 + 35 = 41$ | 41 |

- (i) From the calculation, it is apparent that the nichrome-constantan thermocouple pair gives the maximum sensitivity.
Ans. (b)
- (ii) The last row in the above calculation shows that the thermocouple pair copper-constantan gives a sensitivity of 41×10^{-6} V/°C. Since the output is 2.050 V after being amplified 1000 times, the actual output of the thermocouple is 2.050×10^{-3} V. That corresponds to a temperature of $T = \frac{2.05 \times 10^{-3}}{41 \times 10^{-6}} = 50$ °C.
Ans. (d)

Common Thermocouples

A variety of thermocouples today cover a range of temperatures from -270°C to $+1800^{\circ}\text{C}$. They are given letter designations of B , E , J , K , N , R , S and T . Table 10.5 gives an idea about their composition, and typical attributes. The letter type identifies a specific temperature-voltage relationship, not a particular chemical composition. Manufacturers may fabricate thermocouples of a given type with variations in composition; however, the resultant temperature versus voltage relationships must conform to the thermoelectric voltage standards associated with the particular thermocouple type.

Table 10.5 Relevant data of common thermocouples and their characteristics

| ANSI type | Junction materials (+/-) | Typical range (°C) | Nominal sensitivity (µV/°C) | Atmospheric media ^f | | | |
|-----------|---|--------------------|-----------------------------|--------------------------------|-----|-----|-----|
| | | | | I | R | O | V |
| <i>E</i> | Chromel ^a /Constantan ^b | 0 to 900 | 76 | Yes | No | Yes | No |
| | Highest sensitivity among thermocouples commonly used. Low drift. Good corrosion resistance. Use is less widespread than other base-metal thermocouples owing to their low useful range. | | | | | | |
| <i>J</i> | Iron/Constantan | -200 to 760 | 55 | Yes | Yes | Yes | Yes |
| | Iron oxidises rapidly above 540°C. Should not be used above 760°C due to an abrupt magnetic transformation at the Curie point of iron (~770°C) which changes its characteristic and can cause permanent de-calibration. Widely used in industry due to their high sensitivity and low cost. | | | | | | |
| <i>K</i> | Chromel/Alumel ^c | -200 to 1260 | 39 | Yes | No | Yes | No |
| | Not recommended in sulphur environments. Cycling at high temperatures can cause calibration drift. Most commonly used base-metal thermocouple. | | | | | | |
| <i>T</i> | Copper/Constantan | -200 to 400 | 45 | Yes | Yes | Yes | Yes |
| | Rust and corrosion resistant. Best for sub-zero temperatures. | | | | | | |
| <i>N</i> | Nicrosil ^d /Nisil ^e | 0 to 1100 | 10.4 | Yes | Yes | Yes | Yes |
| | Not recommended in sulphur environments. Improved resistance to drift and better stability over <i>K</i> and <i>E</i> at elevated temperatures. | | | | | | |
| <i>R</i> | Platinum-13% Rhodium/ Pure Platinum | 0 to 1593 | 6 | Yes | No | Yes | Yes |
| | High temperature use. Usually with a ceramic sheath. Granular precipitation from metal protection tubes can cause failure or calibration drift. | | | | | | |
| <i>S</i> | Platinum-10% Rhodium/ Pure Platinum | 0 to 1538 | 10.4 | Yes | No | Yes | Yes |
| | Because of its high stability, used as the standard for calibrating the melting point of Gold. | | | | | | |
| <i>B</i> | Platinum-30%Rhodium/ Platinum-6% Rhodium | 50 to 1800 | 7.7 | Yes | No | Yes | Yes |
| | Owing to its increased Rhodium content, it is not so stable as the <i>R</i> or <i>S</i> types. | | | | | | |

^a chromium-nickel alloy^b copper-nickel alloy^c nickel-aluminium-silicon-manganese alloy^d nickel-chromium-silicon-magnesium alloy^e nickel-silicon alloy^f I means *inert atmosphere* which refers to an environment of a gaseous mixture that contains little or no oxygen and primarily consists of non-reactive gases, or gases that have a high threshold before they react, R means *reducing atmosphere* which refers to an environment in which oxidation is prevented by the removal of oxygen and other oxidising gases or vapours, O means *oxidising atmosphere* which is a gaseous environment in which the oxidation of solids readily occurs due to the presence of excess oxygen, and V means a vacuum.

Three additional thermocouple types used for high temperature measurements are *C*, *D*, and *G*. Their designation letters (*C*, *D*, *G*) are not recognised as standards by ANSI²⁰. Nevertheless, they are commercially available. Their wire compositions are:

| <i>Type</i> | <i>Composition</i> |
|-------------|---|
| <i>C</i> | Tungsten-5%Rhenium/ Tungsten-26%Rhenium |
| <i>D</i> | Tungsten-5%Rhenium/Tungsten-25%Rhenium |
| <i>G</i> | Tungsten/Tungsten-26%Rhenium |

Construction

The basic wire type thermocouple is constructed by joining homogeneously two dissimilar metals at one end to form the measuring junction. Junctions may be formed by welding, soldering or crimping the two metals, but the emf generated will be identical in all cases. All wire-type thermocouples have an exposed junction. While wire-type thermocouples offer good response time, ruggedness, and high temperature use, they are susceptible to environmental conditions and therefore must be protected.

Mineral insulated (MI) thermocouples. Mineral insulation is provided to overcome the disadvantages of wire type construction by embedding the thermocouple wires in ceramic insulation and protecting them with a metallic sheath. The mineral insulated cable (MI cable) design is based on small mass and high thermal conductivity which in turn promotes rapid heat transfer from the heat source to the measuring junction.

Impervious to most liquids and gases, the sheaths can withstand high external pressures. The seamless design protects against moisture or other contaminants attacking the thermocouple elements.

Since the only materials used to make the MI cable are the thermocouple conductors, the mineral oxide insulation and the metallic sheath, the cables are inherently fireproof thus providing an intrinsically safe temperature measuring system.

Polynomial equation representation

A polynomial equation is generally used to convert thermocouple voltage to temperature ($^{\circ}\text{C}$) over a wide range of temperatures. Standard equations have been developed for each type of thermocouple. These power series equations use unique sets of coefficients a_n which are different for different temperature segments within a given thermocouple type. Unless otherwise indicated, all standard thermocouple equations and tables are referred to a cold junction temperature of 0°C . The general form of the equation is

$$T = \sum_{n=0}^N a_n E^n \quad (10.18)$$

where T is the temperature in $^{\circ}\text{C}$ corresponding to the thermo-emf E generated by the thermocouple.

²⁰American National Standards Institute, see www.ansi.org

The coefficients are tabulated in many places²¹. We present the polynomial coefficients for a type E thermocouple in Table 10.6.

Table 10.6 Polynomial coefficients for type E thermocouple

| n | $a_n (^{\circ}\text{C}/\text{mV})$ | |
|-----|------------------------------------|------------------------------|
| | -270 to 0 $^{\circ}\text{C}$ | 0 to 1000 $^{\circ}\text{C}$ |
| 0 | 0.00000×10^{00} | 0.00000×10^{00} |
| 1 | 1.69773×10^{01} | 1.70570×10^{01} |
| 2 | -4.35150×10^{-01} | -2.33018×10^{-01} |
| 3 | -1.58597×10^{-01} | 6.54356×10^{-03} |
| 4 | -9.25029×10^{-02} | -7.35627×10^{-05} |
| 5 | -2.60843×10^{-02} | -1.78960×10^{-06} |
| 6 | -4.13602×10^{-03} | 8.40362×10^{-08} |
| 7 | -3.40340×10^{-04} | -1.37359×10^{-09} |
| 8 | -1.15649×10^{-05} | 1.06298×10^{-11} |
| 9 | 0.00000×10^{00} | -3.24471×10^{-14} |

Nonlinearity of characteristics

Almost all practical temperature ranges can be covered using thermocouples. But their full-scale voltage output is only millivolts with sensitivities in the microvolts per degree range and their response is nonlinear. Figure 10.24 displays typical voltage-temperature characteristics of common thermocouples. These curves provide a visual indication of thermocouple ranges and sensitivities (from the slopes of the curves).

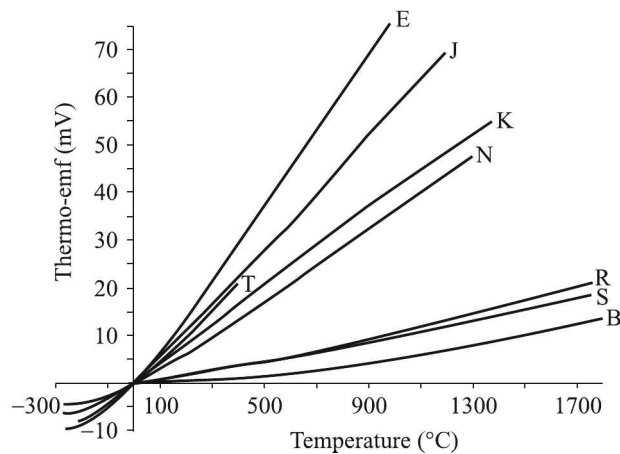


Fig. 10.24 Thermo-emf vs. temperature curves for common thermocouples.

²¹Complete sets of temperature vs. voltage tables (cold junction at 0 $^{\circ}\text{C}$) and polynomial equations for all popular industry standard thermocouples are downloadable from NIST (National Institute of Standards and Testing, USA, see <http://srdata.nist.gov/its90/main/>).

Although the curves appear to be linear because they are plotted in millivolt scale, they are actually nonlinear as may be seen from Fig. 10.25 which illustrates thermocouple nonlinearity by plotting the difference between an ideal linear response and the response of a type J thermocouple over the range of 0 to 150°C. The nonlinearity is also apparent from Eq. (10.18).

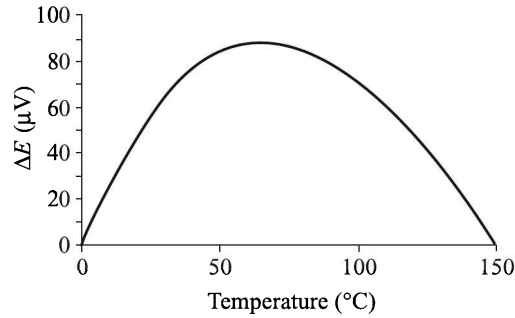


Fig. 10.25 Difference between an ideal linear response and the response of a type J thermocouple over the range of 0 to 150°C.

But the nonlinearity poses no problem to temperature measurement by thermocouples because of the existence of excellent thermo-emf vs. temperature relations as well as look-up tables supplied by manufacturers.

Cold Junction Compensation

The emf generated by a thermocouple, as shown by Eq. (10.18), is dependent not only on the temperature of the measuring junction, but also on that of the cold junction. So, if we make an arrangement as shown in Fig. 10.26(a) for temperature measurement of T_1 , it may generate different emf at different times although T_1 may be steady, because we have unwittingly left the other junction to an uncertain temperature T . In industry, it is not practical to maintain

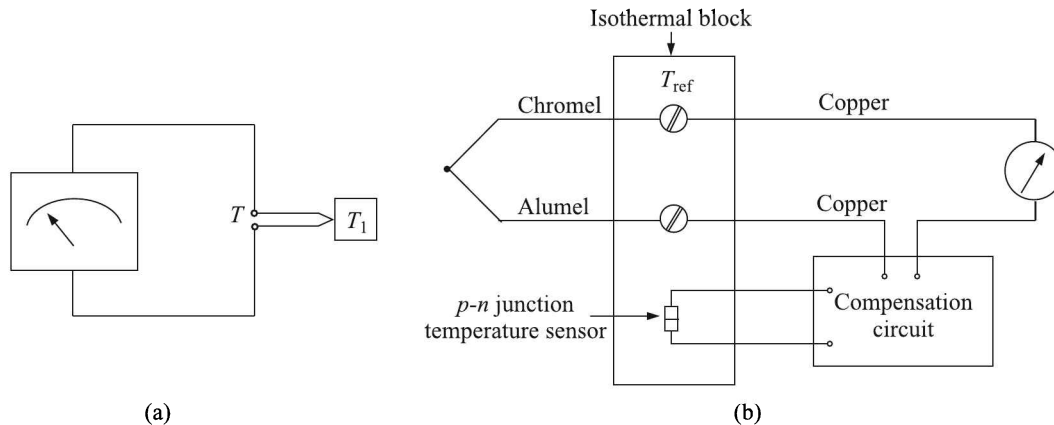


Fig. 10.26 Temperature measurement by thermocouple: (a) cold junction left at an arbitrary temperature, (b) cold junction temperature compensated for.

an ice point and put the reference junction there. In that case, apart from the compensation shown in Fig. 10.26(b), another type of compensation can be provided through emf generation by a bridge circuit as discussed in Section 16.1 at page 753.

Practical Considerations

The following points need to be considered when measuring temperatures with thermocouples.

1. Connectors should have no thermal gradients across the individual connections.
2. Thermocouple behaviour depends on the molecular structures of materials. Environmental conditions such as strain, chemical corrosion, radiation, etc. that affect molecular structure anywhere along the length of the thermocouple wire can create errors. For example, thermocouples with iron composition are subject to rust. That can cause errors.
3. Twisted pair extension wires and signal conditioning modules with adequate filtering should be used to avoid EMI and RI errors.
4. Thermocouple lead lengths should be short. Extension wires recommended by manufacturer should be used if long thermocouple leads are necessary.
5. Hostile corrosive environments combined with moisture and heat may stimulate galvanic action and create electrochemical voltage errors.

Thermocouples are suitable for measuring quickly varying temperatures as well as for measuring temperatures at a point. However, they are suitable for neither precision measurements nor remote measurements. But they are inexpensive, rugged, reliable, and can be used over a wide temperature range.

Example 10.10

A copper-constantan thermocouple was found to have linear calibration between 0°C and 400°C with emf at maximum temperature (w.r.t. cold junction temperature at 0°C) equal to 20 mV.

- (a) Determine the correction which must be made to the indicated emf if the cold junction temperature is 20°C .
- (b) If the indicated emf is 8.90 mV (w.r.t. cold junction temperature at 20°C) in the thermocouple circuit, determine the temperature of the hot junction.

Solution

- (a) Sensitivity of the thermocouple is

$$\frac{20}{400} \text{ mV}/^\circ\text{C} = 0.05 \text{ mV}/^\circ\text{C}$$

The emf corresponding to $20^\circ\text{C} = (20 \times 0.05) \text{ mV} = 1.0 \text{ mV}$. Therefore, 1.0 mV has to be added to the indicated emf if the cold junction is at 20°C .

- (b) Corrected emf = $8.90 + 1.0 \text{ mV} = 9.9 \text{ mV}$. The corresponding temperature of the hot junction is

$$\frac{9.9}{0.05} = 198^\circ\text{C}$$

Integrated Circuit Sensors

Integrated circuit (IC) temperature sensors are ubiquitous nowadays. Apart from household body temperature measuring thermometers, PC and automotive applications, they are used as temperature sensors in many electronic gadgets. Mobile phones usually include one or more sensors in the battery pack, and notebook computers might have four or more sensors for checking temperatures in the CPU, battery, ac adapter, and PCMCIA card cage. These applications do not cover the enormous number of thermal-shutdown and thermal-protection circuits that designers build into all sorts of ICs as a final defence against short circuits and over-clocking (exceeding the IC's specified clock speed).

Of course, they cannot always replace the traditional temperature sensors like RTDs, thermistors and thermocouples, but IC temperature sensors offer many advantages. They require no linearisation or cold-junction compensation, for instance. Rather, they often provide cold-junction compensation for thermocouples. They generally provide better noise immunity through higher-level output signals, and some provide logic outputs that can interface directly to digital systems.

These ICs generate electrical output proportional to the temperature. Their principle of operation is as follows.

Principle of operation

The sensor works on the principle that the forward voltage of a silicon diode depends on its temperature. The following simplified equation shows the voltage-temperature relationship:

$$V_F = \frac{kT}{e} \ln \frac{I_F}{I_S} \quad \text{for } I_F \gg I_S \quad (10.19)$$

where T is the ambient temperature in degrees kelvin

k is the Boltzmann's constant (1.3807×10^{-23} J/K)

e is the charge of an electron (1.602×10^{-19} coulomb)

I_F is the forward current

I_S is the saturation current

I_S is a constant defined by the diode size. Now, if a constant forward current I_F is used to bias the diode, only T remains a variable in Eq. (10.19). However, I_S is dependent not on T only, it varies considerably from diode to diode. To obviate this difficulty, a two-diode solution is resorted to. If both diodes are biased with constant forward currents I_{F1} and I_{F2} , and $I_{F2}/I_{F1} = n$, the difference between the two forward voltages ΔV_F has no dependence on the saturation currents of the two diodes as can be seen from the following

$$\Delta V_F = V_{F1} - V_{F2} = \frac{kT}{e} \cdot \ln \left(\frac{I_{F1}/I_S}{nI_{F1}/I_S} \right) = \left| \frac{kT}{e} \cdot \ln(n) \right|$$

ΔV_F is also called the voltage proportional to absolute temperature, V_{PTAT} . V_{PTAT} provides a linear voltage change with a slope of

$$86 \cdot \ln(n) \Big|_{n=10} = 200 \mu\text{V}/^\circ\text{C}$$

at the room temperature. This voltage is either amplified for an analogue output or interfaced with an ADC to produce a digital output.

The excessive leakage currents characteristic of silicon p-n junctions limits the temperature for IC-based sensors to about 200°C. As a rule of thumb, these currents double with every 10°C rise in temperature. Excessive leakage current causes malfunctions in the bandgap reference ΔV_F and signal-conditioning circuitry.

The accuracy of V_{PTAT} over the specified range depends on the matching of I_F and I_S of the two diodes. Any mismatch contributes to the error in the indicated temperature and nonlinearity.

IC temperature sensors are available in both voltage and current output configurations. The current output units are usually set for an output change of 1 $\mu\text{A}/\text{K}$ while the voltage output configuration generates 10 mV/K. With this background, we will briefly discuss three representative IC temperature sensors—LM335, LM 35²², and AD592²³— here.

LM335. The LM335 is a precision temperature sensor which can be easily calibrated. It operates as a 2-terminal Zener diode with a voltage output of 10 mV/K. That means, at 25°C (298.2 K) it acts like a 2.982 V Zener (Fig. 10.27).

It comes with an accuracy of $\pm 1^\circ\text{C}$ and it can be externally trimmed. A single point calibration improves its accuracy to $\pm 0.5^\circ\text{C}$ over a range of -55°C to $+125^\circ\text{C}$. Unlike other sensors, the LM335 has a linear output.

LM35. The output voltage of LM35 is linearly proportional to the Celsius (centigrade) scale of temperature.

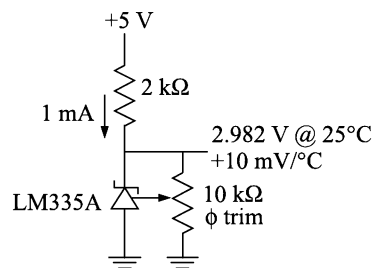


Fig. 10.27 LM335 Zener diode temperature sensor.

The LM35 covers a temperature range of -55°C to 150°C . Its overall accuracy is $\pm 0.75^\circ\text{C}$ over the entire range and that near the room temperature is $\pm 0.25^\circ\text{C}$. It has an internal offset that produces 0 V at 0°C . Therefore, no external trimming or calibration is necessary. Its other features include:

1. Its output of + 10.0 mV/°C is nearly linear.
2. It behaves as a 3-terminal reference (rather than a 2-terminal Zener) powered by +4 to +20 V applied to a third terminal. But plus and minus supplies with a ‘pulldown’ resistor are necessary to operate it near or below 0°C [Fig. 10.28(b)].
3. As it draws only 60 μA from the supply, its self-heating is very low (0.08°C in still air) .

Its cousin, the LM34, works similarly but with a readout in Fahrenheit.

²²Produced by National Semiconductor, USA

²³Produced by Analog Devices, USA

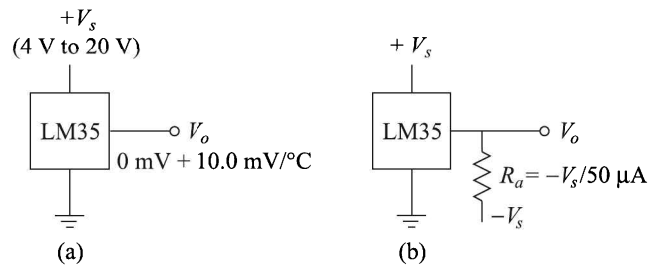


Fig. 10.28 LM35 connections: (a) single power supply (for $+2^{\circ}\text{C}$ to $+150^{\circ}\text{C}$), and (b) plus-minus supplies (for full range).

AD592. AD592 is a 2-terminal device that acts like a constant current element passing a constant current in microamps equal to the absolute temperature. That means, at 25°C (298.2 K) it behaves like a constant current regulator of $298.2 \pm 0.05\ \mu\text{A}$. Its operating range is from -25°C to $+105^{\circ}\text{C}$.

It comes with an accuracy of $\pm 1^{\circ}\text{C}$ and it can be externally trimmed. A single point calibration improves its accuracy to $\pm 0.8^{\circ}\text{C}$ over its entire range.

Of course, there is a host of other IC temperature sensors manufactured by other vendors. But they belong to either of the three categories we have discussed here.

Table 10.7 gives a comparison of advantages and disadvantages of widely used temperature measurement technologies.

Table 10.7 Advantages and disadvantages of widely used temperature sensors

| Device | Advantages | Disadvantages |
|--------------|--|--|
| RTD | Linear. High stability. Wide range of operating temperature. Interchangeable over wide temperature range. | Rather low sensitivity. Relatively slow response. Low resistance requires three- or four-wire measurement. Sensitive to shock and vibration. Voltage source required. Expensive. |
| Thermistor | High stability. Fast response. High sensitivity. High resistance eliminates the need for four-wire measurement. Small size. Interchangeable. | Nonlinear. Limited operating temperature. Interchangeable over relatively narrow temperature ranges. Voltage source required. Inexpensive. |
| Thermocouple | Simple. Wide range of operating temperature. No external power supply required. Rugged. Inexpensive. | Nonlinear. Relatively low stability. Low sensitivity. Low voltage output can be affected by RI and EMI. Cold junction compensation required. |
| IC | Linear. High sensitivity. Inexpensive. | Limited range of operating temperature. Power supply required. Subject to self-heating. Limited configurations. |

10.5 Fibre-optic Sensors

Optical fibres are affected by temperature in several ways. These include:

1. Microbending
2. Change of refractive index

3. Change of polarisation of the incident light
4. Change of length of the fibre
5. Interferometric effects
6. Diffraction grating effects of the fibre
7. Raman scattering effect
8. Sagnac effect

These characteristics allow them to be used for sensing temperatures.

Microbending Type

The microbending type fibre-optic sensors utilise the lowering of detected light intensity to measure temperature. A self-explanatory schematic diagram of such a device, using a bimetallic element, is shown in Fig. 10.29.

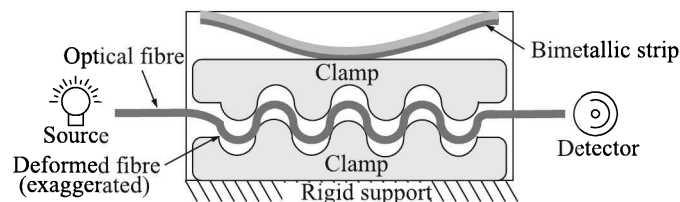


Fig. 10.29 Schematic diagram of microbending type fibre-optic temperature sensor.

Change of Refractive Index, Polarisation and Length Measurement Type

To construct a fibre-optic temperature sensor for types (b), (c) and (d) as stated above, the sensing element is usually deposited directly on the cleaved end of the fibre and the temperature is deduced from the phase or spectrum (Fig. 10.30) of the reflected beam.

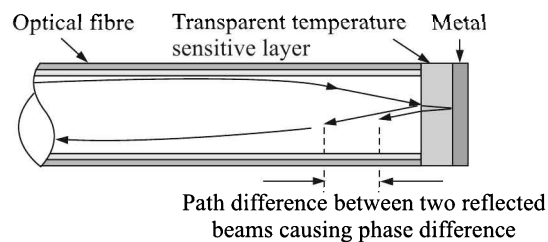


Fig. 10.30 Fibre-optic temperature sensor showing path difference between the incident and reflected beams.

Interferometric Sensors

In essence, a Fabry-Pérot²⁴ fibre-optic temperature sensor provides a temperature-sensitive reflectance spectrum. We have already discussed these sensors in detail in Section 7.3.

²⁴An interferometer named after its inventor Maurice Paul Auguste Charles Fabry (1867–1945), and Jean-Baptiste Alfred Pérot (1863–1925), both French physicists.

Figure 10.31 illustrates an extrinsic Fabry-Pérot interferometric temperature measurement system.

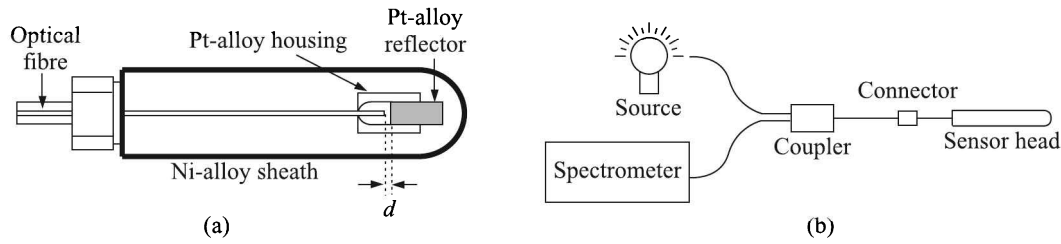


Fig. 10.31 Fabry-Pérot fibre-optic temperature sensing system: (a) sensor cutaway view, and (b) diagram of the arrangement.

The sensor head comprises an inconel (a nickel-alloy) sheath that contains the Fabry-Pérot interferometer, which is located at the tip of an optical fibre that connects the sensor head to the external instrumentation. A small platinum-alloy housing holds a reflector (made of the same alloy) at a short distance d from the fibre-optic tip, which is polished. The gap d defines the interferometer cavity.

White light incident on the optical fibre travels along the fibre to the sensor head. About 4% of the incident light is reflected back along the fibre from the polished tip. The remainder of the incident light travels on to the platinum-alloy reflector. About 90% of the light incident on the reflector re-enters the fibre and propagates back along the fibre, along with the light reflected from the fibre tip.

The fibre-optic coupler directs the two backward-propagating light beams to a spectrometer that analyses the characteristic interference fringes, i.e. a reflected intensity vs. wavelength spectrum. Because of the difference between the coefficients of thermal expansion of the optical fibre and platinum-alloy housing and reflector, d varies with temperature, giving rise to a change in the spectrum. The increase in length may be determined from Eq. (7.41).

This type of temperature sensors have been successfully used to measure temperatures in the range of -50 to $+600^\circ\text{C}$. The sensitivity can be as low as 10^{-8}°C .

Diffraction Grating Sensors

We have discussed in Section 7.3 how to measure strain with the help of fibre Bragg gratings (FBGs). The same FBG can be employed to measure temperature as well. There we have seen that the sensing parameters are usually calculated from the bandwidth change of the principal maximum of the diffraction spectrum. Alternatively, the wavelength shift of the centre of the reflection spectrum can be monitored to yield the same result. The wavelength of the centre of the reflection spectrum is given by the resonance or Bragg condition²⁵

$$\lambda_B = 2n_{\text{eff}}\Lambda$$

where λ_B is the Bragg wavelength

n_{eff} is the effective refractive index of the mode in consideration

Λ is the spatial period of perturbation of the refractive index.

²⁵Proposed by William Lawrence Bragg (1890–1971), who was a British physicist and X-ray crystallographer and his father William Henry Bragg (1862–1942).

Sensitivity

Taking the isotropic and homogeneous properties of silica fibres into account, the fractional change in the Bragg wavelength λ_B with temperature can be computed from

$$\frac{\Delta\lambda_{B,T}}{\lambda_B} = \frac{1}{\lambda_B} \frac{d\lambda_B}{dT} = \left[\frac{1}{n_{\text{eff}}} \frac{\partial n_{\text{eff}}}{\partial T} + \frac{n_{\text{eff}}^2}{2} (p_{11} + 2p_{12}) \alpha_T + \alpha_T \right]$$

where $\Delta\lambda_{B,T}$ is the change in centre wavelength per Kelvin temperature

$\partial n_{\text{eff}}/\partial T$ is thermo-optical coefficient

p_{11} , p_{12} are Pockels coefficients representing the electro-optic effect

α_T is thermal expansion coefficient of the fibre material.

For bare silica fibres $\alpha_T \approx 0.5 \times 10^{-6}/\text{K}$, Pockels²⁶ coefficients $p_{11} = 0.113$, $p_{12} = 0.252$, and $n_{\text{eff}} \approx 1.46$ though it depends on the wavelength and refractive index profile. The thermo-optical coefficient is experimentally determined by the refractive index dependence on the temperature variation, which gives the total derivative dn/dT . Knowing all other relevant material parameters, the partial derivative $\partial n_{\text{eff}}/\partial T$ is found to be equal to 9.7×10^{-6} . This results in a general sensitivity of around 10.5 pm/K at a wavelength of 1550 nm.

The mentioned coefficients depend on temperature leading to nonlinearities in sensor response. Though these nonlinearities can be neglected for modest temperature changes around the room temperature, at cryogenic temperatures the nonlinearities lead to a strongly reduced sensitivity. In order to increase the sensitivity there, the fibre may be bonded to materials with high thermal expansion coefficients. With aluminium substrate, a sensitivity of 20 pm/K at a wavelength of 1500 nm has been achieved at 100 K.

Thermal stability

FBGs show a decay in reflectivity under elevated temperatures. This decay decreases with time and settles to a quasi-stable value for long time.

Therefore, to obtain stable sensors, thermal annealing is usually performed prior to installation. The annealing generates a stability of less than 0.3% at 80°C in reflectivity for about 25 years. For applications exceeding 500°C, the FBGs may be coated with metals to make them suitable for long-term high temperature measurements.

Sensors based on Raman Scattering

While discussing Brillouin scattering (Section 7.3 at page 264), we talked about Raman scattering as well. Unlike Brillouin scattering, the Raman signal is sufficiently strong and distinct to be used in temperature measurement. Here the frequency shift is proportional to the characteristic vibrational frequencies of the molecules. The frequency shift may be positive or negative. Spectral lines corresponding to negative frequency shift are termed *Stokes*²⁷ *lines* and those corresponding to positive frequency shift are termed the *anti-Stokes*

²⁶The Pockels effect (first described in 1906 by the German physicist Friedrich Carl Alwin Pockels (1865–1913)) is the linear electro-optic effect, where the refractive index of a medium is modified in proportion to the applied electric field strength.

²⁷George Gabriel Stokes (1819–1903), was an English mathematician and physicist, who made important contributions to fluid dynamics, optics, and mathematical physics.

lines (see Fig. 7.27). The longer wavelength Stokes line is only weakly temperature sensitive whereas the shorter wavelength anti-Stokes increases with higher temperature.

A short laser pulse is sent along the fibre and the backscattered Raman line is detected with high temporal resolution. The intensity of the Raman line contains information about loss and temperature along the fibre whereas the time between sending the pulse and detecting the backscattered signal provides a measure of the distance along the fibre. The method is widely known as the optical time-domain reflectometry (OTDR). The measurement set-up is very similar to OTDR for Brillouin scattering (see Fig. 7.28).

Rather than measuring temperature at a point, the FBG and Raman scattering fibre-optic sensors are used for distributed temperature sensing (DTS) over a cable length of say a few kilometres. As the DTS system is based on optical fibre technology, it can be used in a wide range of conditions, including hazardous environment and EMI intensive areas, such as power cable monitoring, fire detection in tunnels and pipeline monitoring.

Sagnac Effect-based Sensors

The Sagnac²⁸ effect (also called *Sagnac Interference*), is a phenomenon encountered in interferometry that is elicited by rotation. A beam of light is split and the two beams are made to follow a trajectory in opposite directions. To act as a ring the trajectory must enclose an area. On return to the point of entry the light is allowed to exit the apparatus in such a way that an interference pattern is obtained. The position of the interference fringes is dependent on the angular velocity of the setup. This arrangement is also called a Sagnac interferometer. The schematic set-up of a Sagnac interferometer is shown in Fig. 10.32.

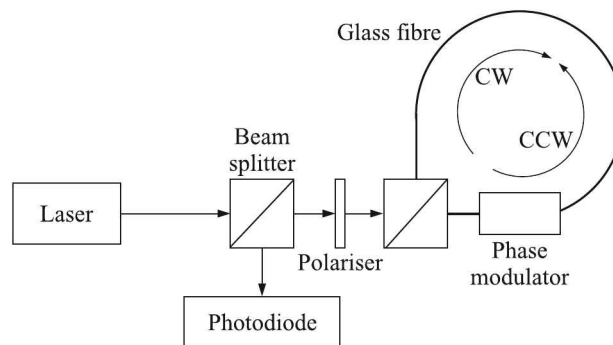


Fig. 10.32 Schematic set-up of a Sagnac interferometer.

Light is decomposed into two beams by a 50:50 beam splitter, with one travelling clockwise (CW) and the other counter-clockwise (CCW) around a polarisation-maintaining single-mode glass-fibre loop. The two beams interfere after passage through the loop, and the interference signal is measured with a photodiode. If only reciprocal effects are involved in the experiment, then the two beams interfere constructively (relative phase shift $\Delta\phi = 0$). If $\Delta\phi \neq 0$, then non-reciprocal effects occur, one of them being the Sagnac effect that results from the rotation of the fibre loop during the measurement.

²⁸Named after Georges Sagnac (1869–1928), a French physicist.

Let r be the radius of the fibre-optic loop
 N be the number of windings
 λ be the wavelength of the incident radiation
 c be the velocity of light, and
 ω be the angular velocity of the rotating loop.

If light is injected into the loop at time $t = 0$, at $t = (2\pi r N/c)$, the CW and CCW beams meet again at the starting point. However, due to the rotation of the loop, they have travelled different path lengths as seen from an inertial frame. The difference in path length can be expressed as (see Fig. 10.33)

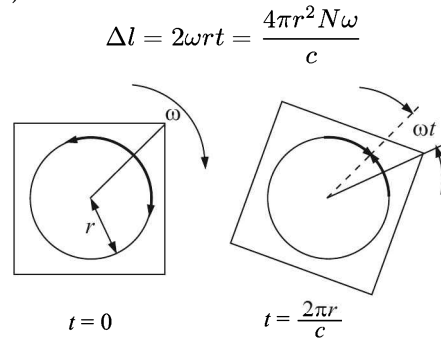


Fig. 10.33 Path difference between the CW and CCW beams.

and the corresponding phase difference between the two beams is

$$\Delta\phi = \frac{8\pi^2 r^2 N \omega}{\lambda c}$$

It is interesting to note that although the above calculation is over-simplified (e.g., the speed of light in vacuum c was assumed instead of that in the glass fibre), it yields the correct result. Exact relativistic calculations show that the phase shift is indeed independent of the material of the wave guide and the above equations apply.

In order to distinguish a CW from a CCW rotation of the loop, and to improve upon the sensitivity of the interferometer, a phase modulator is inserted into the loop providing a phase modulation

$$\phi(t) = \phi_m \sin \omega_m t$$

Since the modulator is located just behind the beam splitter (see Fig. 10.32), the CW beam first runs through the fibre coil before it passes the modulator. Thus, it passes the modulator with a time delay of

$$\Delta t = \frac{2\pi r N n}{c}$$

with respect to the CCW beam, where n is the refractive index of the material of the fibre. The modulation is most effective if Δt is half of the period time of the modulation, i.e. $\Delta t = \pi/\omega$, so that

$$\phi(t) = -\phi(t - \Delta t)$$

Under this condition, the non-reciprocal phase shift between the two beams is

$$\Delta\phi = \Delta\phi_0 + 2\phi_m \sin \omega_m t$$

with $\Delta\phi_0$ the phase shift due to a mechanical rotation of the fibre coil.

In a polarisation maintaining fibre (PMF) Sagnac loop interferometer, two different polarisation modes exhibit different responses to temperature. This property is utilised to construct Sagnac effect based temperature sensors. Of late, it has attracted much attention owing to its advantages such as easy manufacture, great flexibility, and good stability.

Sometimes temperatures need to be measured in places that do not allow conventional sensors to be employed. For example, consider the situation while measuring the temperature of the windings of a high voltage oil cooled power transformer. The voltage may be as high as 500 kV peak. Wired sensors would be hazardous to anyone near the measuring device. However non-contact sensors cannot be used, because the transformer windings are not visible. In situations like these, the fibre-optic sensors become the only option.

But fibre-optic temperature sensors and associated measuring devices are expensive and hence applied only when they have a compelling advantage for specialist applications like the one we just talked about. For this reason, we have not covered them in detail here. However, maybe in future this type of sensors will become significantly cheaper and more widely deployed as and when optical integrated circuits become a reality.

10.6 Quartz Thermometer

The natural frequency of vibration of a piezoelectric crystal like quartz depends on its temperature. This property is utilised to construct quartz thermometers.

The resonant frequency of a quartz oscillator as a function of temperature T is given by

$$f_T = f_0(1 + \alpha T + \beta T^2 + \gamma T^3) \quad (10.20)$$

where f_0 is the frequency at 0°C , and α , β and γ are coefficients. By choosing the cutting plane of the crystal, β and γ coefficients can be made zero, thus making the frequency linearly dependent on the temperature.

The cutaway schematic view of a quartz thermometer is shown in Fig. 10.34. The quartz resonator is shaped like a tuning fork. This shape enables the thermometer to cover a temperature span of 4.2 K to 523 K, with an α -value of -54 ppm/K at 273 K and hysteresis below 0.001 K.

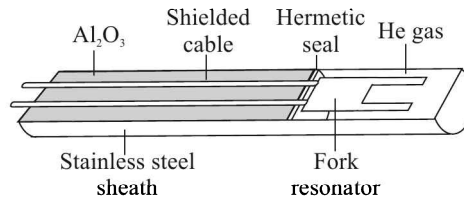


Fig. 10.34 Cutaway schematic view of a quartz thermometer.

The typical size of the vibrating part of the fork resonator is about 2.5 mm (L) \times 0.25 mm (W). The resonant frequency is on the order of 50 kHz. The resonator and its two terminals are hermetically sealed in a stainless steel sheath filled with helium gas at a pressure of about 5 Torr.

The thermometer can be interfaced to an analogue or digital recorder or a computer. Periodical calibration at the ice-point ensures a stable operation. Typical accuracy of measurement is ± 0.075 K in the range of 193 K to 523 K.

Quartz thermometers are typically used in calibration of thermometers, in precision colorimetry and in remote measurements. Their digital operating principle makes them free from problems caused by lead resistance and noise pick-up.

10.7 Change in the Velocity of Sound Propagation

When sound travels through a medium—gas, liquid or solid—its velocity v depends on the temperature of the medium. The relevant relations, as enunciated by Newton more than 4 centuries ago, are:

$$\text{For gases} \quad v = \sqrt{\frac{\gamma RT}{M}} \quad (10.21)$$

$$\text{For liquids} \quad v = \sqrt{\frac{\beta(T)}{\rho(T)}}$$

$$\text{For solids} \quad v = \sqrt{\frac{M}{\rho(T)}}$$

where γ is the the ratio of the specific heats at constant pressure and volume, c_p/c_v

R is the gas constant

T is the absolute temperature

M is the molecular weight

$\beta(T)$ is the temperature dependent bulk modulus of elasticity

$\rho(T)$ is the temperature dependent specific gravity.

Instead of ascertaining precise values of $\beta(T)$ and $\rho(T)$ for liquids and solids, thermometers are better calibrated with some ideal gas for which all the parameters of Eq. (10.21) are known precisely.

This property is utilised to construct two kinds of temperature sensors:

1. SAW thermometers
2. Ultrasonic thermometers

SAW Thermometers

Surface acoustic wave velocities depend on the temperature of material. The propagating medium changes with temperature, affecting the output. The orientation and type of crystalline material used to fabricate the sensor determine its sensitivity.

SAW temperature sensors have a resolution of 0.001°C , good linearity and low hysteresis. Their additional advantage is that they require no power. So they can be wireless, making them well suited for use in remote locations.

However, they must be sealed in a hermetic package, because they are highly sensitive to mass loading.

Ultrasonic Thermometers

Ultrasonic pulses of 0.1 MHz to 3 MHz are generated with the help of piezoelectric or magnetostrictive transmitting transducers. These are transmitted either directly or through sound conductors to the investigated medium (Fig. 10.35). The pulses arrive at the receiving transducer with a certain lag depending upon the medium and its temperature. The time-delay is measured electronically and converted to sound velocity in the medium. Normally single pulses are sent and received. But in case the noise level is high, a series of pulses may be sent and received.

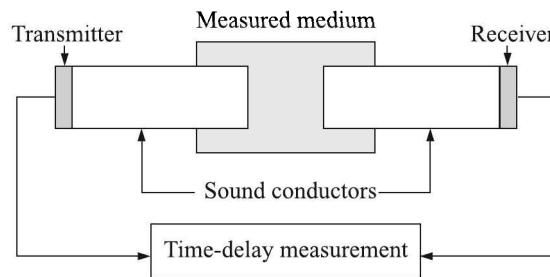


Fig. 10.35 Schematic set-up of ultrasonic temperature measurement.

Temperatures in the range of 300 to 17000 K may be measured with an accuracy of 0.001 K with the help of ultrasonic thermometers.

Because of its high range, this thermometer is suitable for direct measurement of temperatures of ionised gases or plasma which is not feasible by other methods.

Also, since it is not a point measurement, it gives an average value of the temperature over a rather large area of the investigated medium.

Ultrasonic temperature measurement typically finds application in measurement of temperatures in nuclear reactors or liquids in tanks or for direct measurement of temperatures over 2000 K.

10.8 Radiation Pyrometers

So far we have discussed thermometers that depend on the heat conduction process to acquire the temperature of the measurand. But this kind of measurement fails in the following situations:

1. In a contact type temperature measurement, the measuring sensor has to be at the same temperature as the measurand; when the measurable temperature is very high, this may melt or even burn the sensor.
2. The measuring environment is highly corrosive that may destroy the thermometer.
3. It is necessary to measure the temperature of a moving object like a missile.
4. It is necessary to measure the average temperature, rather than the temperature at a point of the object.
5. A fast temperature measurement is necessary; conduction being a slow process, conduction-based temperature measuring devices have to be in contact with the object for quite some time to attain its temperature.

In such situations we have to have recourse to radiation pyrometry²⁹.

Radiation pyrometry is the method of measuring the temperature of a body by measuring the radiation emitted by it. A radiation thermometer can measure the temperature of an object without physical contact and has many advantages over other contact-type measurement devices. Such a measurement does not contaminate, damage, or interfere with the object being monitored. It can be mounted remotely from the hot target enabling it to operate for long periods of time with minimal maintenance.

Theoretically, all bodies above 0 K emit radiation and hence it is possible to measure temperature of a body by radiation pyrometry at all temperatures. But in practice, the method is applied to measure temperatures above 700°C when

1. Direct temperature measurement becomes difficult, and
2. The intensity of the emitted radiation is substantial to allow good measurements

Radiation Fundamentals

Before we discuss methods of radiation pyrometry we make a short detour to discuss radiation fundamentals.

Electromagnetic spectrum

We have said that any object whose temperature is above 0 K is capable of radiating electromagnetic energy which is propagated through space at the speed of light. The energy spectrum contains many forms of electromagnetic emissions, including radio waves, infrared, light, X-rays, etc. (Fig. 10.36).

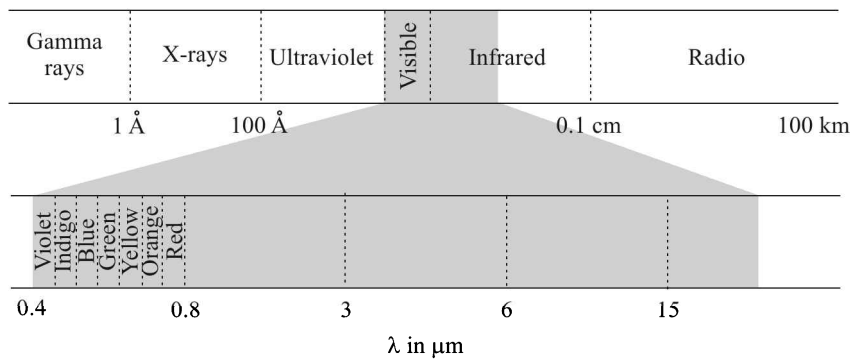


Fig. 10.36 The electromagnetic spectrum (not to scale). The lower part shows the portion of the spectrum that can be used for radiation pyrometry.

Radiation thermometers are designed to respond to wavelengths within the infrared portion of the spectrum. In principle, temperature measurement can be made using thermometers operational over many different ranges of wavelength which generally lie somewhere between 0.2 mm and 20 mm. In practice, infrared radiation thermometers/pyrometers, by specifically measuring the energy radiated from an object in the 0.7 μm to 20 μm wavelength range, are

²⁹ *pyro* is a Greek word meaning *fire*.

a subset of radiation thermometers. The human eye responds to infrared emissions within the visible region of the infrared portion of the spectrum. It is this response which enables the eye to observe the temperature of a metal change its colour from dull red to bright white while being heated. The human eye does not respond to most infrared emissions and therefore we cannot see them. However, they can still be focussed by an optical system on to a detector inside a radiation thermometer in a way similar to visible light.

Absorption, transmission and reflection

When the infrared radiated by an object strikes another body, it will

1. Absorb a portion of the energy
2. Reflect a portion, and
3. If the body is not opaque, will transmit a portion

The sum of the three individual fractions must always add up to the initial value of 1 which left the source. Thus, if a_λ , r_λ and t_λ indicate absorbed, reflected and transmitted fractions respectively then

$$a_\lambda + r_\lambda + t_\lambda = 1 \quad (10.22)$$

Blackbody and blackbody radiation. In Eq. (10.22), if $r_\lambda = t_\lambda = 0$, then $a_\lambda = 1$. That means, if we have a body which is totally non-reflective and completely opaque then the entire radiated infrared energy received by this body will be absorbed. This type of body is a perfect absorber and it can be proved that it will also be a perfect emitter of radiation. A perfect absorber and hence emitter of radiated energy is called a *blackbody*.

The word blackbody is a technical term to describe an object capable of absorbing all radiation falling on it and emitting maximum energy for a given temperature. It would not necessarily appear to be black in colour. In practice surfaces of materials are not perfect absorbers and tend to emit and reflect infrared energy. A non-blackbody would absorb less energy than a blackbody under similar conditions and hence would radiate less infrared energy even though it was at the same temperature. The knowledge of a surface's ability to radiate infrared is important when using an infrared thermometer. We will revert to it after we discuss the laws of radiation.

Laws of radiation

Three laws of radiation are of importance to us in radiation pyrometry. They are

1. Planck's³⁰ law
2. Stefan³¹-Boltzmann³² law
3. Wien's³³ displacement law

³⁰Max Karl Ernst Ludwig Planck (1858–1947) was a German physicist who is regarded as the founder of the quantum theory. He received the Nobel Prize in Physics in 1918.

³¹Joseph Stefan (1835–1893) was a physicist, mathematician, and poet of Austria. He experimentally established the law.

³²Ludwig Eduard Boltzmann (1844–1906) was an Austrian physicist famous for his initiating the fields of statistical mechanics and statistical thermodynamics. He explained the law theoretically.

³³Wilhelm Carl Werner Otto Fritz Franz Wien (1864–1928) was a German physicist who used theories about heat and electromagnetism to deduce Wien's displacement law in 1893.

Planck's law. Max Planck showed³⁴ that for a black radiator the intensity of radiation \mathcal{R}_λ at a particular wavelength λ is given by

$$\mathcal{R}_\lambda = \frac{c_1 \lambda^{-5}}{\exp(c_2/\lambda T) - 1} \quad (10.23)$$

where $c_1 = 3.746 \times 10^{-16} \text{ W}\cdot\text{m}^2$
 $c_2 = 0.0144 \text{ m}\cdot\text{K}$
 λ is the wavelength of radiation, in m
 T is the absolute temperature of the blackbody
 \mathcal{R}_λ is the energy, in $\text{W}\cdot\text{m}^{-3}$.

Equation (10.23) thus gives the distribution of radiant intensity with wavelength, which is depicted in Fig. 10.37.

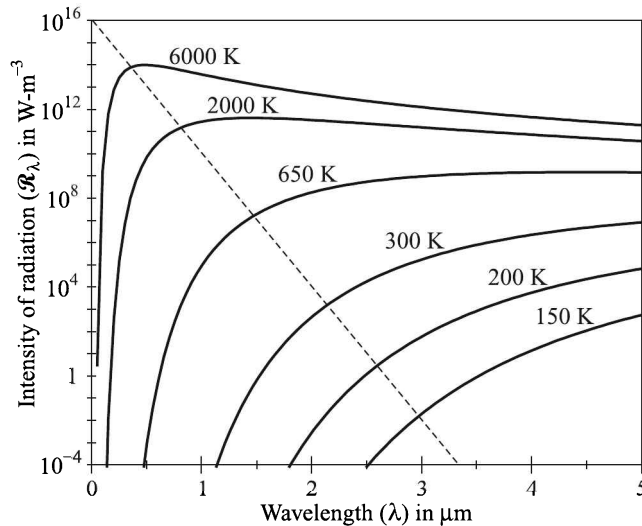


Fig. 10.37 Distribution of radiant intensity at different temperatures.

The following characteristics of the curves may be noted:

1. As the temperature of the radiant object increases, the peak height increases and shifts towards shorter wavelengths.
2. The rate of rise of a curve is faster than that of its decay.
3. The decay of a curve is roughly linear.

Wien's displacement law. Since $\exp(c_2/\lambda T) \gg 1$, Wien wrote

$$\mathcal{R}_\lambda = c_1 \lambda^{-5} \exp\left(-\frac{c_2}{\lambda T}\right) \quad (10.24)$$

By differentiating Eq. (10.24) with respect to λ and equating it to zero, Wien found an expression for the maximum value of λ for a particular value of T . Thus, a relationship between

³⁴See, for example, *Physics* (Part II), D Halliday and R Resnick, Wiley Eastern, New Delhi (1990), p 1173ff.

the peak wavelength and the corresponding temperature can be derived. The relationship is known as *Wien's displacement law*.

The law states that

$$\lambda_{\max}T = \text{constant} = 2.898 \times 10^{-3} \text{ m-K} \quad (10.25)$$

where λ_{\max} is the wavelength at which the maximum energy is emitted by a blackbody at temperature T K.

Equation (10.25) is useful in predicting the wavelength at which peak energy will occur for any given target temperature.

Example 10.11

The expected temperature of a target is 1000°C. At what wavelength should we set our pyrometer to obtain maximum accuracy in the measurement of temperature?

Solution

The maximum accuracy in the temperature measurement can be achieved if we measure at the wavelength at which the target is radiating at its peak energy. From Eq. (10.25), this wavelength is

$$\lambda_{\max} = \frac{2.898 \times 10^{-3}}{273 + 1000} \text{ m} = 2.28 \mu\text{m}$$

Stefan-Boltzmann law. If we calculate the area under the curve corresponding to any temperature, say 650 K, we get the total emitted power at that temperature. The amount of total emitted power of a blackbody at a given temperature is given by the *Stefan-Boltzmann law*.

The law states that the total power \mathcal{R}_T emitted by a blackbody at a temperature T is given by

$$\mathcal{R}_T = \sigma T^4 \quad (10.26)$$

where

$$\sigma = 5.67 \times 10^{-8} \text{ W-m}^{-2}\text{K}^{-4}$$

Emissivity

There is a technical distinction between the terms *emittance* and *emissivity* though both are often used in the same connotation. Whereas emittance refers to the properties of a particular *object*, emissivity refers to the properties of the *material which the object is made of*. In other words, emissivity is only one of the factors that determine emittance. Other factors include shape of the object, oxidation and surface finish.

In reality, objects are not blackbodies, or, perfect emitters of infrared energy. Because, as the energy strikes the surface, about 40% is reflected back, and this internally reflected energy will never leave by radiative means (see Fig. 10.38).

The ability of an object to radiate energy at a wavelength, say infrared, depends upon its emissivity which is really a comparison between the energy emitted by the target object and an ideal emitter or blackbody at the same temperature. Hence emissivity at a particular wavelength ε_λ at a given temperature T may be expressed as follows:

$$\varepsilon_\lambda = \frac{\text{Radiation emitted by the target}}{\text{Radiation emitted by the blackbody}} \Big|_{T,\lambda}$$

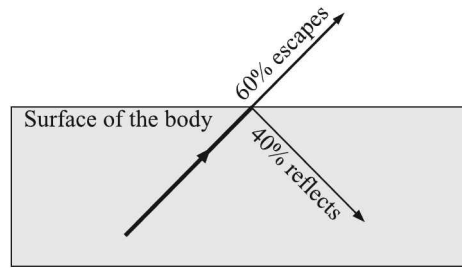


Fig. 10.38 Phenomena that occur when energy strikes a surface.

The apparent emittance of a material depends on

1. Temperature at which it is determined
2. Wavelength at which the measurement is made
3. Surface condition of the emitter, with lower values for polished surfaces and higher values for rough or matte surfaces; however, as materials oxidise, emittance tends to increase and the surface condition dependence decreases

Grey bodies. Emisivity of all materials does not change equally at different wavelengths. Materials for which emissivity remains the same at different wavelengths are called *grey bodies*. Materials for which this is not true are called non-grey bodies. Representative emissivity values for a range of common metals and nonmetals at various temperatures are given in Table 10.8.

Table 10.8 Emissivity values for a few common materials

| <i>Material</i> | <i>Condition</i> | ϵ_λ (at 1 μm) (approx.) |
|-----------------|------------------|--|
| Aluminium | Unoxidised | 0.13 |
| | Oxidised | 0.40 |
| Copper | Unoxidised | 0.06 |
| | Oxidised | 0.80 |
| Iron and Steel | Unoxidised | 0.35 |
| | Oxidised | 0.85 |
| Tin | Unoxidised | 0.3 |
| | Oxidised | 0.6 |
| Asbestos | | 0.90 |
| Asphalt | | 0.85 |
| Brick | | 0.8 |

We have already stated that several factors influence the emissivity of a material. They are as follows:

Wavelength. At longer wavelengths the emissivity of polished metals tends to decrease. However, nonmetallic materials often show an increase in emissivity with increasing wavelength. Semi-transparent materials, such as plastic films, show strong variations with wavelength and require special consideration. The variations of emissivity with wavelength for iron and a grey body are shown in Fig. 10.39.

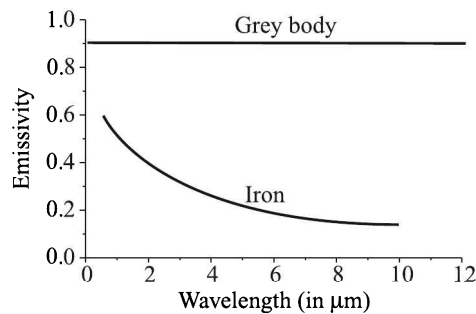


Fig. 10.39 Variation of emissivity of iron and grey body with wavelength.

Surface condition. The nonmetals with rough surfaces, such as brick, tend to have high values of emissivity. Metals with unoxidised surfaces tend to have rather low emissivities. We should keep it mind that for an opaque object

$$\text{Emissivity} + \text{Reflectivity} \approx 1.0$$

This means that a target surface which is rather non-reflective, such as asphalt, would have a high emissivity, and a highly reflective material, such as rolled aluminium, would have a low value of emissivity.

Temperature. If the radiation thermometer operates over a narrow waveband, the emissivity of materials does not tend to change very much with temperature.

Viewing angle. The dependence of emissivity on the angle that the radiation thermometer subtends to the normal of the sample surface is shown in Fig. 10.40.

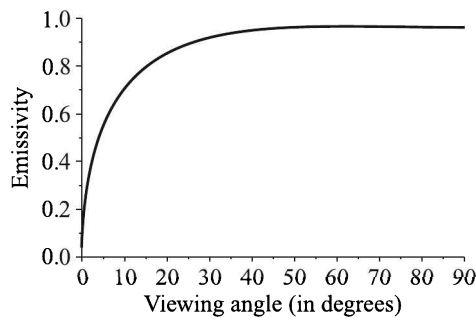


Fig. 10.40 Variation of emissivity with viewing angle.

The emissivity of most materials is not strongly dependent on the viewing angle provided the measurement is made within about 45° of the normal to the sample surface.

The values for the emissivities of almost all substances are known and published in reference literatures. However, the emissivity determined under laboratory conditions seldom agrees with the actual emittance of an object under real operating conditions. For this reason, the published emissivity data are used only as guidance.

As a rule of thumb, most opaque nonmetallic materials have a high and stable emissivity value of 0.85 to 0.90. Most unoxidised metallic materials have a low to medium emissivity value of 0.2 to 0.5. Gold and silver are exceptions, with emissivity values in the 0.02 to 0.04 range. The temperature of these metals is very difficult to measure with a radiation thermometer.

Classification of Radiation Pyrometers

The temperature measurement by radiation thermometers can be classified into four categories as follows:

1. Broadband pyrometers
2. Narrow-band pyrometers
3. Ratio pyrometers
4. Fibre-optic pyrometer

Broadband pyrometers

Broadband pyrometers have usually been the simplest infrared thermometers, with spectral responses from $0.3\ \mu\text{m}$ to 2.5 or even $20\ \mu\text{m}$ as determined by the lens or window material. They may be considered as total radiation thermometers, because in the temperature ranges of normal use they measure a significant fraction of all the thermal radiation emitted by the object whose temperature is being measured.

Fery's pyrometer. Figure 10.41 shows the essential elements of Fery's broadband pyrometer which, incidentally, was the first such instrument ever designed.

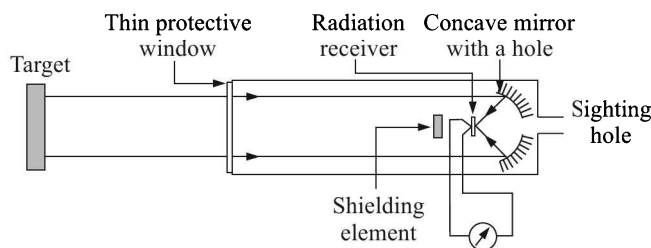


Fig. 10.41 Fery's broadband pyrometer.

Radiation from the target falls on the concave mirror, which can be moved back and forth (arrangement not shown here) to focus radiation on the surface of the radiation receiver (normally platinum black because of its absorptivity of around 0.98). The focussing could be done by a convex lens system though with the possibility of selective absorption by the lens system, while a concave mirror is free from such sources of error. The hot junction of a thermocouple is attached to the radiation receiver. A shielding element protects the thermocouple junction from receiving direct radiation. The developed emf is read on a millivoltmeter which may be calibrated to a temperature scale. Because of the fourth-power law, such calibrations are evidently nonlinear which renders the device unsuitable for measurement below 650°C because of poor sensitivity.

Precautions. The following precautions are necessary while measuring temperatures by means of a broadband pyrometer.

1. The measurement of temperature by a broadband pyrometer is dependent on the total emittance of the surface being measured. Figure 10.42 shows the error in reading for various emissivities and temperatures when a broadband device is calibrated for a blackbody. So long as the emittance does not change, an emissivity control allows the user to compensate for this error.
2. The path to the target must be unobstructed. Water vapour, dust, smoke and radiation absorptive gases present in the atmosphere can attenuate emitted radiation from the target and cause the thermometer to read low.
3. The optical system must be kept clean, and the sighting window protected against any corrosives in the environment.

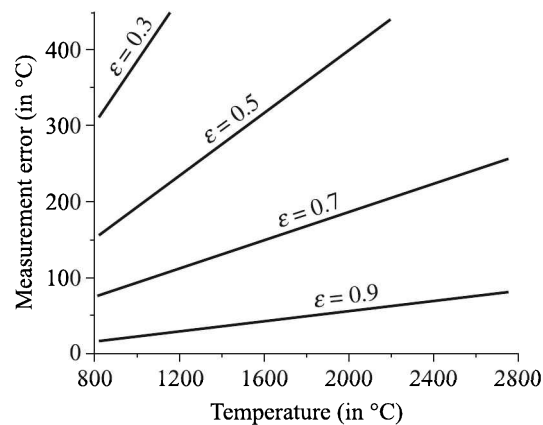


Fig. 10.42 Error in temperature measurement by a broadband IR thermometer caused by non-blackbody emissivity.

Standard ranges of commercially available instruments include 0 to 1000°C, and 500 to 900°C although theoretically there is no upper limit to the temperatures that can be measured in this way. Typical accuracy is 0.5 to 1% full scale.

Advantages and disadvantages. These are given in following table:

| <i>Advantages</i> | <i>Disadvantages</i> |
|-------------------|--|
| 1. Wide range | 1. Low sensitivity |
| 2. Inexpensive | 2. Absorption in the sight path may cause errors |
| 3. Simple | |

Narrow-band pyrometers

Narrow-band pyrometers are not much different in construction from the simple, broadband thermometers except that the lens, window, or filter characteristics are selected to view only a selected portion of the spectrum. A $5 \pm 0.2 \mu\text{m}$ narrow-band pyrometer is typically used to measure the surface temperature of glass which emits strongly in this region, but poorly below

or immediately above this band. However, an 8–14 μm band-pass is preferred to measure the surface temperature of glass if its surroundings are cooler than itself. For low temperature, general-purpose use, the same 8–14 μm narrow-band pyrometer is also useful because the IR radiation in this band is not absorbed by the atmospheric moisture.

Disappearing-filament pyrometer. The classical form of this type is the disappearing-filament optical pyrometer which utilises the photometric principle of comparison of the intensity of incoming radiation at a particular wave band to that of a lamp. Though it is more accurate than Fery's pyrometer, it cannot be used below 650°C since it requires a visual brightness match by a human operator.

Here an image of the target (Fig. 10.43) is superimposed on a heated tungsten filament. The brightness of the tungsten lamp, which is made very stable, has been calibrated previously so that when the current through the lamp is known, the temperature corresponding to the generated brightness of the lamp is known.

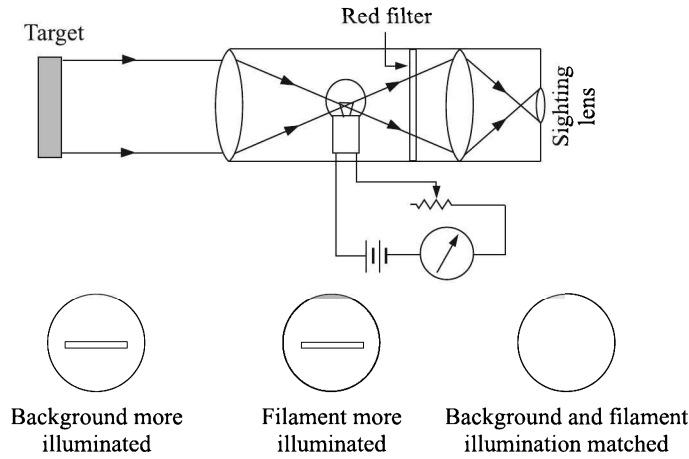


Fig. 10.43 Disappearing-filament optical pyrometer.

A red filter allows only a narrow band of wavelengths around $0.65\mu\text{m}$ to the eye of the observer who controls the lamp current until the filament disappears in the superimposed target image. Then the brightness of the target and lamp are equal and we can write from Eq. (10.23),

$$\frac{\varepsilon_\lambda c_1 \lambda^{-5}}{\exp(c_2/\lambda T_t) - 1} = \frac{c_1 \lambda^{-5}}{\exp(c_2/\lambda T_l) - 1} \quad (10.27)$$

where subscripts t and l correspond to target and lamp respectively and ε_λ is the emissivity of the target at wavelength λ (here $0.65\mu\text{m}$). For $T < 4000^\circ\text{C}$, the exponential terms are much greater than 1. Hence neglecting 1 from the denominators of both sides, we get from Eq. (10.27) after doing some algebra

$$\varepsilon_\lambda = \exp \left\{ -\frac{c_2}{\lambda} \left(\frac{1}{T_l} - \frac{1}{T_t} \right) \right\}$$

and finally,

$$\frac{1}{T_t} - \frac{1}{T_l} = \frac{\lambda \ln \varepsilon_\lambda}{c_2} \quad (10.28)$$

If the target is a blackbody, $\varepsilon_\lambda = 1$ and hence $T_t = T_l$. Otherwise, the temperature can either be calculated from the Eq. (10.28) by knowing ε_λ , or from a calibration curve.

Figure 10.44 shows the schematic representation of an automatic disappearing filament type narrow-band infrared pyrometer.

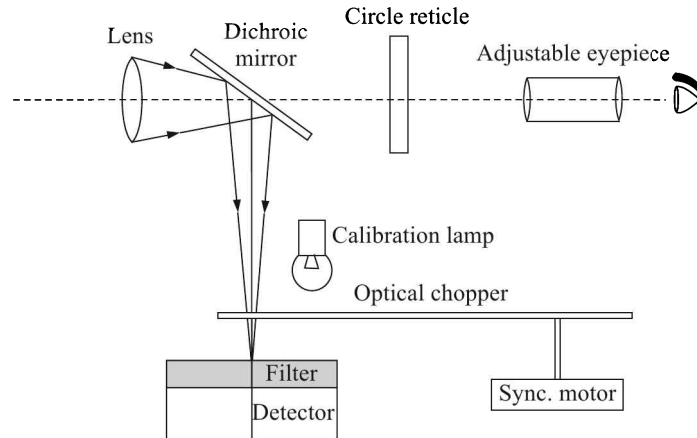


Fig. 10.44 Automatic disappearing-filament type infrared pyrometer.

Here, radiant energy passes through the lens into the dichroic³⁵ mirror, which reflects infrared radiation to the detector, but allows visible portion of the emitted light to pass to an observer through an adjustable eyepiece. A chopper, driven by a motor, is used to alternately expose the detector to incoming radiation and reference radiation. In some models, the human eye is used to adjust the focus.

The device compares the amount of radiation emitted by the target with that emitted by the controlled calibration lamp. The instrument output is proportional to the difference in radiation between the target and the reference. The instrument may have a wide or narrow field of view. All the components can be packaged into a gun-shaped, hand-held instrument (Fig. 10.45). Pressing the trigger energises the reference standard and read-out indicator.

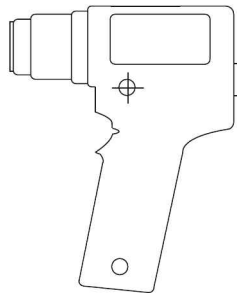


Fig. 10.45 Profile of a hand-held infrared pyrometer.

³⁵Means two-coloured (from the Greek *dikhroos*). It refers to any optical device which can split a beam of light into two beams that differ in wavelengths. Such devices include mirrors and filters. They are usually given optical coatings, which reflect light over a certain range of wavelengths, and transmit light which is outside that range.

Narrow-band pyrometers have typical accuracy in the 1% to 2% of full scale range. The spectral response of narrow-band thermometers is determined by the optical filter used. They are used as general purpose instruments over the temperature range of interest. For example, a thermometer using a silicon cell detector has a response that peaks at 0.9 μm . The upper limit of usefulness is about 1.1 μm . Such a thermometer can only be used at temperatures above 1000°C. Thus, the temperature range of a narrow-band thermometer not only depends on its optical materials but also on the sensitivity of the detectors used.

Advantages and disadvantages. These are given in the following table:

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|--|
| 1. High accuracy even when the emissivity is unknown or varying provided a short wavelength unit is chosen | 1. Narrow temperature spans |
| 2. No interference from ambient open air radiation which is seldom IR | 2. Difficult to select an optimum unit that meets all criteria |
| 3. Specific wave-bands for measuring temperatures of specific materials | |

Example 10.12

A total radiation pyrometer, calibrated with respect to blackbody conditions, is used to measure temperature of a surface with emissivity of 0.8. It indicates a surface temperature of 1050 degree celsius.

- (a) Determine the true temperature of the surface.
- (b) If an optical pyrometer calibrated with respect to blackbody conditions were used to measure the surface temperature of the abovementioned surface, what would have been the indicated value? Assume, emissivity of 0.8 at wavelength $\lambda = 0.65 \mu\text{m}$, constant c_2 in Planck's law = 14338 $\mu\text{m}\cdot\text{K}$.

Solution

(a) Given: $\varepsilon = 0.8$, measured temperature $T_m = 1050^\circ\text{C}$. Since it is a total radiation (or broadband) pyrometer, we have

$$\mathcal{R}_T|_{\text{true}} = \varepsilon\sigma T_{\text{true}}^4$$

whereas, owing to the blackbody assumption, the temperature has been figured out using the relation

$$\mathcal{R}_T|_{\text{true}} = \sigma T_m^4$$

Thus,

$$T_m = \left(\frac{\mathcal{R}_T|_{\text{true}}}{\sigma} \right)^{1/4} = \left(\frac{\varepsilon\sigma}{\sigma} \right)^{1/4} T_{\text{true}}$$

or

$$T_{\text{true}} = \frac{T_m}{(\varepsilon)^{0.25}} = \frac{1050 + 273}{(0.8)^{0.25}} = 1399 \text{ K} = 1126^\circ\text{C}$$

(b) For an optical (or narrow-band) pyrometer, we get from Eq. (10.28)

$$\begin{aligned}\frac{1}{T_t} - \frac{1}{T_m} &= \frac{(0.65 \times 10^{-6}) \ln(0.8)}{14338 \times 10^{-6}} \\ &= -1.0116 \times 10^{-5} \text{ K}^{-1}\end{aligned}$$

Thus

$$\begin{aligned}T_m &= \left(\frac{1}{T_t} + 1.0116 \times 10^{-5} \right)^{-1} \\ &= \left(\frac{1}{1399} + 1.0116 \times 10^{-5} \right)^{-1} \\ &= 1379 \text{ K} = 1106^\circ\text{C}\end{aligned}$$

Ratio pyrometers

Also called *two-colour pyrometers*, these devices measure the radiated energy of an object between two narrow wavelength bands, and calculate the ratio of the two energies. The schematic diagram of a ratio pyrometer is shown in Fig. 10.46.

The two wavelengths are selected by two filters mounted on a rotating wheel. The detector measures the intensities of radiations corresponding to the two wavelengths and the following electronics (not shown) determines the ratio of the intensities. This ratio is a function of the temperature of the object.

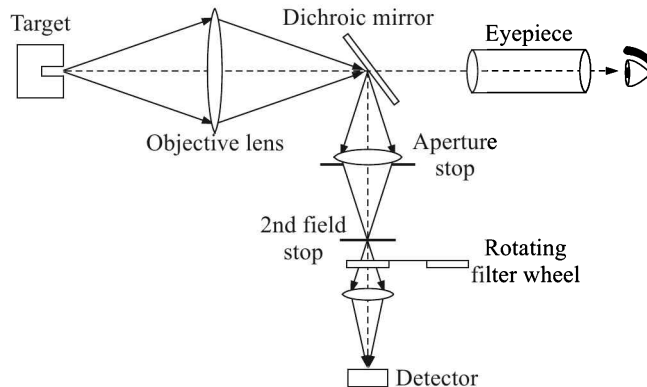


Fig. 10.46 Schematic diagram of a ratio pyrometer.

Because the temperature measurement is dependent only on the ratio of the two energies measured, and not their absolute values, any *static* parameter, such as the target size, which affects the amount of energy in each band by an equal percentage, has no effect on the temperature indication.

The ratio technique also eliminates, or reduces, errors in temperature measurement caused by *dynamic changes* in emissivity, surface finish, and presence of energy absorbing materials, such as water vapour or CO_2 , between the thermometer and the target, because these changes are seen identically by the detector at the two wavelengths being used.

For these reasons, a ratio thermometer is inherently more accurate.

Advantage and disadvantages. The following table lists them:

| <i>Advantage</i> | <i>Disadvantages</i> |
|---|--|
| 1. Rather insensitive to varying target size or intermittent blockage of sight path by smoke, particles, etc. | 1. More expensive 2. Sensitive to changes in ratio of the emissivities 3. It is necessary to know the ratio of the emissivities in the two wave bands of measurement |

Fibre-optic pyrometers

A section of these devices merely use an optical fibre to direct the emitted radiation to the detector. Strictly speaking, this section of fibre-optic devices is not a class by itself. The first such sensors used a sapphire rod of 3 mm diameter to pick up the energy from the target and transmit it to a detector. Contemporary fibre-optic pyrometers use a flexible bundle of glass fibres with or without a lens. The spectral response of these fibres extends to about $2\ \mu\text{m}$, though some materials such as fluorides have a wider band-pass. Some are useful at target temperatures as low as 100°C .

Beyond collection of radiant energy, fibre-optic glasses can be doped to serve directly as radiation emitters at hot spots so that the fiber-optics serve as both the sensor and the media. A sapphire probe is commercially available that has the sensing end coated by a refractory metal forming a blackbody cavity. The thin, sapphire rod thermally insulates and connects to an optical fibre as is shown in Fig. 10.47. A ratio method may be utilised to determine the temperature.

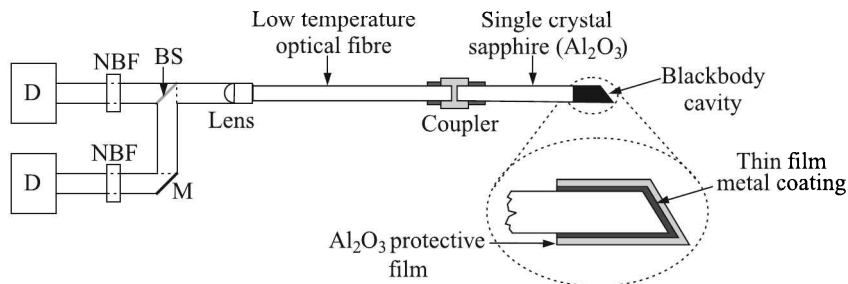


Fig. 10.47 Schematic diagram of a fibre-optic ratio pyrometer. NBF: narrow-band filter, D: detector, M: mirror, BS: beam splitter.

Fibre-optic pyrometers are especially useful where it is difficult, dangerous or impossible to obtain or maintain a clear sighting path to the target, as in pressure or vacuum chambers. Fibre-optic pyrometers have also been used to measure temperatures of turbine blades in gas turbines, and the temperature of small objects in induction heating coils.

Next we pass on to an important aspect of temperature measurement which is considered necessary for all kinds of thermal probes. Of course, remote measurements by pyrometers, having no probes, do not come under its purview.

10.9 Thermowells

Thermowells or protection tubes are used in industrial temperature measurement to provide isolation to a temperature probe from the environment, either liquid, gas or slurry whose temperature is to be measured. Apart from protecting the probe from corrosive process materials, they also facilitate removing, changing, checking or replacing probes without compromising either the ambient region or the process.

Design Factors

In designing a system using thermowells, the factors to be considered are:

1. Material of construction
2. Insertion length
3. Flow velocity of the process fluid
4. Well shank, and
5. Bore size

In the following paragraphs, we discuss them in a little more detail.

Material of construction

Thermowell material must endure the corrosive conditions in the well environment and provide mechanical strength needed to withstand operating pressure and process flow. In addition to selecting the proper base material, coatings may be used to improve a thermowell's resistance to abrasion or the chemical process. A few materials are listed in Table 10.9 to give an idea about the selection of material for a thermowell.

Table 10.9 Choice of materials for thermowells

| <i>Material</i> | <i>Operating temperature</i> °C (max.) | <i>Suitability</i> |
|-----------------|---|--|
| Carbon steel | 700.0 | May be used in oxidising environment. |
| 310 SS | 1150.0 | Resistant to carburising and reducing environments. Carbide precipitates in the 900°C to 1150°C range. |
| Alloy 800 | 1150.0 | Sulphur and corrosion resistant. |
| Hastelloy C | 1200.0 | Oxidation resistant to 950°C. Resistant to corrosion by ferric and cupric chlorides, contaminated mineral acids, wet chlorine gas. |
| Nickel | 750.0 | Should be used in sulphur-free oxidising environment. |
| Tantalum | 2750.0 | Mostly used as a sheath material for stainless flanged wells. Resistant to corrosion from most chemicals. High thermal conductivity. |

The thermowell wall should have an optimum thickness. It should not be too thick to generate errors caused by thermal conduction and slow sensor response. Also, it should not

be too thin to collapse under process pressure, to be eroded by abrasive media or to bend by the process flow. Spring-loaded mounting styles ensure positive contact to maximise thermal transfer and minimise sensor vibration within a thermowell.

Insertion length

The insertion length or *U-length* in common jargon, is measured from the bottom of the threads or flange to the tip of the well. The accuracy of the sensor may be affected by the U-length of the well. Thermocouples, which are tip sensitive, are less likely to be affected by short U-lengths, while stem-sensitive RTDs would require longer U-lengths for the same process condition. A rule of thumb is to immerse a thermocouple at least 7.5 cm in gases and 2.5 cm in liquids. In the case of RTDs, the U-length should be 10 times the sensor diameter plus 2.5 cm.

Flow velocity of the process fluid

As the fluid flows past the well, a turbulent wake forms around which causes the well to vibrate. Known as the *von Kármán trail*³⁶, it is the most common cause of well failure. This vibration frequency is a function of the diameter of the well and the velocity of the fluid. The well must have sufficient stiffness to ensure that the wake frequency never equals the natural frequency of the well. In case of a resonance, the amplitude of vibration may be high enough to destroy the well.

Well shank

Tapered wells provide a greater protection against breaking in high velocity fluid applications. The tapered well has a higher strength to weight ratio as well as a higher natural frequency. Moreover, reduced tip or step down wells provide increased sensitivity.

The common connection types include threaded, flanged and socket weld types with standard bore sizes.

Bore size

Selection of a standard bore size throughout the plant permits the use of several types of temperature measuring instruments in the same wells. Most applications use 6.6 mm or 9.8 mm diameter bores. These bore sizes accommodate most commercially available sheathed thermocouples, RTDs and thermometers.

Disadvantages

While thermowells offer many advantages as mentioned above, their disadvantages include

1. Response of the sensor to temperature changes becomes slower
2. The measurement error increases due mostly to the heat loss through the stem down the length of the thermowell
3. Extra costs for purchase and installation

³⁶See Section 11.3 at page 467.

Review Questions

- 10.1 Elucidate the laws of thermocouple behaviour.
- 10.2 (a) In what way does specifying a temperature scale differ from specifying, say, a scale for measuring length?
- (b) As shown in Fig. 10.48, two thermocouples are used to measure a temperature difference (A, B and C are different metals). Will both the arrangements yield the same emf? Justify your answer.
- (c) State the reason for using a potentiometric arrangement for measuring a static thermo-emf.
- (d) Show schematically an arrangement for measuring a dynamic thermo-emf.
- (e) State two advantages of using a thermocouple for measurement of temperature.

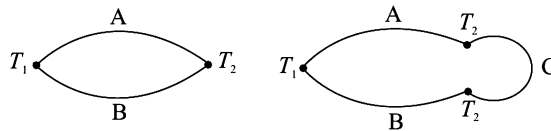
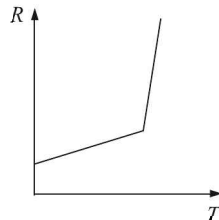


Fig. 10.48

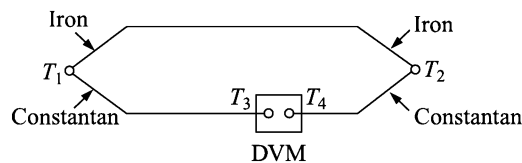
- 10.3 What are thermistors? How are they constructed? Discuss their resistance-temperature characteristics.
- The typical value of β for a thermistor around 25°C is 4000 K. Find the corresponding resistance-temperature coefficient (rate of change of resistance with temperature per unit resistance) for a thermistor.
- 10.4 Suggest an instrumentation circuit which will produce a linear output with temperature.
- 10.5 Choose the correct answer out of the suggested answers:
- (a) Thermocouple measurement error arises from
- poor junction connections
 - de-calibration of thermocouple wire
 - thermal gradients
 - all of the above
- (b) A thermostatic cut-out works on the principle of
- thermal expansion of fluids
 - variation of resistance with temperature
 - expansion caused by air pressure
 - thermal expansion of metals
- (c) A platinum resistance thermometer has a specific resistance of $0.9 \times 10^{-5} \Omega\text{-cm}$ at -50°C . The value of the specific resistance at 50°C would be near to
- $0.9 \times 10^{-6} \Omega\text{-cm}$
 - $1.05 \times 10^{-5} \Omega\text{-cm}$
 - $0.8 \times 10^{-5} \Omega\text{-cm}$
 - $0.9 \times 10^{-5} \Omega\text{-cm}$

- (d) A thermocouple arrangement is to be used to measure temperature in the range of 700–800°C. Point out the pair that would be the most suitable for this application
- (i) Copper-constantan
 - (ii) Iron-constantan
 - (iii) Chromel-alumel
 - (iv) Platinum-platinum-rhodium
- (e) In a thermocouple element heat energy transferred to the hot junction is converted to electrical energy by
- (i) Johnson's effect
 - (ii) Seebeck effect
 - (iii) Hall effect
 - (iv) Faraday effect
- (f) Which of the following should be incorporated in RTD to make a temperature sensing bridge most sensitive to temperature?
- (i) Platinum
 - (ii) Nickel
 - (iii) Thermistor
 - (iv) Copper
- (g) A thermocouple with its reference junction exposed to room temperature of 20°C gives an open circuit voltage of 5 mV. If the thermocouple has temperature sensitivity of 50 $\mu\text{V}/^\circ\text{C}$, the measured temperature is
- (i) 100°C
 - (ii) 12°C
 - (iii) 8°C
 - (iv) 2°C
- (h) The range in which the mercury-in-glass thermometer can be used is
- (i) 0° to 100°C
 - (ii) –20° to 340°C
 - (iii) –50° to 560°C
 - (iv) –100° to 500°C
- (i) PTC thermistor shows
- (i) positive resistance characteristics
 - (ii) negative temperature characteristics
 - (iii) positive temperature characteristics
 - (iv) negative resistance characteristics

- (j) T-type thermocouple is made of
- (i) chromel-alumel
 - (ii) copper-constantan
 - (iii) iron-constantan
 - (iv) none of these
- (k) Non-contact type temperature sensor is
- (i) thermocouple
 - (ii) radiation pyrometer
 - (iii) thermistor
 - (iv) none of these
- (l) The emf developed by a thermocouple depends on
- (i) the length of wires and temperature difference between the hot and cold junctions
 - (ii) materials used, diameter of wires used and the temperature difference between the hot and cold junctions
 - (iii) materials used, temperature of hot junction and temperature of cold junction
 - (iv) materials used, shape and size of materials, resistance of the wires and temperature difference between the hot and cold junctions
- (m) Choose the thermocouple that can measure a temperature in the range of 1300°C to 1400°C
- (i) iron-constantan
 - (ii) copper-constantan
 - (iii) platinum-platinum rhodium
 - (iv) chromel-alumel
- (n) The temperature of fixed points used to define International Temperature Scale are determined by using
- (i) platinum resistance thermometer
 - (ii) platinum-platinum rhodium thermocouple
 - (iii) vapour pressure thermometer
 - (iv) gas thermometer to which corrections are applied for non-ideal behaviour of the gas
- (o) The resistance temperature characteristics of a temperature transducer is shown in Fig. 10.49.

**Fig. 10.49**

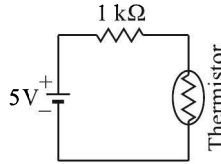
- The transducer is
- (i) nickel RTD
 - (ii) platinum RTD
 - (iii) NTC thermistor
 - (iv) PTC thermistor
- (p) The International Temperature Scale in the range 0–630°C is defined by means of a
- (i) mercury pressure spring thermometer
 - (ii) platinum-platinum 10% rhodium thermocouple
 - (iii) platinum resistance thermometer
 - (iv) total radiation pyrometer
- (q) The sensing element of a thermocouple at its hot junction is provided with a shield while taking measurements in a high temperature gas. The principal reason for providing the shield is
- (i) to reduce conduction and convection errors
 - (ii) to reduce radiation error
 - (iii) to provide temperature compensation to the Seebeck voltage
 - (iv) to improve air supply to the sensing element for better response
- (r) A temperature measuring system consists of a sensor and a thermal well. Each of them can be considered as a single capacity system with the capacity of the thermal well higher than that of the sensor. The overall measuring system will
- (i) have equal time constants
 - (ii) have zero damping
 - (iii) be a non-interactive system
 - (iv) be an interactive system
- (s) A type J (iron-constantan) thermocouple has a voltage sensitivity of $55 \mu\text{V}/^\circ\text{C}$. A digital voltmeter (DVM) is used to measure the voltage under the condition shown in the following figure.



Given that $T_1 = 300^\circ\text{C}$, $T_2 = 100^\circ\text{C}$, and $T_3 = T_4 = 20^\circ\text{C}$, the meter will indicate a voltage of

- (i) 11.0 mV
- (ii) 15.4 mV
- (iii) 16.5 mV
- (iv) 17.6 mV

- (t) A thermistor has a resistance of $10\text{ k}\Omega$ at $25\text{ }^\circ\text{C}$ and $1\text{ k}\Omega$ at $100\text{ }^\circ\text{C}$. The range of operation is $0\text{ }^\circ\text{C}$ to $150\text{ }^\circ\text{C}$. The excitation voltage is 5 V and a series resistor of $1\text{ k}\Omega$ is connected to the thermistor.



The power dissipated in the thermistor at $150\text{ }^\circ\text{C}$ is

- (i) 4.0 mW
 - (ii) 4.7 mW
 - (iii) 5.4 mW
 - (iv) 6.1 mW
- (u) The temperature being sensed by a negative temperature coefficient (NTC) type thermistor is linearly increasing. Its resistance will
- (i) linearly increase with temperature
 - (ii) exponentially increase with temperature
 - (iii) linearly decrease with temperature
 - (iv) exponentially decrease with temperature
- 10.6 Explain giving example as to why the input effective resistance of the measuring instrument should be very high like that of a potentiometer or a voltage follower if we want to measure temperature by a thermocouple accurately.
- Neatly draw and clearly explain the operation of a practical self-balancing type thermocouple-actuated potentiometer circuit for measurement of temperature. The measuring circuit should be capable of automatically compensating the error caused by variation of room temperature at which the reference junction of the thermocouple is kept.
- 10.7 Compare and contrast a thermocouple with a thermistor as a temperature transducer.
- 10.8 (a) Why is platinum normally used in the construction of precision standard thermometers for calibration work? State its measuring temperature range and the reason for selecting such a range. Describe with a neat diagram the construction details of a platinum resistance thermometer.
- (b) Briefly explain with diagrams the bridge circuits using RTDs with 3 or 4 lead wires commonly used for measurements of temperatures employing the deflection or the null-mode of operation. What do you mean by sensitivity and sensing error in respect of such a temperature sensing bridge? How does the sensing error prohibit the use of higher excitation voltage to obtain higher accuracy?
- 10.9 (a) What is the resistance of a Pt-100 type RTD at $0\text{ }^\circ\text{C}$?
- (b) Draw the schematic diagram for connection of a 3-wire and a 4-wire RTD.

- (c) Explain the advantage of a 3-wire or 4-wire RTD over a 2-wire RTD
- (d) Why is cold junction compensation required for a thermocouple? Explain with suitable diagram one method of implementation of the same.
- 10.10 (a) Suggest five numbers of commonly used metals and alloys for making resistance coils for the measurement of temperature.
- (b) What are 'thermistors'? Show what the resistance-temperature characteristic of a thermistor looks like and comment on its suitability for temperature measurement.
- 10.11 What do you understand by the self-heating error of a resistance temperature detector? How is this error eliminated?
- 10.12 (a) Explain the working principle of a filled system thermometer.
- (b) What are the various types of filled system thermometers? Explain in detail the working of any of them.
- (c) What are the possible sources of errors in filled system thermometer? How are they minimised?
- (d) What are the advantages and disadvantages of filled system thermometer?
- 10.13 (a) What do you mean by the K-type thermocouple? What is the measuring range of K-type thermocouple?
- (b) What is the advantage of the law of intermediate junction? Explain the working principle of the bead-type thermistor with a neat sketch.
- (c) Describe the operating principle of radiation pyrometer for measurement of temperature and mention the range of this meter.
- 10.14 A thermistor of resistance $1\text{ k}\Omega$ with a temperature coefficient of resistance $4.5\%/^{\circ}\text{C}$ and having an internal temperature rise of $0.2^{\circ}\text{C}/\text{mW}$ around 27°C is included in a dc bridge with three resistors each of value $1\text{ k}\Omega$.
- (a) Calculate the maximum voltage that can be applied to the bridge if the internal temperature rise is not to exceed 0.1°C .
- (b) Calculate the open circuit bridge voltage for a change of 1°C around 27°C .
- 10.15 Figure 10.50 shows a bridge for temperature measurement using two platinum resistances R_A and R_B . R_A is maintained at the reference temperature T_1 while R_B senses an

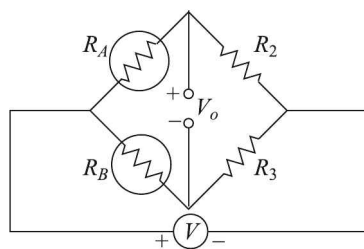


Fig. 10.50

unknown temperature T_2 . Both resistances are equal to R_0 at T_1 and $R_2 = R_3$.

- (a) Find an expression for the bridge output voltage V_o in terms of the bridge ratio R_2/R_0 , temperature differences $(T_2 - T_1)$ and temperature coefficient α .
- (b) Given that $R_2/R_0 = 10$, $T_1 = 100^\circ\text{C}$ and $\alpha = 4 \times 10^{-3}/^\circ\text{C}$, show that an almost linear relationship exists between V_o and $(T_2 - T_1)$ varying from 50°C to 150°C .

10.16 Figure 10.51 shows a bridge circuit for a 2-wire platinum RTD with resistance $R_0 = 10\ \Omega$ at 0°C . It is used to measure temperature between 0 and 100°C . The other bridge arms are resistances of fixed value $R = 10\ \Omega$. To avoid error due to self-heating, the power dissipation through the RTD should be less than $1\ \text{mW}$.

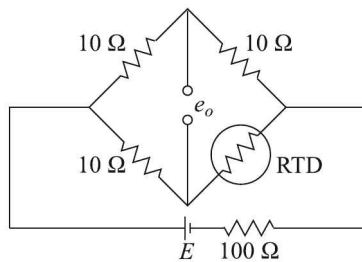


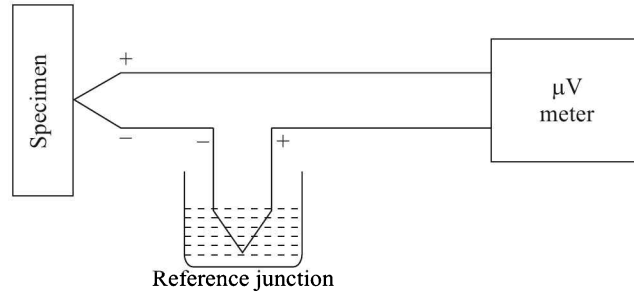
Fig. 10.51

- (a) Calculate the maximum supply voltage E that can be applied in the circuit.
 - (b) Redraw the circuit if the RTD is a three-wire type. Point out its advantage.
- 10.17 An RTD has a resistance of $500\ \Omega$ at 20°C and a temperature coefficient of 0.005°C at 0°C . The RTD is used in a Wheatstone bridge circuit with $R_1 = R_2 = 500\ \Omega$. The variable resistor R_3 nulls the bridge. If the bridge supply is $10\ \text{V}$ and the RTD is in a bath of 0°C , find the value of R_3 to null the bridge when no self-heating of the RTD is considered.
- 10.18 A thermopile having a resistance of $100\ \Omega$ and consisting of 20 copper-constantan thermocouples is used to measure the temperature difference between two points. The temperature of the first point measured separately is 25°C . The emf generated by a voltage measuring device having an internal resistance of $1000\ \Omega$ is $1.47\ \text{mV}$. The emf-temperature relationship for the copper-constantan thermocouple with the reference junction at 0°C is as follows.

Determine the temperature difference between the two points after applying correction for loading effect in the measurement of the emf.

| Temperature ($^\circ\text{C}$) | emf (mV) | Temperature ($^\circ\text{C}$) | emf (mV) |
|----------------------------------|----------|----------------------------------|----------|
| 0 | 0 | 33 | 1.318 |
| 10 | 0.389 | 35 | 1.401 |
| 20 | 0.787 | 37 | 1.485 |
| 25 | 0.990 | 39 | 1.568 |
| 27 | 1.071 | 41 | 1.652 |
| 29 | 1.153 | 43 | 1.737 |
| 31 | 1.235 | 45 | 1.821 |

10.19 A pair of identical thermocouples is employed for measuring the temperature of a specimen as shown below. The characteristic of the thermocouples is tabulated below. The reference junction is at 2°C . The meter reads $48\ \mu\text{V}$.



| <i>Temperature ($^\circ\text{C}$)</i> | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
|--|----|----|----|----|----|----|----|-----|-----|-----|
| <i>Output (μV)</i> | 35 | 45 | 55 | 65 | 75 | 85 | 95 | 105 | 115 | 125 |

The correct temperature of the specimen is

- (a) 13°C (b) 46°C
 (c) 48°C (d) 50°C

10.20 Match the type of thermocouple with its characteristic

- (a) Type E Chromel-constantan (e) Low cost
 (b) Type J Fe-constantan (f) Good stability
 (c) Type T Cu-constantan (g) High precision
 (d) Type S Pt-PtRh (h) High output

Flow Measurement

The flow measuring devices can be broadly divided into two categories, namely

1. Quantity meters, and
2. Rate-of-flow meters

The distinction is based on the sensing device that interacts with the fluid flow. A turbine-type gas meter, for example, indicates the quantity of gas which has flowed past the meter at different times. The rate-of-flow meter on the other hand is concerned with a continuous stream of fluid flow. Of course, a rate-of-flow meter, when integrated over time, works as a quantity meter.

The flow of fluid affects certain physical properties which are sensed by the transducers and accordingly, flow meters can be classified into a few types as shown in Table 11.1.

Table 11.1 Different categories of flowmeters

| <i>Type</i> | <i>Flowmeter</i> |
|----------------------------|--|
| Head type | Orifice plate Venturi tube Flow nozzle Dall flow tube Pitot tube Rotameter (variable area) |
| Velocity measurement type | Electromagnetic Turbine Ultrasonic: Doppler frequency shift Ultrasonic: Transit time Vortex shedding Hot-wire: Constant-current Hot-wire: Constant-temperature |
| Mass-flow measurement type | Coriolis Thermal Impact |
| Positive displacement type | Nutating disc Sliding vane Lobed impeller |
| Open channel type | Weir Flume |

Of course, this list is just indicative and not exhaustive. There are many more flowmeters belonging to each category.

But before we study different types of flowmeters, let us be familiar with a very important number, called *Reynolds number* Re , which categorises different types of flow of fluids.

11.1 Reynolds Number and Flow Patterns

The Reynolds number¹ is defined as the ratio of the fluid's inertial forces to its drag forces. The flow rate, the pipe diameter and the density of the fluid are inertial forces while the viscosity of the fluid is the drag force. Thus,

$$Re = \frac{vd\rho}{\mu}$$

where Re is the Reynolds number

v is the velocity

ρ is the density

μ is the viscosity of the fluid

d is the pipe diameter

The pipe diameter and the density remain constant for most liquid applications. At very low velocities or high viscosities, Re is low, and the liquid flows in smooth layers with the highest velocity at the centre of the pipe and low velocities at the pipe wall where the viscous forces restrain it. This type of flow is called the *laminar flow* [Figs. 11.1(a) and (b)]. The corresponding Re values are below 2000 if CGS units are used for the parameters. A characteristic of laminar flow is the parabolic shape of its velocity profile.

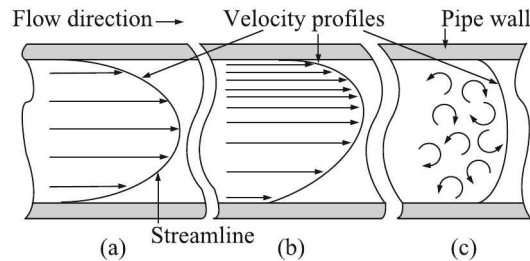


Fig. 11.1 Different types of flow: (a) laminar flow (symmetric to the axis), (b) laminar flow (asymmetric), and (c) turbulent flow.

However, most applications involve the *turbulent flow*, with Re values above 3000. Turbulent flow occurs at high velocities or low viscosities. The flow breaks up into turbulent eddies that flow through the pipe with the same average velocity [Fig. 11.1(c)]. Fluid velocity is less significant, and the velocity profile is much more uniform in shape. A transition zone exists between turbulent and laminar flows. Depending on the piping configuration and other installation conditions, the flow may be either turbulent or laminar in this zone.

With this background on types of flows, we now move on to consider the flowmetering devices.

¹Named after an English engineer Osborne Reynolds (1842–1912).

11.2 Head-type Flowmeters

This type of flow measurement follows from Bernoulli's theorem² according to which for a fluid in laminar flow along the streamline

$$gh + \frac{v^2}{2} + \frac{p}{\rho} = \text{constant} \quad (11.1)$$

where g is the acceleration due to gravity
 h is the head, i.e. the height of the streamline from a datum level
 v is the velocity of the fluid in motion
 p is the pressure acting on the fluid
 ρ is the density of the fluid.

The three terms of Eq. (11.1) are called *elevation head*, *velocity head* and *pressure head*, respectively.

Consider a flow-tube of varying cross-sectional area and having a difference in level as shown in Fig. 11.2.

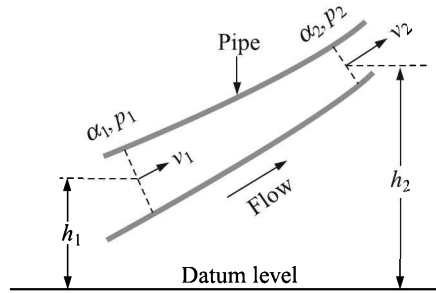


Fig. 11.2 Flow of fluid through a pipe.

From Bernoulli's theorem, we have

$$gh_1 + \frac{v_1^2}{2} + \frac{p_1}{\rho} = gh_2 + \frac{v_2^2}{2} + \frac{p_2}{\rho} \quad (11.2)$$

If the pipe is horizontal, i.e., $h_1 = h_2$,

$$\frac{v_1^2}{2} + \frac{p_1}{\rho} = \frac{v_2^2}{2} + \frac{p_2}{\rho} \quad (11.3)$$

If the flow is continuous, the volume of fluid passing per second Q_t is

$$Q_t = \alpha_1 v_1 = \alpha_2 v_2 \quad (11.4)$$

where α_1 and α_2 are the cross-sectional areas of the tube at the two locations. Equation (11.4) yields

$$v_1 = m v_2 \quad \text{where} \quad m = \frac{\alpha_2}{\alpha_1}$$

²See, for example, *Physics* (Part I), R Resnick and D Halliday, Wiley Eastern, New Delhi (1990), p. 447. Daniel Bernoulli (1700–1782) was a Swiss mathematician. The last name is pronounced as 'ber'nü-lë'.

Now from Eq. (11.3)

$$v_2^2 - v_1^2 = \frac{2(p_1 - p_2)}{\rho}$$

or
$$v_2^2(1 - m^2) = \frac{2(p_1 - p_2)}{\rho}$$

\Rightarrow
$$v_2 = \frac{1}{\sqrt{1 - m^2}} \sqrt{\frac{2(p_1 - p_2)}{\rho}}$$

Therefore,
$$Q_t = \alpha_2 v_2 = \frac{\alpha_2}{\sqrt{1 - m^2}} \sqrt{\frac{2(p_1 - p_2)}{\rho}} \quad (11.5)$$

While deriving Eq. (11.5), it has been assumed that

1. The fluid flow is one-dimensional
2. The fluid is incompressible
3. The flow is frictionless, and
4. There is no elevation change

In practice, situations deviate from these assumptions and hence corrections are necessary. It is found that the relation between the actual flow rate Q_a and Q_t is

$$Q_a = C_d Q_t \quad (11.6)$$

where C_d is called the *discharge coefficient*.

The discharge coefficient of a given installation depends mainly on the Reynolds number Re at the orifice. The variation of C_d with Re for two values of the constriction factor β (defined as, orifice diameter \div pipe diameter) is shown in Fig. 11.3.

The term $1/\sqrt{1 - m^2}$ of Eq. (11.5) is often called the *velocity of approach factor* (E), and $C_d/\sqrt{1 - m^2}$, the *flow coefficient*.

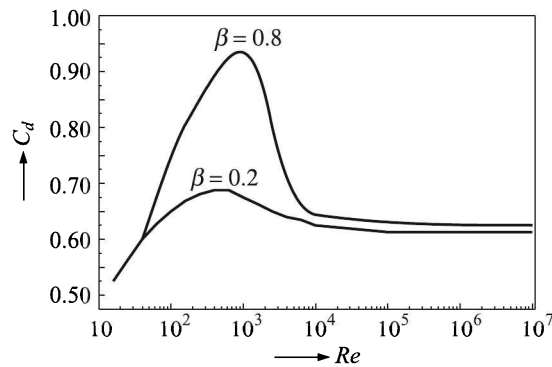


Fig. 11.3 Discharge coefficient vs. Reynolds number for different values of the constriction or ratio factor (orifice diameter \div pipe diameter) β .

- Note:*
- Equations (11.5) and (11.6) are true as long as the flow is laminar, i.e. $Re < 2000$ (approx.). Turbulence sets in at higher values of Re . Usually, the flow rates corresponding to $10^4 < Re < 10^6$ are used in industries. C_d takes care of those situations to some extent. However, ready-reckoner charts are available from which a knowledge of Re , m and C_d helps us to calculate Q_a more accurately. Or else, the resulting error may be considerable³.
 - For compressible fluids, such as gases and vapours, another factor, called the *expansion ratio*, B , is added to Eq. (11.6) so that it is modified to

$$Q_a = BC_d E \alpha_2 \sqrt{\frac{2\Delta p}{\rho}}$$

- For liquids, $B \simeq 1$ while for gases and vapours $B < 1$. There are quite a few equations relating B to m , $\Delta p/\rho$, and the specific heat ratio γ of gases. But for all practical purposes, B is figured out from the available standard tables.

We now discuss six flowmeters which function on the basis of this principle.

Orifice Plate

The orifice plate is the most widely used flow-metering element mainly because of its simplicity and low cost. A diagram of the arrangement, along with the flow profile inside it is shown in Fig. 11.4(a).

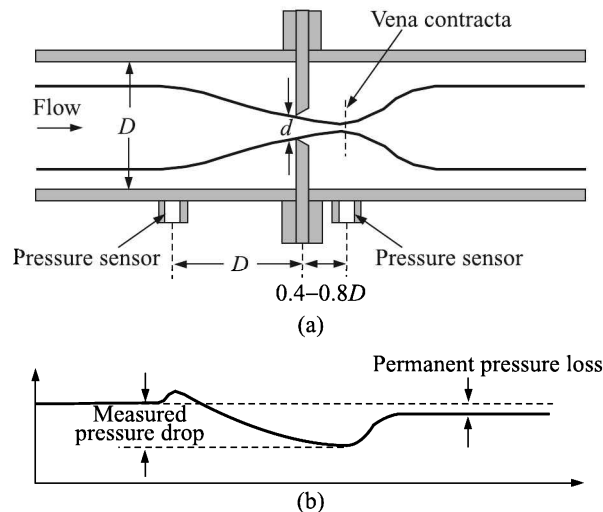


Fig. 11.4 (a) An orifice-plate arrangement along with the flow profile inside the pipe, and (b) liquid pressure along the pipe.

³See, for example, *Instrument Technology* (Vol. I), E B Jones, Butterworths, London (1974), p 178ff.

Construction. The orifice plate is basically a metal plate with, usually, a circular opening. The opening (or orifice) may be concentric, eccentric or segmented as shown in Fig. 11.5(a). The concentric type is by far the most widely used. The orifice has a sharp edge cut as per dimensions given in Fig. 11.5(b).

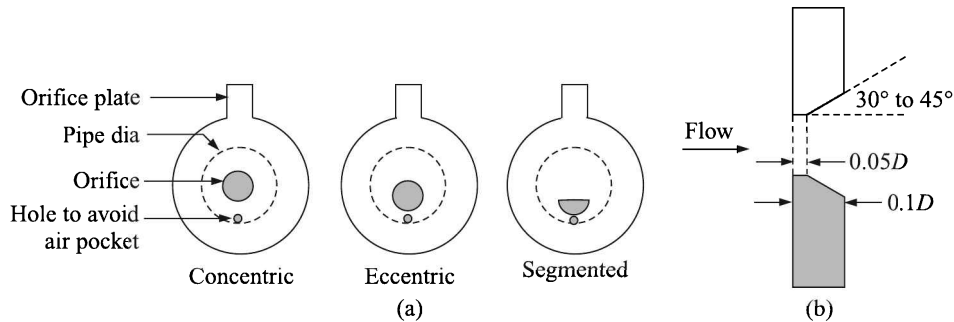


Fig. 11.5 (a) Different openings of orifice plates, and (b) sharp edge cut of orifice plate. D indicates the diameter of the pipe.

When inserted in a pipeline, the plate causes an increase in flow velocity and a consequent decrease in downstream pressure [see Fig. 11.4(b)]. The flow pattern shows an effective decrease in the cross-section of flow a little beyond the orifice plate. This position where the flow velocity is maximum and the fluid pressure minimum, is known as the *vena contracta*⁴ [see Fig. 11.4(a)].

Pressure trapping. There are in general three methods of placing the taps:

| Location | Tap |
|----------------|--|
| Flange | 25 mm upstream and 25 mm downstream from the face of orifice |
| Vena contracta | $1D^5$ upstream and $0.4D$ to $0.8D$ downstream from the face of orifice |
| Pipe | $2.5D$ upstream and $8D$ downstream from the face of orifice |

As discussed before, the flow rate can be determined from the pressure-drop measurement.

Features. A widely used device, the orifice plate has the following features:

| | |
|---------------------------------------|--|
| <i>Recommended fluids</i> | Both clean and dirty liquids and some slurry |
| <i>Rangeability</i> ⁶ | 4:1 |
| <i>Head loss</i> | For $\beta = 0.5$, about 70–75% of the orifice differential |
| <i>Typical accuracy</i> | 2 to 4% of FS |
| <i>Required upstream straight run</i> | $10D$ to $30D$ |
| <i>Discharge coefficient</i> | Typically 0.6 |
| <i>Viscosity effect</i> | High |
| <i>Relative cost</i> | Low |

⁴Meaning *contracted portion of a liquid jet*.

⁵Indicates the diameter of the pipe.

⁶See Section 11.7 at page 483 for definition.

Venturi Tube

Some of the disadvantages of the orifice plate are minimised in the Venturi⁷ tube, the design of which comprises three sections, namely

1. Converging conical section at the upstream
2. Cylindrical throat
3. Diverging recovery outlet cone at the downstream

A sectional view of the Venturi tube is shown in Fig. 11.6.

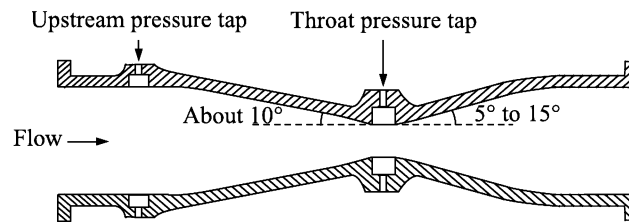


Fig. 11.6 Venturi tube.

Because of the cone and the gradual reduction in the area, there is no vena contracta in a Venturi tube flow. The flow area is at a minimum at the throat. The discharge coefficient of a Venturi tube is nearly 0.98 and this remains substantially constant for $\beta = 0.25$ to 0.75 where $\beta = \text{throat diameter} \div \text{pipe diameter}$. If the discharge coefficient is high, it means the deviation from the theoretical flow rate value is less.

That happens because the Venturi tube has a streamlined construction for which there is minimum energy loss owing to the impact of the fluid flow at the throat area. This energy loss is pretty high in an orifice plate where the flowing fluid has to negotiate an obstacle in the form of the orifice plate all of a sudden. Also, the vortices formed behind the orifice plate cause some energy loss. These energy losses at the obstacle is also responsible for substantial head loss in an orifice meter.

Features. The features of the Venturi tubes are:

| | |
|---------------------------------------|--|
| <i>Recommended fluids</i> | Clean, dirty and viscous liquids and some slurry |
| <i>Rangeability</i> | 4:1 |
| <i>Head loss</i> | Low |
| <i>Typical accuracy</i> | 1% of FS |
| <i>Required upstream straight run</i> | 5D to 20D |
| <i>Discharge coefficient</i> | Typically 0.975 |
| <i>Viscosity effect</i> | High |
| <i>Relative cost</i> | Medium |

Example 11.1

A Venturi tube of throat diameter 6 cm is placed in a water pipe of diameter 10 cm to measure the volumetric flow of rate which is found to be $0.08 \text{ m}^3/\text{s}$. If the density and viscosity of water are 10^3 kg/m^3 and $10^{-3} \text{ Pa}\cdot\text{s}$ respectively, determine

⁷Named after Giovanni Battista Venturi (1746–1822), an Italian physicist.

- (a) Reynolds number for these conditions
 (b) Upstream-to-throat differential pressure developed (given, discharge coefficient = 0.99)

Solution

(a) Since, $Q_a = (\pi d_{\text{pipe}}^2/4)v$, we have $v = 4Q_a/(\pi d_{\text{pipe}}^2)$, and therefore,

$$\begin{aligned} Re &= \frac{v\rho d_{\text{throat}}}{\eta} = \frac{4Q_a\rho d_{\text{throat}}}{\pi\eta d_{\text{pipe}}^2} \\ &= \frac{4 \times 0.08 \times 10^3 \times 0.06}{\pi \times 10^{-3} \times 0.1^2} \simeq 6.112 \times 10^5 \end{aligned}$$

(b) From Eqs. (11.5) and (11.6), we get

$$\begin{aligned} \Delta p &= (1 - m^2) \left[\frac{Q_a}{C_d \alpha_{\text{pipe}}} \right]^2 \frac{\rho}{2} \\ &= (1 - 0.6^2) \left[\frac{0.08}{0.99 \times \pi \times 0.05^2} \right]^2 \frac{10^3}{2} \text{ Pa} \simeq 33.9 \text{ kPa} \end{aligned}$$

Note: In SI units, the standard atmospheric pressure, i.e. 760 mm Hg, is nearly 101 kPa.

Example 11.2

A Venturi tube of throat diameter 5 cm has a discharge coefficient of 0.98, and with a flow rate of 10 dm³/s the pressure differential is 12.5 kPa. Determine the flow rate when an orifice of 5 cm is used in the same pipe (discharge coefficient 0.60) and the pressure differential is the same.

Solution

Other conditions remaining unchanged, we get from Eq. (11.6)

$$\begin{aligned} [Q_a]_{\text{orifice}} &= \frac{[C_d]_{\text{orifice}}}{[C_d]_{\text{Venturi}}} [Q_a]_{\text{Venturi}} \\ &= \frac{0.60}{0.98} \times 10 \text{ dm}^3/\text{s} = 6.12 \text{ dm}^3/\text{s} \end{aligned}$$

Example 11.3

A Venturimeter is used to measure the volume flow rate of an oil having a density of 850 kg/m³. It is fitted in a vertical pipe line with oil flowing downwards. Its diameters at the inlet and throat are 0.3 m and 0.2 m respectively. The pressures at the inlet and throat are measured by pressure transducers and are found to be 1.8 × 10⁵ Pa and 1.4 × 10⁵ Pa respectively. The difference in height between the inlet and throat is 0.5 m. The discharge coefficient of the Venturi tube is 0.95. Determine the volume flow rate of the oil.

Solution

Given:

$$\begin{aligned} \rho &= 850 \text{ kg/m}^3 & p_1 &= 1.8 \times 10^5 \text{ Pa} & p_2 &= 1.4 \times 10^5 \text{ Pa} \\ d_1 &= 0.3 \text{ m} & d_2 &= 0.2 \text{ m} & \Delta h &= h_{\text{inlet}} - h_{\text{throat}} = 0.5 \text{ m} \\ C_d &= 0.95 \end{aligned}$$

$$\begin{aligned} \therefore m &= \frac{\alpha_2}{\alpha_1} = \frac{d_2^2}{d_1^2} \\ &= \frac{0.2^2}{0.3^2} = \frac{4}{9} \end{aligned}$$

$$\text{and } \alpha_2 = \frac{\pi(0.2)^2}{4} = 0.01\pi$$

From Eq. (11.2), we get on rearranging

$$v_2 = \frac{1}{\sqrt{1-m^2}} \sqrt{\frac{2\Delta p}{\rho} + 2g\Delta h}$$

where $\Delta p = p_1 - p_2 = 1.8 \times 10^5 - 1.4 \times 10^5 = 0.4 \times 10^5$ Pa.

Therefore, the volume flow rate is given by

$$\begin{aligned} Q &= C_d \alpha_2 v_2 = \frac{C_d \alpha_2}{\sqrt{1-m^2}} \sqrt{\frac{2\Delta p}{\rho} + 2g\Delta h} \\ &= \frac{(0.95)(0.01\pi)}{\sqrt{1-(4/9)^2}} \sqrt{\frac{2(0.4 \times 10^5)}{850} + 2(9.81)(0.5)} \\ &\cong 0.34 \text{ m}^3/\text{s} \end{aligned}$$

Flow Nozzle

Flow nozzles (shown in Fig. 11.7) are similar to orifice meters and they work on the same principle.

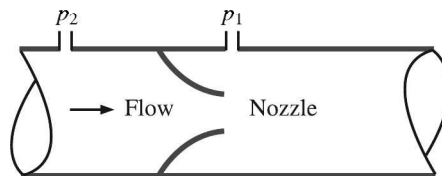


Fig. 11.7 Flow nozzle.

But at high velocities, flow nozzles can handle approximately 60% greater liquid flow than orifice plates, despite having the same pressure drop. They are less dust sensitive than orifices and require a smaller straight run of pipe for installation. They have less wear at high velocities and temperature and the pipe roughness has smaller influence on their function. However, use of the units is not recommended for highly viscous liquids or those containing large amounts of sticky solids.

Features of flow nozzles compare well with Venturi tubes though they are costlier than orifice plates.

Dall Flow Tubes

Dall flow tubes are somewhat similar to Venturi tubes except that they do not have the entrance cone. They have a tapered throat, but the exit is elongated and smooth. The distance between the front face and the tip is approximately one-half the pipe diameter. Pressure taps are located about one-half pipe diameter downstream and one pipe diameter upstream.

Since these meters have significantly lower permanent pressure losses than the orifice meters, the Dall flow tubes have widely been used for measuring the flow rate of large pipework.

Pitot Tube

A Pitot⁸ tube (Fig. 11.8) is a cylindrical probe inserted into the fluid stream which converts the velocity head to an impact (or stagnation) pressure.

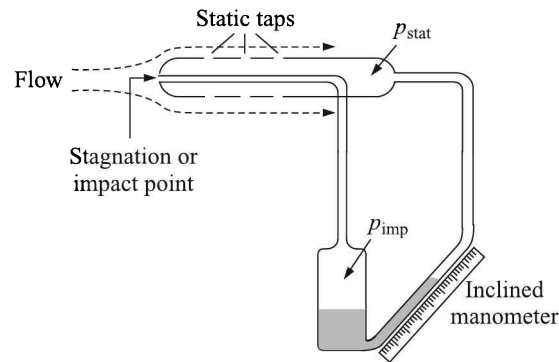


Fig. 11.8 Schematic diagram of a Pitot tube.

It consists of two coaxial tubes. The open end of the inner tube faces the incoming fluid. The outer tube has a closed end but it contains a few holes in its walls. With this kind of arrangement a situation is created where the flow velocity of the fluid is zero in one of the two tubes which are interconnected through a manometer.

The outer tube senses the static pressure of the fluid as well as its velocity head, while the inner tube senses only the impact pressure. The difference between the static pressure and the impact pressure is a measure of the flow rate. We may work out the quantitative relation between the flow rate and the pressure difference as follows. From Bernoulli's theorem [Eq. (11.1)], we have

$$\frac{p_{\text{stat}}}{\rho} + \frac{v^2}{2} = \frac{p_{\text{imp}}}{\rho} \quad (11.7)$$

which gives

$$v = \sqrt{\frac{2(p_{\text{imp}} - p_{\text{stat}})}{\rho}}, \quad (11.8)$$

where, p_{stat} is the static pressure within the tube and p_{imp} is the impact or stagnation pressure.

⁸Invented by Henri Pitot (1695–1771), a French hydraulic engineer. Pronounced as 'pi-tow'.

Note: Equation (11.8) has been derived with an assumption that the fluid is incompressible. While this assumption may hold good for liquids, it is not true for gases which are highly compressible. That is why corrections are necessary when the air velocity is measured by the Pitot tube.

In an actual Pitot tube, deviations from the theoretical result arise from a number of sources, such as

- (a) Misalignment of the tube-axis and the velocity vector
- (b) Non-zero tube diameter. Streamlines next to the tube must be longer than those in the undisturbed flow, which indicates an increase in velocity, and a consequent decrease in static pressure.

An important application of the Pitot tube is found in aircrafts and missiles. Here p_{imp} and p_{stat} readings of a tube fastened to a vehicle are used to determine the air velocity and the Mach number⁹ while the static reading alone is utilised to measure altitude.

Example 11.4

- (a) Determine the flow velocity of water of density 1000 kg/m^3 at the head of a Pitot tube if it produces a pressure differential of 10 kPa between the outlets.
- (b) If the same pressure differential is obtained in air at an altitude where the density of air is 0.650 kg/m^3 , determine the velocity of air flow.

Solution

(a) From Eq. (11.8), we have

$$v_{\text{water}} = \sqrt{\frac{2 \times 10 \times 10^3}{10^3}} \text{ m/s} \simeq 4.47 \text{ m/s}$$

(b) From the same equation, we get

$$v_{\text{air}} = \sqrt{\frac{2 \times 10 \times 10^3}{0.65}} \text{ m/s} \simeq 175.41 \text{ m/s}$$

Air being compressible, the measured value of the air velocity value is rather approximate.

Rotameter

From Eqs. (11.5) and (11.6), we get that the flow rate of an incompressible fluid through a horizontal pipe is given by

$$Q_a = \frac{C_d \alpha_2}{\sqrt{1 - m^2}} \sqrt{\frac{2(p_1 - p_2)}{\rho}} \quad (11.9)$$

where the symbols have their usual meaning.

In orifice meter or Venturi tube, the cross-section of the orifice α_2 is held constant and therefore $Q_a \propto \sqrt{p_1 - p_2}$. Now, if we make arrangements such that α_2 is allowed to vary and $(p_1 - p_2)$ is held constant, then Q_a will vary as α_2 . This concept is utilised to construct a rotameter.

⁹ $N_M = v/v_s$, where v_s = velocity of sound in air $\simeq 330 \text{ m/s}$. It is named after Ernst Mach (1838–1916), an Austrian physicist.

A rotameter consists of a vertical tube with a tapered bore [Fig. 11.9(a)] in which a float assumes a vertical position depending on the rate of flow through the tube. The float remains at an equilibrium position because the vertical forces acting on it—differential pressure, weight, viscosity and buoyancy—are in balance.

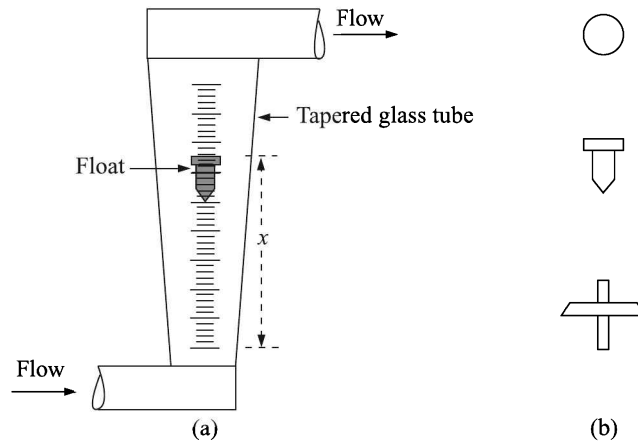


Fig. 11.9 (a) Rotameter, and (b) float shapes.

Seen quantitatively, the downward forces on the float are:

1. weight of the float = $V\rho g$, where V is the volume and ρ is the density of the float;
2. due to pressure on its upper surface = $p_2\alpha$, where α is the cross-sectional area of the float;

and the upward forces on it are:

1. buoyancy = $V\rho_f g$ where ρ_f is the density of the fluid;
2. due to pressure on its lower surface = $p_1\alpha$;
3. viscous drag due to fluid flow = $Kv\mu$ where v is the velocity of fluid flow, μ is the viscosity coefficient and K is a constant.

Neglecting the viscous drag which is small and equating the upward and downward forces, we get for equilibrium of the float

$$V\rho g + p_2\alpha = V\rho_f g + p_1\alpha$$

$$\text{or} \quad p_1 - p_2 = \frac{Vg}{\alpha}(\rho - \rho_f) \quad (11.10)$$

Thus from Eqs. (11.9) and (11.10), we get

$$Q_a = \frac{C_d(\alpha_1 - \alpha)}{\sqrt{1 - m^2}} \sqrt{\frac{2Vg(\rho - \rho_f)}{\alpha\rho_f}} \quad (11.11)$$

where $(\alpha_1 - \alpha)$ is the difference in the cross-sectional areas of the tube and the float and which eventually is the cross-sectional area of the orifice, i.e. α_2 .

If C_d is nearly constant and $m \ll 1$, then from Eq. (11.11)

$$Q_a = C(\alpha_1 - \alpha)$$

where C is a constant.

Again, if the geometry of the tube is such that $\alpha_1 \propto x$, where x is the displacement of the float from a datum level, then

$$Q_a = K_1 + K_2x$$

where K_1 and K_2 are constants. Thus, the flow rate has a linear relation with the float position.

Density compensation. Equation (11.11) shows that the flow rate depends on the fluid density. That means, the rotameter calibration for one fluid will not remain valid for another of different density. This difficulty can be obviated from the consideration of invariance of mass-flow rate, i.e. $d(\rho_f Q_a)/d\rho_f = 0$. With this condition, Eq. (11.11) yields

$$\rho = 2\rho_f \quad (11.12)$$

On substituting this condition in Eq. (11.11), we observe that the right-hand side of the equation becomes free from all density factors. Therefore, to keep the calibration intact, the material of the float should be changed in such a way that Eq. (11.12) is satisfied. This is called the *density compensation* of the float.

Construction. A conical tube made of high-strength glass allows direct observation of the float position. In case greater strength is required, metal tubes can be used and the float position can be monitored with the help of a suitable, say magnetic, transducer.

Floats are made of brass, stainless steel, monel or special plastics. Some float shapes, such as spheres, require no guiding in the tube; others are kept central by guide wires. Often, floats are made with sharp edges [see Fig. 11.9(b)] which create uniform turbulence at both high and low flow rates.

Range and accuracy. The typical range of a rotameter is nearly 10:1, accuracy, $\pm 2\%$ of full scale and repeatability, about 0.25% of reading. Sharp edge floats may increase the range to 100:1.

Example 11.5

Design a rotameter taper for flow of water up to 40 litres/min with float volume = 8 cm³, float diameter = 2 cm, tube length = 25 cm and tube inlet diameter = 2 cm. Assume flow coefficient = 1 and that the float is compensated for fluid density.

Solution

Let the rotameter taper be γ . Then, if D is the tube diameter at the top, d the tube diameter at the bottom and l the tube length,

$$D = d + \gamma l$$

The annular area,

$$\alpha_1 - \alpha = \frac{\pi}{4}(D^2 - d^2) = \frac{\pi}{4}\gamma(\gamma l^2 + 2ld)$$

Given:

$$\begin{array}{llll} d = 2 \text{ cm} & l = 25 \text{ cm} & (\rho - \rho_f)/\rho_f = 1 & V = 8 \text{ cm}^3 \\ g = 980 \text{ cm/s}^2 & Q_a = 40 \text{ L/min} & C_d/\sqrt{1 - m^2} = 1 & \end{array}$$

So

$$\alpha = \frac{\pi d^2}{4} = \frac{\pi(2)^2}{4} = \pi$$

Now we have

$$\begin{aligned} Q_a &= \frac{40 \times 10^3}{60} = 6.667 \times 10^2 \text{ cm}^3/\text{s} \\ &= (\alpha_1 - \alpha) \sqrt{\frac{2Vg(\rho - \rho_f)}{\alpha\rho_f}} && \text{[from Eq. (11.11)]} \\ &= \frac{\pi}{4} \gamma (\gamma l^2 + 2ld) \sqrt{\frac{2Vg(\rho - \rho_f)}{\alpha\rho_f}} \end{aligned}$$

Substituting the values of l , d , V , g , $(\rho - \rho_f)/\rho_f$ and α , we get

$$6.667 \times 10^2 = \frac{\pi}{4} \gamma (625\gamma + 100) \sqrt{\frac{2 \times 8 \times 980}{\pi}}$$

This yields

$$625\gamma^2 + 100\gamma - 12.0155 = 0$$

the positive root of which gives

$$\gamma = 0.08$$

Example 11.6

Figure 11.10 shows the float of a rotameter stationary at level xx for a certain flow rate of water. The specific gravity of the float is 2.0, mass 10^{-2} kg and base area $1.0 \times 10^{-2} \text{ m}^2$.

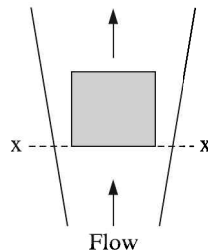


Fig. 11.10 Example 11.6.

Neglecting the effect due to viscosity,

- Calculate the pressure drop across the float.
- If the velocity of water before the float is $v_1 = 0.1 \text{ m/s}$, then using Bernoulli's equation find the velocity just after the section xx.

Solution

Given specific gravity of the float $s = 2.0$, mass of the float $m = 10^{-2} \text{ kg}$, base area $\alpha = 1.0 \times 10^{-2} \text{ m}^2$.

- (a) If ρ and ρ_w are the densities of the float and water respectively, the volume of the float $V = (m/\rho)$. Therefore from Eq. (11.10), we have

$$\begin{aligned} p_1 - p_2 &= \frac{mg}{\rho\alpha}(\rho - \rho_w) \\ &= \frac{mg}{\alpha} \left(1 - \frac{\rho_w}{\rho}\right) = \frac{mg}{\alpha} \left(1 - \frac{s_w}{s}\right) \\ &= \frac{(10^{-2})(9.81)}{10^{-2}} \left(1 - \frac{1}{2}\right) = 4.9 \text{ N/m}^2 \end{aligned}$$

- (b) The difference in height of the top and bottom of the float is

$$h_1 - h_2 = \frac{V}{\alpha} = \frac{m}{\rho\alpha}$$

Therefore from Bernoulli's equation, i.e. Eq. (11.2), we have on rearranging

$$\frac{v_2^2}{2} = g(h_1 - h_2) + \frac{p_1 - p_2}{\rho_w} + \frac{v_1^2}{2}$$

On substituting the values of the variables that we have so far obtained and utilising Eq. (11.10), we get

$$\frac{v_2^2}{2} = \frac{mg}{\rho\alpha} + \frac{mg}{\rho\rho_w\alpha}(\rho - \rho_w) + \frac{v_1^2}{2}$$

Thus,

$$\begin{aligned} v_2 &= \sqrt{\frac{2mg}{\rho\alpha} + \frac{2mg}{\rho\rho_w\alpha}(\rho - \rho_w) + v_1^2} \\ &= \sqrt{\frac{2mg}{\rho_w\alpha} + v_1^2} \\ &= \sqrt{\frac{2(10^{-2})(9.81)}{(10^3)(10^{-2})} + (0.1)^2} = 0.14 \text{ m/s} \end{aligned}$$

11.3 Velocity Measurement-type Flowmeters

Electromagnetic Flowmeter

These flowmeters utilise the principle of electromagnetic induction. Consider a rectangular loop of wire, one end of which is in a uniform field B at right angles to the loop. This field of \mathbf{B} is produced in the gap of a large electromagnet [Fig. 11.11(a)]. Suppose, the loop is being pulled to the right at a constant velocity v . The flux Φ_B enclosed by the loop is

$$\Phi_B = Blx \quad (11.13)$$

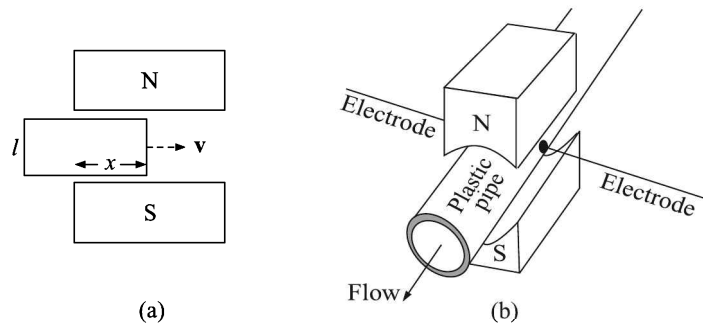


Fig. 11.11 (a) Rectangular loop in a magnetic field, and (b) the electromagnetic flowmeter.

where lx is the area of that part of the loop where B is not zero. The emf \mathcal{E} is

$$\begin{aligned} |\mathcal{E}| &= \frac{d\Phi_B}{dt} = \frac{d}{dt}(Blx) = Bl \frac{dx}{dt} \\ &= Blv \text{ (volt)} \end{aligned} \quad (11.14)$$

where B is the flux density, in tesla
 l is the length, in m
 v is the velocity, in m/s.

- Note:*
1. The only dimension of the loop that enters in Eq. (11.14) is the length of the conductor.
 2. The induced emf \mathcal{E} develops perpendicular to both B and v .

Thus, an electric field \mathcal{E} acts on conductor when it moves in a magnetic field and that \mathcal{E} acts in direction which is perpendicular to both the direction of the magnetic field and the direction of motion of the conductor.

Now consider a cylindrical jet of conducting fluid with a uniform velocity profile, traversing a magnetic field. An electric field will act on it and will polarise it to produce positive and negative ions. These ions, forced to opposite sides of the jet, will give rise to a potential difference as discussed above. If, l is the diameter of the pipe, and if Q_v is the volume flow rate in litre/s ($= 10^{-3} \text{ m}^3/\text{s}$), we have

$$Q_v \times 10^{-3} = \frac{\pi l^2}{4} v$$

or
$$v = \frac{4Q_v}{\pi l^2} \times 10^{-3} \text{ m/s} \quad (11.15)$$

Substituting Eq. (11.15) in Eq. (11.14), we get

$$|\mathcal{E}| = \frac{4B}{\pi l} Q_v \times 10^{-3} \text{ V} \quad (11.16)$$

The emf is measured by means of two electrodes built into a non-magnetic length of the pipe [Fig. 11.11(b)]. In practice, an alternating magnetic field is used to avoid permanent polarisation of the elements, i.e. collection of gas bubbles on the electrodes owing to electrolytic action.

Advantages and disadvantages. The advantages and disadvantages of electromagnetic flowmeters are listed below:

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|---|
| 1. Electromagnetic flowmeter does not obstruct the fluid flow. | 1. The liquid has to be conducting (conductivity $> 10^{-6}$ mho-cm); this requirement eliminates its use for all gases and for most of the hydrocarbons. |
| 2. It does not cause any permanent pressure drop in the fluid flow. | 2. The pipe should always be full. |
| 3. It is very suitable for metering corrosive acids, cement slurries, detergents, greasy and sticky fluids, etc. | 3. The pipe has to be non-magnetic and non-conducting. |
| 4. It can be used to measure bi-directional flow. | |
| 5. It can measure very low flow rates. | |

Example 11.7

Calculate the generated emf (rms value) by the magnetic flowmeter if the ordinary tap water flows at the rate of 400 litres per minute through its tube of 8 cm diameter. Assume the maximum flux density of the alternating magnetic field established in the flowmeter to be 400 lines per cm^2 (gauss).

Solution

Given:

$$Q_v = \frac{400}{60} = 6.667 \text{ litre/s} \quad l = 8 \text{ cm} = 0.08 \text{ m} \quad B = 400 \text{ gauss} = 400 \times 10^{-4} \text{ tesla}$$

Substituting these values in Eq. (11.16), we get

$$\mathcal{E} = \frac{4(6.667)(400 \times 10^{-4})}{\pi(0.08)} \times 10^{-3} \text{ V} \cong 4.24 \text{ mV}$$

or

$$\mathcal{E}_{\text{rms}} = \frac{2 \times 4.24}{\sqrt{2}} \text{ mV} = 6.0 \text{ mV}$$

Turbine-type Flowmeter

Meteorological observatories use a cup-type anemometer for measuring the wind velocity. The rotational speed of the anemometer is calibrated against the wind velocity. A turbine-type meter is a variant of this anemometer where a turbine wheel is placed in the pipe through which the fluid flows (Fig. 11.12).

As the fluid flows, the turbine wheel rotates, its angular velocity depending on the flow rate of the fluid. The angular velocity of the turbine wheel is conveniently measured with the help of a magnetic pick-up which produces voltage pulses as each blade of the turbine moves past it. Either these pulses can be counted by a digital counter over a fixed period of time to determine the pulse rate, or the pulses can be fed to a frequency-to-voltage converter to produce an analogue voltage which can be calibrated against the flow rate. It can be seen that

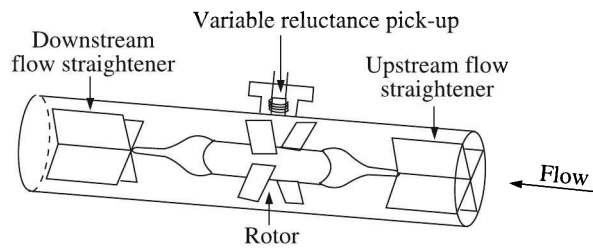


Fig. 11.12 Turbine-type flowmeter.

the angular velocity (or pulse rate) of the rotor is directly proportional to the volume flow rate of the fluid.

Let us consider the turbine as a screw.

If p is the pitch of the screw

v is the velocity of fluid flow

ω is the angular velocity of the turbine rotor

D is the outer diameter of the turbine

d is the rotor hub diameter

A is the annular area through which fluid flows = $\frac{\pi}{4}(D^2 - d^2)$

Q_v is the volume flow rate of the fluid

then we have

$$v = \frac{Q_v}{A} = \frac{4Q_v}{\pi(D^2 - d^2)}$$

$$\omega = \frac{v}{p} = \frac{4Q_v}{\pi p(D^2 - d^2)}$$

Therefore

$$Q_v = \frac{\pi p(D^2 - d^2)}{4} \omega \quad (11.17)$$

Since all other quantities, except ω , on the right-hand side of Eq. (11.17) are constant for a particular turbine, it follows that

$$Q_v = k\omega = K\nu \quad (11.18)$$

where ν is the pulse rate detected at the magnetic pick-up, k and K being constants.

Linearity of a turbine-type flowmeter is good at higher flow rates while viscous drag of the fluid and magnetic pick-up drag degrade it at low flow rates.

Commercially available turbine-type flowmeters range from about 1 litre/min to 100,000 litre/min for liquids and 5 litre/min to 100,000 litre/min for gases. The accuracy is sometimes better than 0.1% FS, and since they are essentially first-order instruments, they can follow the flow transients rather accurately.

However, these meters can be used in fluid which is free from suspended particles. Also, it needs a straight run of upstream pipe of length 20 times the diameter of the pipe.

Ultrasonic Flowmeters

Ultrasonic flowmeters are of two types:

1. Doppler frequency shift type
2. Transit time type

Doppler frequency shift flowmeter

Suppose, a stationary source of sound is emitting a sound of a certain frequency. To a listener who is moving away from the source of sound, the pitch (frequency) is lower than when he is at rest. Conversely, the pitch will be higher if the listener moves towards the source of sound.

A similar phenomenon results if the listener is stationary but the source of sound is moving. Known as *Doppler effect*¹⁰ after the name of its discoverer¹¹, the frequency shift observed in this phenomenon is related to the relative velocity of the listener and the source.

This effect is used (Fig. 11.13) to measure the flow rate of a fluid, carrying suspended particles. Continuous-wave ultrasonic signal of frequency around 10 MHz is generated by a piezoelectric crystal oscillator. This signal is scattered by moving suspended particles to produce a signal of changed frequency which is detected by the receiver. It can be shown that the frequency shift is proportional to the velocity of fluid flow, which, in turn, is proportional to the volume flow rate of the fluid.

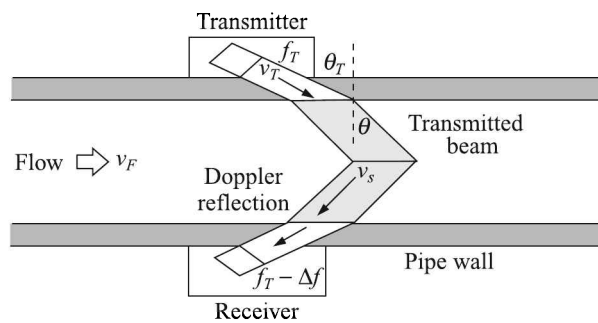


Fig. 11.13 Doppler frequency shift flowmeter principle.

Equations governing phenomena in a Doppler flowmeter are

$$\Delta f = 2f_T \left(\frac{v_F}{v_s} \right) \sin \theta \quad [\text{Doppler's law of frequency shift}] \quad (11.19)$$

$$\frac{\sin \theta_T}{v_T} = \frac{\sin \theta}{v_s} \quad [\text{Snell's law of refraction}] \quad (11.20)$$

where v_F is the flow velocity

v_T is the sonic velocity of transmitter material

v_s is the sonic velocity of the fluid

¹⁰See *Physics* (Part I), R Resnick and D Halliday, Wiley Eastern, New Delhi (1990), p 512ff.

¹¹Christian Andreas Doppler (1803–1853) was an Austrian mathematician and physicist.

f_T is the transmission frequency

Δf is the Doppler shift of frequency

θ_T is the angle of the transmitted sonic beam

θ is the angle of entry of the transmitted beam in the fluid

Solving Eqs. (11.19) and (11.20) simultaneously, we get

$$v_F = \frac{v_T}{f_T \sin \theta_T} \Delta f \equiv K \Delta f$$

where K is the sensitivity factor.

Advantages and disadvantages. Like any other flowmeter, a Doppler frequency shift flowmeter has distinct plus and minus points as given below in a table form

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|--|
| 1. Low pressure loss because of almost no obstruction to flow | 1. Highly dependent on physical properties of the fluid, such as the sonic conductivity, particle density, and flow profile |
| 2. Good accuracy, linear calibration curve, fast response | 2. Accuracy is sensitive to the velocity profile variations and to the distribution of acoustic reflectors, such as suspended particles or entrained air bubbles, in the measurement section |
| 3. Small in size (6 to 10 mm), corrosion resistant, relatively low power consuming | 3. Sensitive to changes in density and temperature of the fluid |

Transit time flowmeter

Suppose, a pulse of ultrasonic sound of a certain frequency is emitted by a source which is fixed on a pipe (Fig. 11.14). If the fluid inside the pipe is stationary, the sound will be refracted

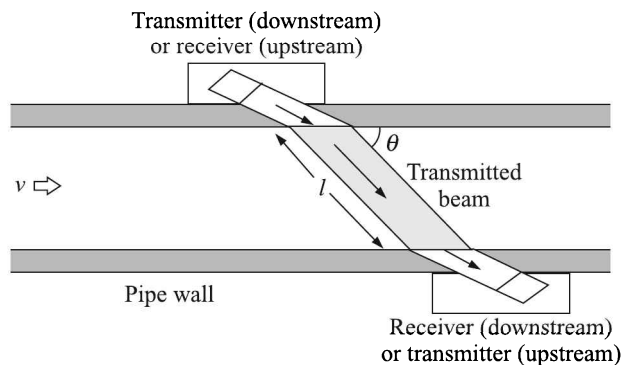


Fig. 11.14 Transit time flowmeter.

by the fluid and will be detected at the receiver after a certain time. Now, if the fluid starts flowing with a certain velocity, the velocity of propagation of the sound will change depending on the direction of fluid flow and therefore, the transit time will change. The change in the

transit time will depend on the velocity of fluid flow. This principle is utilised to construct transit time flowmeters.

Let v be the velocity of fluid flow
 c be the velocity of ultrasound in the stationary fluid
 l be the distance between transmitter and receiver
 Δt be the transit time difference

then transit times for the upstream and downstream transmissions of ultrasound are

$$t_1 = \frac{l}{c + v \cos \theta}$$

$$t_2 = \frac{l}{c - v \cos \theta}$$

Therefore,

$$\begin{aligned} \Delta t = t_2 - t_1 &= \frac{l}{c - v \cos \theta} - \frac{l}{c + v \cos \theta} \\ &\cong \frac{l}{c} \left[\left(1 + \frac{v}{c} \cos \theta\right) - \left(1 - \frac{v}{c} \cos \theta\right) \right] \quad [\because v \ll c] \\ &= \frac{2lv \cos \theta}{c^2} \end{aligned} \quad (11.21)$$

The velocity of ultrasound in the stationary fluid c is highly susceptible to temperature and pressure of the fluid. As can be seen from Eq. (11.21), a simple transit time measuring system will incorporate c which is a source of error.

However, c can altogether be eliminated by using a self-excited oscillating system in which a received pulse is fed back to trigger a transmitted pulse. Then

$$f_1 = \frac{c + v \cos \theta}{l}$$

$$f_2 = \frac{c - v \cos \theta}{l}$$

where f_1 is the repetition frequency of the received pulse at the receiver (downstream) and f_2 is the repetition frequency of the received pulse at the receiver (upstream). Thus,

$$\Delta f = f_1 - f_2 = \frac{2v \cos \theta}{l} \quad (11.22)$$

Actually, a train of pulses can be generated and the repetition frequency of these pulses will depend on the velocity of the fluid flow.

This technique had been in vogue for quite some time. But now, with the advent of the advanced techniques of digital electronics, it has been replaced with highly accurate time measurement. Since

$$\frac{1}{t_1} - \frac{1}{t_2} = \frac{2v \cos \theta}{l}$$

now v is directly measured using the relation

$$v = \frac{l}{2 \cos \theta} \left(\frac{1}{t_1} - \frac{1}{t_2} \right) = \frac{l}{2 \cos \theta} \left(\frac{t_2 - t_1}{t_1 t_2} \right)$$

Advantages and disadvantages. The ultrasonic transit time flowmeters have many advantages as well disadvantages as listed below:

| <i>Advantages</i> | <i>Disadvantages</i> |
|---|---|
| 1. Low pressure drop because of almost no obstruction to flow | 1. Operating principle demands reliability in high frequency sound transmission across the pipe |
| 2. Unaffected by changes in temperature, density or viscosity of the fluid | 2. Liquid slurries with excess solids or with entrained gases may block the transmission of ultrasonic pulses |
| 3. Bi-directional flow measurement capability, corrosion-resistance, accuracy about 1% of flow rate, relatively low power consumption | 3. Not recommended for primary or aerobically digested or dissolved air or septic or activated carbon sludge and mixed liquor |

Example 11.8

In an ultrasound flowmeter, the transducers are separated by a distance $l = 25 \text{ mm}$ with the line joining the transducers making an angle of 60° with the direction of flow. The transit time difference between upstream and downstream measurements is 10 ns with the sound velocity in the medium being 1000 m/s . Assuming that the size of the transducers is very small as compared to the diameter of the pipe, what is the flow velocity?

Solution

Given:

$$l = 25 \times 10^{-3} \text{ m} \quad \theta = 60^\circ \quad c = 1000 \text{ m/s} \quad \Delta t = 10 \times 10^{-9} \text{ s}$$

Therefore,

$$\text{Upstream velocity} = c - v \cos \theta = c - v/2$$

$$\text{Downstream velocity} = c + v \cos \theta = c + v/2$$

where v is the flow velocity. So,

$$\Delta t = \frac{l}{c - (v/2)} - \frac{l}{c + (v/2)} = \frac{lv}{c^2 - (v^2/4)} = 10 \times 10^{-9} \text{ (given)}$$

This yields the quadratic equation

$$\frac{v^2}{4} \Delta t + lv - c^2 \Delta t = 0$$

Accepting the positive root of this equation, we get

$$v = 0.4 \text{ m/s}$$

Vortex Shedding Flowmeter

All of us have seen flags fluttering and making a sound in high wind although at low velocities of air they fly steadily. Even kites flying in high wind vibrate. This occurs owing to vortex formation behind the obstructed wind path.

Consider a flow of fluid around a *bluff* (i.e. non-streamlined) *body*. At very low Reynolds numbers (based on the characteristic size of the body) the streamlines of the resulting flow are perfectly symmetric around the bluff body. However as the Reynolds number increases the flow becomes asymmetric and the so called *von Kármán*¹² *vortex street* (Fig. 11.15) occurs.

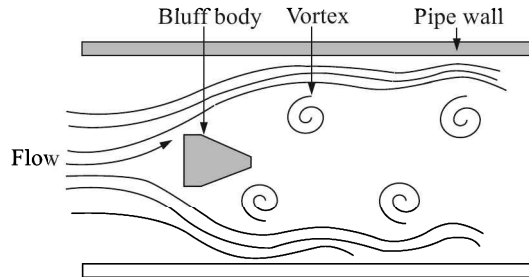


Fig. 11.15 von Kármán vortex street behind a tapered cylindrical bluff body.

This phenomenon of an unsteady flow that takes place in special flow velocities (according to the size and shape of the bluff body) is generally called *vortex shedding*. In this flow, vortices are created at the back of the body and they detach periodically from either side of the body.

Therefore, if a bluff body (aka *shedder bar*) is immersed in a flowing fluid, it vibrates owing to the vortex shedding. The frequency of vibration f is given by

$$f = N_{St} \frac{v}{L} \quad (11.23)$$

where v is the velocity of fluid flow

L is the characteristic length of the bluff body

N_{St} is the Strouhal¹³ number

Thus, the frequency of vortex induced vibration is essentially proportional to the flow rate of the fluid. This principle is utilised to construct vortex flowmeters.

Bluff body

Though any bluff body can be used to generate vortices in a flow, a careful design is necessary to make vortices regular and well defined. Essentially, the body must be

1. Non-streamlined
2. Symmetrical, and
3. Capable of generating vortices for a wide Reynolds number range

A few of the most common designs are shown in Fig. 11.16.

The last design on the bottom right hand side consists of two pieces. The second piece is used to reinforce and stabilise the shedding.

The width of the bluff body is determined by the pipe size and the rule of thumb is that the ratio of the width of the body to the pipe diameter should not be less than 0.2.

¹²Theodore von Kármán (1881–1963) was a Hungarian-American aerospace engineer and physicist. He was active primarily in the fields of aeronautics and astronautics.

¹³Named after the Czech physicist Vincenc Strouhal (1850–1922) who specialised in experimental physics.

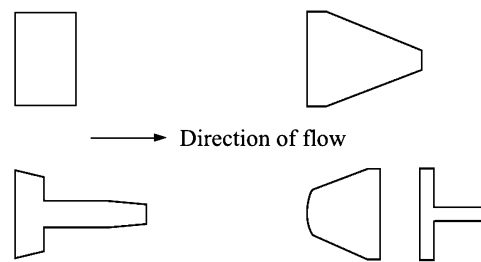


Fig. 11.16 Common bluff body shapes.

Sensing methods

Various types of methods are adopted to sense the frequency of the vortex induced vibrations:

1. Ultrasonic
2. Piezoelectric
3. Thermal
4. Shuttle ball
5. Strain gauge

Ultrasonic. An ultrasonic beam is passed through a section of the von Kármán vortex street (Fig. 11.17). The passing vortices cause modulation of the beam amplitude owing to differential refraction.

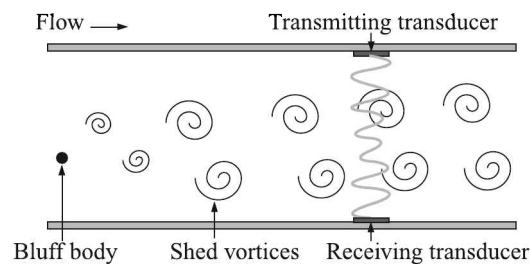


Fig. 11.17 Ultrasonic detection of vortex shedding frequency.

Piezoelectric. Inside, atop, or downstream of the shedder bar is placed a piezoelectric crystal. This crystal produces a small, but measurable, voltage pulse every time a vortex is created. The frequency of the pulse is measured and the flow rate is calculated by the flowmeter electronics.

Thermal. A thermistor is placed in a through passage across the heated shedder bar or behind its face. The thermistor will sense alternating vortices due to the cooling effect caused by their passage.

Shuttle ball. Alternating pressures caused by the vortex shedding drives a magnetic shuttle up and down the axis of the flow element (Fig. 11.18). A magnetic pick up detects this motion.

Strain gauge. The bluff body is so designed that the alternating pressures related to vortex shedding are applied to a cantilevered portion at its rear. This cyclic variation of the deformation of the rear is sensed by an internal strain gauge placed there.

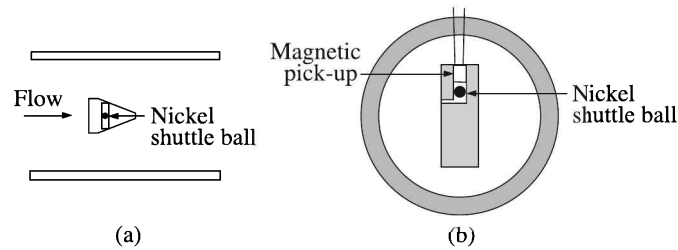


Fig. 11.18 Shuttle ball sensing of vortex shedding frequency: (a) The shuttle ball in the bluff body, and (b) cross-sectional view of placement of the shuttle ball.

Signature curve

We have seen from Eq. (11.23) that the frequency of the vortex induced vibration is linearly related to the velocity of flow of the fluid, provided the Strouhal number is constant. When considering a long circular cylinder, the Strouhal number is given by the empirical formula

$$N_{St} = 0.198 \left(1 - \frac{19.7}{Re} \right) \quad (11.24)$$

where Re is the Reynolds number. If we plot the Strouhal number vs. the Reynolds number, the resulting curve looks like that shown in Fig. 11.19. Equation (11.24) which remains valid over the range $250 < Re < 2 \times 10^5$ as well as Fig. 11.19 indicates that N_{St} becomes constant at higher Re . This variation of N_{St} , and hence the K -factor [$K = N_{St}/L$, see Eq. (11.23)] of the vortex flowmeter for a given geometry of the shedder bar is sometimes called its *signature curve*.

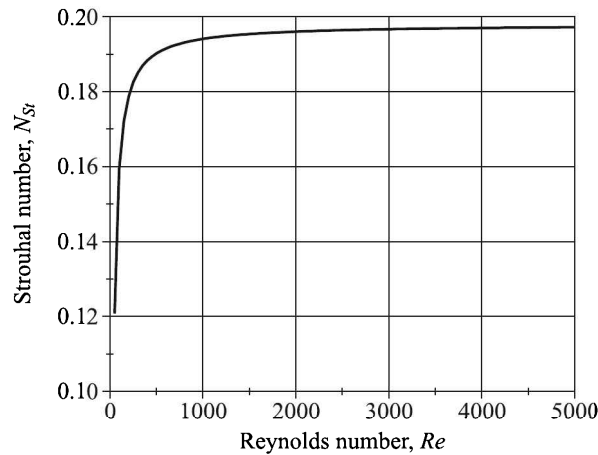


Fig. 11.19 Variation of Strouhal number with Reynolds number for a long circular cylindrical bluff body.

Lock-in. However, another factor limits the range of the vortex flowmeters. The phenomenon of *lock-in* happens when the vortex shedding frequency becomes close to the natural frequency of vibration of the bluff body. When this happens, large and damaging vibrations can result.

Installation

Care has to be taken while installing vortex flowmeters. Pipe flange gaskets upstream and at the transmitter should not protrude into the flow. To ensure uniform flow velocity, there should be straight run of pipes 20 times the diameter on the upstream and 5 times the diameter on the downstream side of the meter.

Advantages and disadvantages

The advantages and disadvantages of the vortex shedding flowmeter are given below.

| <i>Advantages</i> | <i>Disadvantages</i> |
|---|--|
| 1. Since it counts frequency, the output is digital. | 1. Low flow rate cannot be measured because of no vortex formation. |
| 2. Output frequency varies linearly with the flow velocity within the range $2 \times 10^3 < Re < 1 \times 10^5$ for gases and $4 \times 10^3 < Re < 1.4 \times 10^5$ for liquids. This feature makes them highly popular in industries where high flow rates are used. | 2. Pipe diameters below 12 mm are not practical for installation of vortex flowmeters. |
| 3. Measurement is independent of viscosity, density, temperature or pressure of the fluid. | 3. Highly viscous fluid may not produce vortices. |
| 4. Equally suitable for flow rate or flow totalisation measurements. | 4. Not suitable for bi-directional flow measurements. |
| 5. Low cost | |

Anemometers

A hot-wire anemometer consists of a small length of fine resistance wire supported on a probe (Fig 11.20) which is exposed to the fluid flow. Such anemometers are commonly made in two basic forms:

1. Constant-current type
2. Constant-temperature type

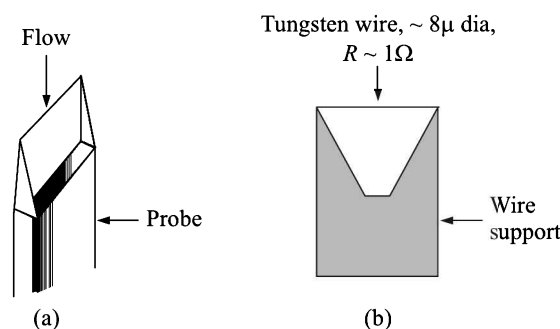


Fig. 11.20 Hot-wire anemometer: (a) perspective view, and (b) cross-sectional appearance.

Hot wire constant-current anemometer

In the constant-current type anemometer, the wire is kept heated by passing a current of fixed value through it. When the device is exposed to the fluid flow, heat is dissipated by the wire through convection, in addition to other losses owing to radiation and conduction along the wire supports. This causes a drop in temperature and a consequent decrease in its resistance. This change in resistance is a measure of the flow rate.

Thus, if the anemometer resistance is connected to one arm of a Wheatstone bridge, the change in flow rate will offset the balance of the bridge and the offset voltage measured by the connected voltmeter will be a measure of the flow rate.

Note: This kind of measurement will be accurate only if the temperature of the fluid remains constant.

Hot wire constant-temperature anemometer

In a hot wire constant-temperature anemometer, the current through the wire is adjusted by changing R_1 [Fig. 11.21(a)] so that the temperature of the wire, and therefore, R_w are kept constant. As a consequence, as shown in the analysis below, the current I becomes a function of the flow velocity v .

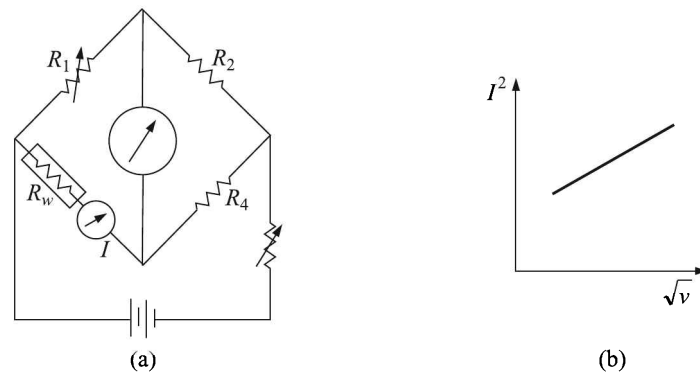


Fig. 11.21 Hot-wire anemometer: (a) measurement bridge circuit, and (b) current-velocity characteristic.

We note that the wire attains an equilibrium temperature when the Joule heat generated in it is balanced by the convective heat loss from its surface. Thus, neglecting radiative and conductive losses which are small, we can write

$$I^2 R_w = hA(T_w - T_f) \quad (11.25)$$

where h is the coefficient of heat transfer

A is the heat transfer area

T_w is the temperature of wire

T_f is the temperature of fluid

R_w is the resistance of the wire.

Now, h is mainly a function of the flow velocity for a given fluid density and its dependence on this velocity is of the form

$$h = \alpha + \beta\sqrt{v} \quad [\text{King's law}] \quad (11.26)$$

where α and β are constants. From Eqs. (i) and (ii) we get

$$I^2 = \frac{A(T_w - T_f)(\alpha + \beta\sqrt{v})}{R_w} \equiv C_1 + C_2\sqrt{v} \quad (11.27)$$

Equation (11.27) indicates that the I^2 vs. \sqrt{v} curve should be a straight line [Fig. 11.21(b)]. This is borne out by experiments as well.

Construction. Anemometer wires are generally made of platinum, tungsten or platinum-iridium alloy having diameters ranging from 2 to 5 μm and length from 2 to 5 mm, the resistance varying between 1 and 5 Ω .

Hot-film is a variation of the hot-wire element where the resistance element is a thin film of platinum deposited on a glass base [Fig. 11.22(a)]. The film takes the place of hot wire—the required circuitry is basically similar to that used in the constant-temperature type. Film elements have greater mechanical strength and may be used at very high temperatures by constructing them with internal cooling water passages [Fig. 11.22(b)].

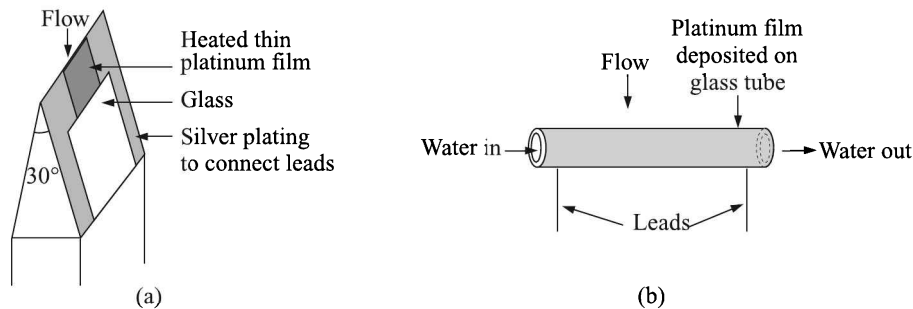


Fig. 11.22 Hot-film anemometer: (a) simple type, and (b) water-cooled type.

Calibration. Although it is possible to calculate a theoretical calibration curve for an anemometer, from the practical point of view it is far simpler to calibrate a given probe with the help of, say, a Pitot tube.

Complications arise when there is a variation of the fluid temperature either because of a steady change from calibration conditions or the dynamic temperature changes during measurement; however, various compensation techniques are available.

Example 11.9

A calibrated orifice meter is used in a pipeline of 103 mm ID to calibrate the probe of a constant temperature hot wire anemometer (CTA). The orifice meter readings are recorded in mm of Hg and the CTA readings in volts. It is independently found that the average velocity in the pipeline is exactly equal to the velocity at its axis. The volume flow rate of the fluid Q can be measured from the orifice meter calibration equation $Q = 6.311 \times 10^{-4} \sqrt{h}$ where h is in mm of Hg and Q is in m^3/s . The readings of the CTA are correlated in the form $(\text{volt})^2 = a + b(\text{velocity})^{1/2}$. Determine the constants a and b in this equation if the voltage readings are 0.284 V and 0.323 V respectively when the corresponding orifice meter readings are 77 and 154 mm Hg.

Solution

Given the relations:

$$Q = 6.311 \times 10^{-4} \sqrt{h} \quad \text{for the orifice meter}$$

$$E^2 = a + b\sqrt{v} \quad \text{for the anemometer}$$

where terms have their usual significance. The diameter of the pipeline is 103 mm = 0.103 m. So, its area of cross-section is

$$\alpha = \frac{\pi(0.103)^2}{4} = 8.3323 \times 10^{-3} \text{ m}^2$$

We know, $v = \frac{Q}{\alpha}$. From these and the supplied data, we construct the following table:

| h (mm) (supplied) | Q (m ³ /s) (calculated) | v (m/s) (calculated) | E (V) (supplied) |
|------------------------|---|---------------------------|-----------------------|
| 77 | 5.5379×10^{-3} | 0.6646 | 0.284 |
| 154 | 7.8317×10^{-3} | 0.9399 | 0.323 |

These data help us to set up the following equations for the anemometer:

$$\begin{aligned} (0.284)^2 &= a + \sqrt{0.6646b} \\ \Rightarrow a + 0.8152b - 0.0806 &= 0 \end{aligned} \quad \text{(i)}$$

$$\begin{aligned} (0.323)^2 &= a + \sqrt{0.9399b} \\ \Rightarrow a + 0.9695b - 0.1043 &= 0 \end{aligned} \quad \text{(ii)}$$

Solving Eqs. (i) and (ii) simultaneously, we get

$$a = 0.0446 \quad b = -0.1536$$

11.4 Mass Flow Measurement Type Flowmeters

The mass flow can be determined by measuring the volume flow rate and multiplying it by the density of the fluid. But a mass flow measurement type flowmeter directly measures the mass flow rate.

Coriolis Mass Flowmeter

The Coriolis mass flowmeter works by allowing a Coriolis¹⁴ force to act on a flowing fluid. The question is, what is a Coriolis force?

¹⁴Named after the French physicist Gustave-Gaspard de Coriolis (1792 – 1843), who first analysed the phenomenon mathematically.

Coriolis force

If a body moves with a translational velocity \mathbf{v} on a surface that rotates with an angular velocity $\boldsymbol{\omega}$ along a direction perpendicular to \mathbf{v} , then, according to the deduction of Coriolis, the body experiences a force given by

$$\mathbf{F}_c = 2m(\boldsymbol{\omega} \times \mathbf{v})$$

From its expression, it is clear that the force acts along the third direction in the right-handed orthogonal coordinate system.

Thus, an object moving with a constant velocity, above or on the surface of the Earth in a generally northerly or southerly direction, will be deflected in relation to the rotating Earth. The deflection is clockwise in the northern hemisphere and anticlockwise in the southern hemisphere. A few examples of the action of the Coriolis force on the natural phenomena are:

1. Circulation of the cyclonic wind in the CW-direction in the northern hemisphere and CCW-direction in the southern hemisphere
2. Circulation pattern of trade winds
3. Circulation pattern of ocean currents
4. Flight paths of missiles and rockets

Working principle

Coriolis-type flowmeters utilise this force to produce a deflection of an arm of a U-shaped flow tube as shown in Fig. 11.23(a). The flow-tube is vibrated sinusoidally by a vibrator when the

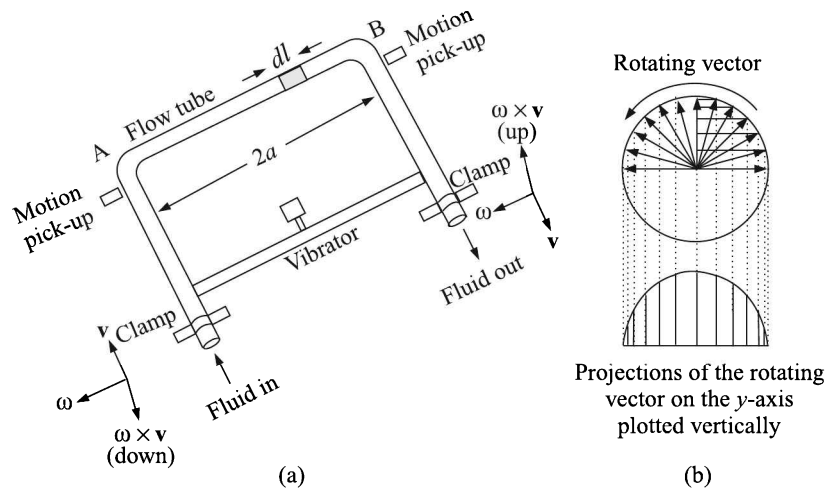


Fig. 11.23 Coriolis mass flowmeter: (a) schematic set-up, and (b) rotating vector representation of a sine-wave.

fluid flows through the tube. A sinusoidal oscillation is equivalent to a rotating vector [see Fig. 11.23(b) where only half of the projections on the y -axis of a rotating vector is shown]. So, the translational velocity \mathbf{v} of the fluid flow is affected by it in the same way as it would be by a rotational motion of the tube in the perpendicular direction. Consequently, the fluid

experiences a Coriolis force given by $2m(\boldsymbol{\omega} \times \mathbf{v})$. At any instant, the direction of this force is opposite in two arms of the U-tube because \mathbf{v} reverses direction after crossing the point B. This will set in another sinusoidal oscillation to cause a phase difference between the two motion pick-up sensors S_1 and S_2 . The time interval Δt required by the same phase to pass between the two sensor locations will be a measure of the mass flow rate G .

For commercial instruments, the sinusoidal excitation frequency typically ranges from 80 Hz to 1100 Hz.

Theory

Many theories have been proposed for the Coriolis mass flowmeter (CMF). We will present here a simple analysis proposed by Apple, Anklin and Drahm¹⁵. To simplify the analysis, only the straight part of the tube (AB in the diagram) is considered. The tube, fixed at both ends is vibrated sinusoidally by the vibrator. When the fluid in the tube is stationary, it oscillates at its fundamental frequency¹⁶ having two nodes at both ends and the vibration amplitude is maximum at the centre [see Fig. 11.24(a)]. Then, the vibration of the tube can be represented by the equation

$$m \frac{d^2 y_e}{dt^2} + ky_e = F_e \quad (11.28)$$

where m is the effective mass of the tube containing fluid

k is the stiffness of the tube

F_e is the excitation force generated by the vibrator.

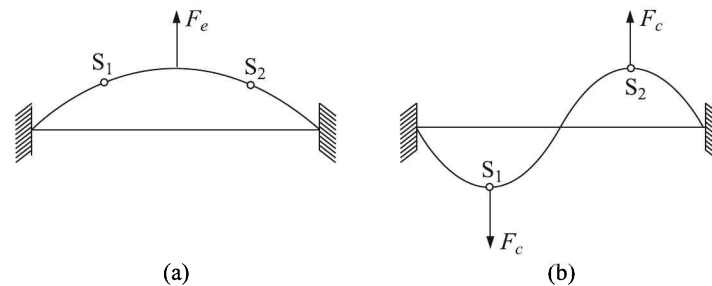


Fig. 11.24 Oscillation of AB caused by (a) F_e , and (b) F_c .

To find the fundamental frequency of vibration of the tube, we put $F_e = 0$ and solve Eq. (11.28). The trial solution is

$$y_e = a_e \sin \omega_e t$$

where a_e is the amplitude of vibration. This yields the natural angular frequency of vibration by excitation as

$$\omega_e = \sqrt{\frac{k}{m}} \quad (11.29)$$

¹⁵ *Process Measurement and Analysis*, Vol. I, B G Liptak (Ed), CRC Press (2003), p. 227ff.

¹⁶ Called *fundamental eigen* (=proper) *mode*.

The natural angular frequency of vibration is obviously a function of the density of the fluid since it incorporates k and m both of which are functions of density. Thus, *the density of the fluid can be determined by measuring ω_e .*

As the liquid starts flowing, the Coriolis force will come into play and the next harmonic of vibration will be excited because, as shown in Fig. 11.24(b), F_c acts in opposite directions on two sides from the middle of the tube. The differential equation representing excitation by Coriolis force is

$$m \frac{d^2 y_c}{dt^2} + k y_c = F_c \quad (11.30)$$

where the terms have their usual significance. The trial function for the longitudinal displacement here is

$$y_c = a_c \sin \omega_c t$$

The frequency ω_c is typically 2.7 times of ω_e , though the theoretical diagram shows it should have been double. Anyway, by putting

$$F_c = f_c \sin \omega_c t$$

we get from Eqs. (11.30) and (11.29)

$$a_c = \frac{f_c}{k[1 - (\omega_c^2/\omega_e^2)]} \quad (11.31)$$

The Coriolis force amplitude f_c is obtained by integrating the Coriolis force $2m\omega_e v$ over the length of the tube as

$$f_c = 2 \int_0^L \omega_e v (dm) = 2 \int_0^L \omega_e v (G dt) \quad (11.32)$$

where G is the mass flow rate through the tube, and $G dt = G dl/v$, dl being an elementary length of the tube and v is the velocity of fluid flow. Then, Eq. (11.32) becomes

$$f_c = 2 \int_0^L G \omega_e dl = 2G \omega_e L$$

Substituting this value in Eq. (11.31), we get

$$a_c = \frac{2G \omega_e L}{k[1 - (\omega_c^2/\omega_e^2)]} \quad (11.33)$$

The displacement of the sensors S_1 and S_2 can be obtained by superposition of the displacements caused by the excitation force and Coriolis force as

$$a_{S1} = a_e - a_c$$

$$a_{S2} = a_e + a_c$$

If $\Delta\phi$ is the phase shift between S_1 and S_2 ,

$$\begin{aligned} \Delta\phi &= \frac{2\pi}{\lambda} (\text{difference in displacement of the tube between } S_1 \text{ and } S_2) \\ &= \frac{2\pi n_e}{\lambda n_e} \cdot [(a_e + a_c) - (a_e - a_c)] \\ &= \frac{\omega_e}{v} \cdot 2a_c \end{aligned} \quad (11.34)$$

where, n_e is the frequency of vibration by excitation. So, the time interval for detection of the same phase between the two sensors is

$$\begin{aligned}\Delta t &= \frac{\Delta\phi}{\omega_e} = \frac{2a_c}{v} && \text{[from Eq. (11.34)]} \\ &= \frac{4G\omega_c L}{vk[1 - (\omega_c^2/\omega_e^2)]} && \text{[from Eq. (11.33)]}\end{aligned}$$

This yields

$$G = \frac{vk[1 - (\omega_c^2/\omega_e^2)]}{4\omega_c L} \cdot \Delta t$$

Thus, by measuring Δt , the mass flow rate can be determined. The CMFs require very accurate measurement of a very small time lag. This implies, they require to be isolated from vibration of the piping system and other environmental factors. This is best achieved by using double tubes and making them to vibrate in counter-phase. Since now both the tubes are affected by extraneous factors in the same way and one of the tubes acts as a reference for the other, it helps in eliminating the extraneous effects.

Advantages

The advantages of these flowmeters are:

1. They are suitable for all kinds of fluids—from clean ones to slurries.
2. They do not interfere with the fluid-flow.
3. They are insensitive to pressure or temperature of fluids.
4. They can be used to measure density of fluids. A recent development indicates that they can be used to measure viscosities as well (see Section 13.7).
5. They require little maintenance because they do not have any moving parts that wear out and need replacement.
6. They can measure forward as well as reverse flows with equal accuracy which is typically 0.1%.

Because of these advantages, CMFs are finding more and more acceptance in food and beverage, chemical and pharmaceutical, and oil and gas industries. A typical tube is about 37 mm long with 25 mm inside diameter having a wall thickness of 1.5 mm. A typical geometry is shown in Fig. 11.25, although other geometries, including straight tubes, are in use. Commercially available CMFs are made of stainless steel, titanium, zirconium, tantalum and some alloys.

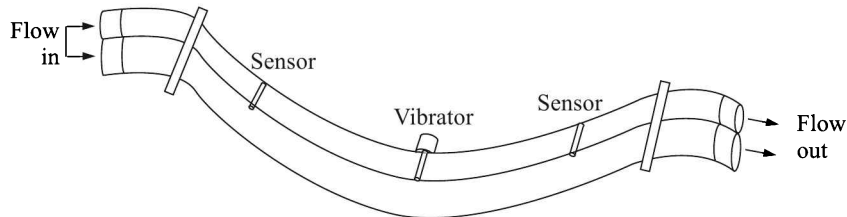


Fig. 11.25 Typical geometry of a double tube Coriolis mass flowmeter.

Thermal Mass Flowmeter

In a thermal mass flowmeter generally temperature sensors are used on both sides of a heating element (Fig. 11.26).

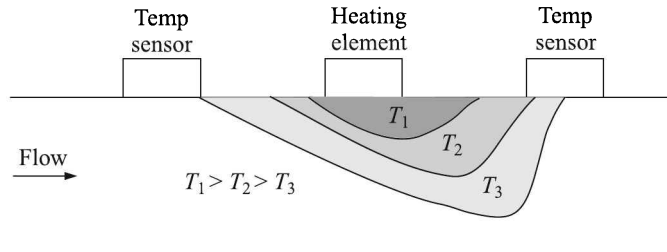


Fig. 11.26 Thermal mass flowmeter.

If the density and the specific heat of the flowing fluid remain the same, the difference of temperature between the two sensors will be proportional to the mass flow of the fluid. The fluid temperature is also measured and compensated for. It does not need any additional pressure temperature compensation over its specified range.

With the advancement of technology, microscopic thermal mass flowmeters can be manufactured as MEMS sensors. These MEMS devices can be used to measure flow rates in the range of nano litres to microlitres per minute.

Thermal-type mass flowmeters have traditionally been used to measure flow of gases such as compressed air, nitrogen, oxygen, helium, argon and natural gas, but designs for liquid flow measurements are available. The accuracy of measurement depends on the reliability of the calibrations of the actual process gas or liquid and variations in the temperature, pressure, flow rate, heat capacity and viscosity of the fluid. The electronics package of commercially available flowmeters includes the flow analyser, temperature compensator, and a signal conditioner that provides a linear output directly proportional to the mass flow.

Impact Flowmeter

The impact flowmeter measures the flow rate of free flowing bulk solids at the discharge of a material chute. The chute directs the material flow so that it impinges on an impact plate [Fig. 11.27(a)]. The impact force exerted on the plate by the material is proportional to the flow rate.

The construction is such that the sensing plate is allowed to turn about the pivoted vertical axis. The impact force is measured by sensing the deflection of the plate by a differential transformer. The voltage output of the differential transformer is converted to a pulse frequency modulated signal. This signal is transmitted as the flow rate to the control system. Typical orientations of the chute and impact plate as well as the drop height are indicated in Fig. 11.27(b).

Impact flowmeters can measure the flow rate of some bulk materials at rates from 1 to 800 tons per hour and with repeatability and linearity within 1%. They can also be used as alternatives to weighing systems to measure and control the flow of bulk solids to continuous processes.

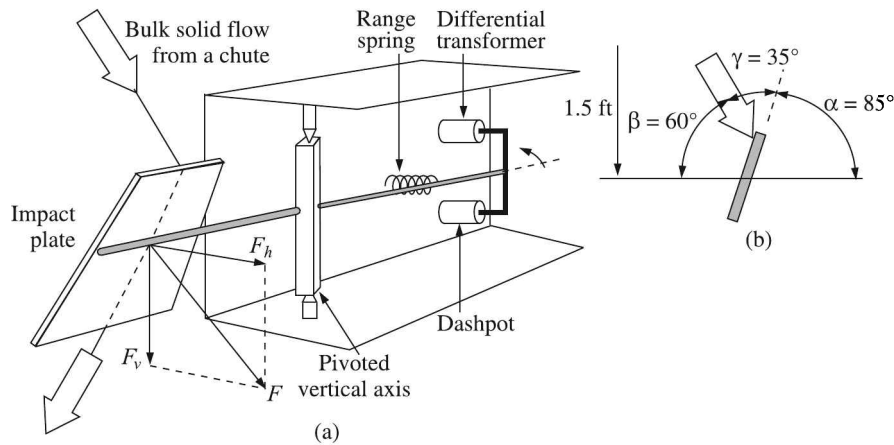


Fig. 11.27 (a) Impact flowmeter, and (b) typical orientations of the chute and impact plate, and drop height.

11.5 Positive Displacement Flowmeters

Positive displacement flowmeters are classified as direct measurement meters as opposed to inferential meters. This is because they operate on the principle of volumetric flow measurement by segmenting the flowing stream into distinct portions and providing an output directly proportional to the number of portions which pass through the meter.

In a positive displacement meter, the fluid which is being driven through the meter is forced to drive the measuring system directly. This system is usually designed as a series of vanes, rotors or pistons which displace a given amount for each revolution. The drive train is then coupled, usually through a calibrator, to either a mechanical gear head, or more commonly to a pulser which converts the rotation into a series of electrical pulses. The electrical pulses are then fed to an electronic register which converts them to a volume indication.

Positive displacement meters are available in several common designs, including the following:

1. Nutating disc
2. Sliding vane
3. Lobed impeller

Nutating Disc Flowmeter

The most common type of positive displacement flowmeter is the nutating¹⁷ disc meter. A disc placed within the confines of a boundary wall at a specific orientation is made to generate a wobbling or nutating motion of the disc. The operating cycle of one such design is shown in Fig. 11.28.

The shaded portion shows the fluid from the inlet and the arrows show the directions of pressure generated by the entering fluid at different parts of the cycle. It will be clear from these arrows how the nutation of the disc is generated.

¹⁷The word *nutating* literally means *nodding the head*, Webster Dictionary.

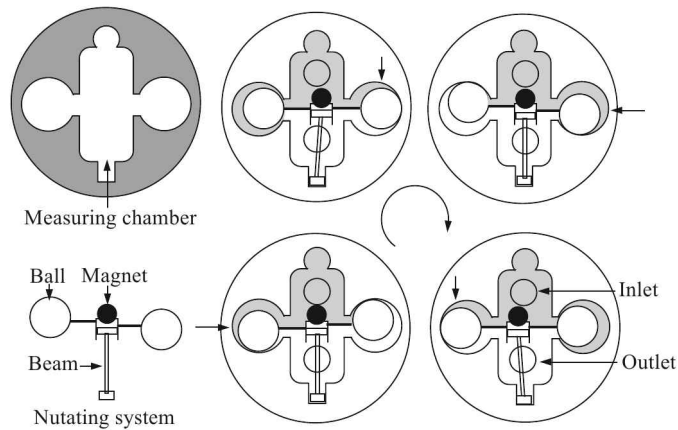


Fig. 11.28 Nutating disc type positive displacement flowmeter.

The outlet side, though not shaded, is also full of fluid all the time. But the fluid which is going out from this portion does not exert any pressure on the nutating disc. The magnet is used to track the nutation cycle from outside through a reluctance type pick-up (not shown).

The entrapment and discharge of the fluid to and from the cylindrical measuring chamber occur during the repetitive cycle of the nutating disc (the one with two balls on two sides). Each cycle is proportional to a specific quantity of flow.

This type of flowmeter is generally used in water service. As is true with all positive displacement meters, viscosity variations below a given threshold will affect measuring accuracies. Many sizes and capacities are available. The units can be made from a wide selection of construction materials.

Sliding Vane Flowmeter

Sliding vane flowmeters are probably the most accurate of all positive displacement flowmeters. Closed measuring chambers are created by means of sliding vanes that move out radially from a slotted rotor (Fig. 11.29).

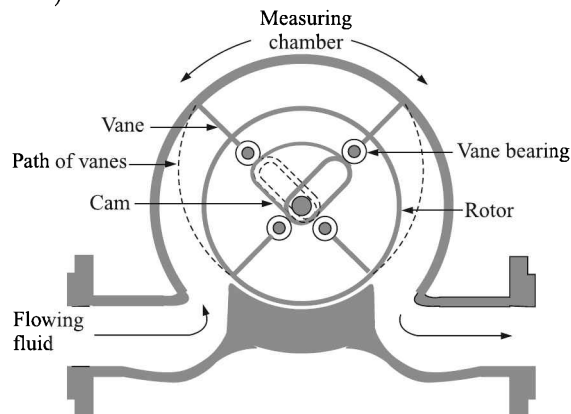


Fig. 11.29 Sliding vane type positive displacement flowmeter.

The movement of the vanes is guided by an internal stationary cam. Rollers, that are mounted on the inside edge of the vanes, follow the contour of the cam. This ensures that during transition through the measuring chamber, the vanes are in contact with the chamber wall.

The liquid impact on the vanes causes the rotor to revolve, allowing a fixed quantity of liquid to be discharged in each rotation. The number of revolutions of the rotor is a measure of the volume of liquid that passed through the meter.

Lobed Impeller Flowmeter

Lobed impeller flowmeter consists of two lobed rotors which are rotated by the passing fluid. The rotors are so lobed that their accurate intermeshing is ensured. The rotors trap the incoming fluid between them and convey it to the outlet as a result of their rotation as shown in Fig. 11.30.

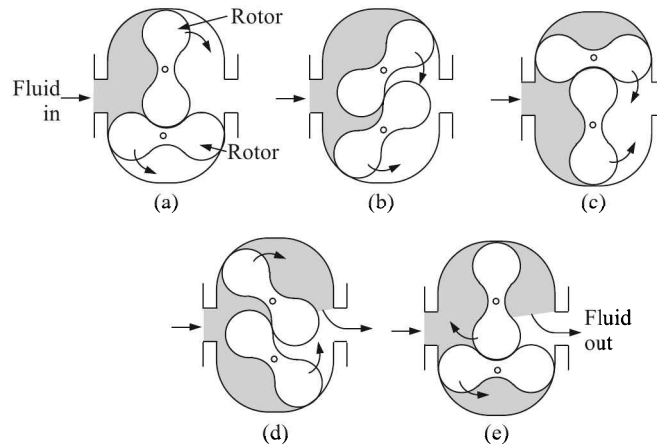


Fig. 11.30 Lobed impeller type positive displacement flowmeter.

This means that for each revolution of the lobed rotors, a specific quantity of fluid is carried through the meter. A spindle extended from one of the shafts can be used to determine the number of rotations and eventually convert it to the quantity of flow.

Advantages and Disadvantages

Positive displacement flowmeters are probably the most widely used meters in the petroleum measurement fields. They have proven to be reliable and cost effective methods for measuring volumetric flow.

They do however have some weaknesses. One of the limitations of positive displacement flowmeters is that, regardless of design, they rely on the fluid being measured to provide lubrication. This is adequate in several applications, but has proven to be a limiting factor in others, most notably the measurement of very dry products such as LPG.

They are also subject to variations in performance due to several other factors which include (i) size of the meter, (ii) fluid pressure and temperature, (iii) ambient temperature, (iv) fluid viscosity, (v) particle content of the fluid and, in some cases, (vi) installation orientation. Their accuracy varies from 0.1% to 1%.

11.6 Open Channel Flowmeters

The *open channel* refers to any conduit in which liquid flows with a free surface. Tunnels, nonpressurised sewers, partially filled pipes, canals, streams, and rivers fall in this category. Low system heads and high volumetric flow rates characterise flow in open channels.

The most common method of measuring the flow rate through open channels is by using hydraulic structures. The hydraulic structure basically changes the level of the liquid. By selecting the shape and dimensions of the structure, the flow rate through or over the restriction can be derived from a single measurement of the liquid level.

The hydraulic structures may be divided into two broad categories:

1. Weirs
2. Flumes

Weirs

The weir is a widely used hydraulic structure for measuring the open channel flow rate. The side view of a weir is shown in Fig. 11.31.

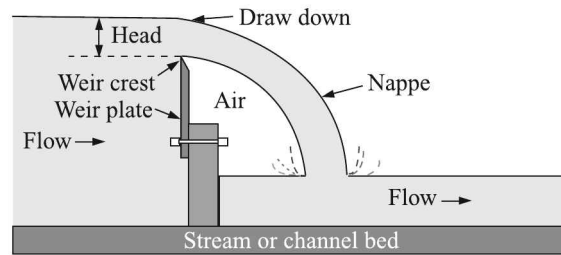


Fig. 11.31 The side view of a weir.

Basically a weir is a dam or obstruction placed in the channel so that liquid backs up behind it and then overflows. The sharp crest allows the liquid to spring clear of the weir plate and fall freely in the form of a nappe.

When the nappe discharges freely into the air, a hydraulic relationship exists between the head and the flow rate. The head can be measured with a meter stick or automatically by float-operated measuring devices. Two common weirs, rectangular and V-shaped, are shown in Fig. 11.32.

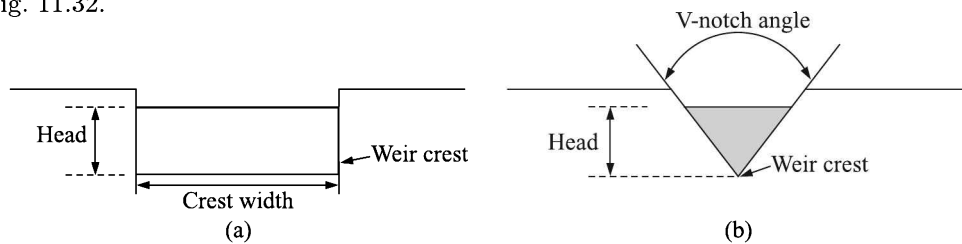


Fig. 11.32 Two common shapes of weirs: (a) rectangular, and (b) V-notch.

The formula used to make rectangular weir computations is

$$Q_R = 3.33wh^{3/2}$$

where Q_R is the flow rate, w is the width of the weir and h is the head. Rectangular weirs are commonly used for large flow rates.

V-notch weirs can have notch angles ranging from 22.5° to 90° . But the right angle notches are more common. The relevant formula for right-angled V-notch weirs is

$$Q_V = 2.5h^{5/2}$$

V-notch weirs are used to measure rather low rates of flow.

Flumes

Flumes are specially constricted sections in an open channel similar to the Venturi tube in a pressure conduit. The special shape of the flume restricts the channel area and/or changes the channel slope, resulting in an increased velocity and a change in the level of the liquid flowing through the flume. The flow rate through the flume may be determined by measuring the head on the flume at a single point. Flumes are generally used when head loss must be kept to a minimum, or if the flowing liquid contains large amounts of suspended solids. Popular flumes are the Parshall and Palmer-Bowlus designs.

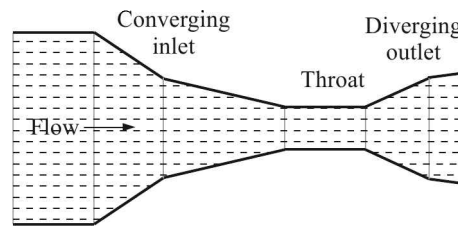


Fig. 11.33 Parshall flume: top view.

The Parshall flume (Fig. 11.33) consists of a converging upstream section, a throat, and a diverging downstream section. Flume walls are vertical and the floor of the throat is inclined downward. Head loss through Parshall flumes is lower than for other types of open channel flow measuring devices. High flow velocities help make the flume self-cleaning. Flow can be measured accurately under a wide range of conditions.

Palmer-Bowlus flumes have a trapezoidal throat of uniform cross-section and a length about equal to the diameter of the channel in which it is installed. It is comparable to a Parshall flume in accuracy and in ability to pass debris without cleaning. A principal advantage is the comparative ease with which it can be installed in existing conduits.

11.7 Turndown and Rangeability of Flowmeters

The *turndown ratio* (or *turndown*) of a flowmeter is defined as the ratio of the maximum flow that the flowmeter will measure *within the stated accuracy* to the minimum flow that can be measured *within the stated accuracy*. Thus,

$$TR = \frac{Q_{\max}}{Q_{\min}} \Big|_{\text{accuracy limit}}$$

where TR is the turndown ratio
 Q_{\max} is the maximum flow
 Q_{\min} is the minimum flow

The flow may be measured in volume for volume flowmeters and in mass for mass flowmeters. The maximum and minimum flow are stated within a specified accuracy and repeatability for the device. Suppose, an orifice meter calibrated for 0 to 100 L/min measures 20 to 100 L/min within 1% of the actual flow rate, which is its stated accuracy. Then,

$$TR = \frac{100}{20} = 5 : 1$$

Similarly, the turndown of a thermal mass flowmeter, calibrated between 0 and 12 kg/s with 0.5% accuracy, when used to measure a maximum flow of 12 kg/s and a minimum flow of 3 kg/s, can be calculated as:

$$TR = \frac{12}{3} = 4 : 1$$

Another term *rangeability* is defined as the ratio of the maximum full scale range to the minimum full scale range of the flowmeter. So, if the range of the aforesaid orifice meter is adjusted so that it can be used to measure flow from 50 L/min to 200 L/min, its rangeability will be

$$\text{Rangeability} = \frac{200}{50} = 4 : 1$$

We may note here that the range has been adjusted without caring for its calibration and stated accuracy.

Although, this subtle difference is made between turndown and rangeability by some authors, it must be mentioned here that in most of the concerned literature, the two terms are used interchangeably.

Review Questions

- 11.1 (a) Starting from Bernoulli's theorem, obtain an expression for the volume flow rate of a one-dimensional incompressible frictionless fluid flow through a horizontal pipe installed with an orifice meter.
 (b) Define: (i) vena contracta, (ii) discharge coefficient, (iii) Reynolds number.
 (c) Explain, with the help of a diagram, the principle of working of a rotameter.
- 11.2 (a) Show, by a diagram, the position of the vena contracta in a laminar flow through an orifice plate placed in a pipe, and the customary placing of pressure-tappings if the pipe diameter is ≥ 5 cm.
 (b) Explain the principle of operation of a hot-wire anemometer.
 (c) Discuss the measurement of flow velocity by a constant-current type hot-wire anemometer.
- 11.3 (a) State the principle of operation of an orifice plate flowmeter.
 (b) Discuss the construction of different kinds of orifice plates and their respective uses.
 (c) What are the disadvantages of orifice plates and how are they minimised in Venturi tubes?

- 11.4 What is the basic difference between Venturi flowmeters and rotameters? Obtain an expression of the mass-flow in terms of systems and other physical parameters. How is the compensation for density variation obtained? Give the flow-range and accuracy of rotameters.
Draw the sharp-edge configuration of the rotameter.
- 11.5 (a) Describe the principle of operation of a head-type flowmeter based on differential pressure measurement. What is Reynolds number?
(b) A Pitot tube is used on an aircraft cruising at a speed of 200 km per hour at an altitude of 3 km above the mean sea level. Calculate the differential pressure at that speed and altitude. Assume the static pressure and density at 3 km altitude to be 700 mb and 0.9 kg/m^3 , respectively.
- 11.6 What are the limitations and advantages of measurement of liquid flow by magnetic flowmeters?
Briefly describe with a diagram as to how the flow rate of liquids can be measured by magnetic flowmeters.
- 11.7 Describe the working of ultrasonic flowmeters. What are the advantages and disadvantages of this type of flowmeter?
- 11.8 (a) Describe the construction and working of turbine flowmeters. Explain how the output is obtained in digital form for both flow rate and total flow.
(b) What are the advantages and disadvantages of Venturi, orifice and flow nozzle meter?
- 11.9 (a) Explain the working principle of an electromagnetic flowmeter.
(b) "Electromagnetic flowmeter is not useful for organic fluids"—explain.
(c) What are the advantages of pulsed dc excitation for electromagnetic flowmeters?
(d) Explain why electromagnetic flowmeters are of 4-wire type instead of 2-wire type?
- 11.10 (a) Explain the difference between volume flow rate and mass flow rate.
(b) Explain the working principle of a Coriolis mass flowmeter.
(c) A flowmeter measures the flow of a liquid as follows:
(i) constant at 10 kg/min for 30 minutes
(ii) constant at 30 kg/min for 20 minutes
What would be the measurement indicated by a mass-flow totaliser for the above period of 50 minutes?
- 11.11 Indicate the correct choice:
(a) The volume flow rate is
(i) directly proportional to the cross-sectional area of pipe through which the fluid is flowing
(ii) inversely proportional to the density of the flowing liquid
(iii) directly proportional to the differential head across the restriction element
(iv) all of these

- (b) The square-root extractor is not required for
 - (i) venturi meter
 - (ii) electromagnetic flowmeter
 - (iii) rotameter
 - (iv) TC
- (c) Which of the following flowmeters works on the constant pressure drop principle?
 - (i) venturi meter
 - (ii) rotameter
 - (iii) turbine flowmeter
 - (iv) vortex flowmeter
- (d) Which of the flowmeters has the lowest pressure loss for a given range of flow?
 - (i) orifice meter
 - (ii) venturi meter
 - (iii) flow nozzle
 - (iv) rotameter
- (e) The flowmeter which cannot measure bidirectional flow is
 - (i) Coriolis mass flowmeter
 - (ii) turbine flowmeter
 - (iii) ultrasonic flowmeter
 - (iv) electromagnetic flowmeter
- (f) Amongst the flowmeters listed below, the one with the minimum straight run requirements is
 - (i) orifice plate + DP transmitter
 - (ii) Pitot tube + DP transmitter
 - (iii) ultrasonic flowmeter
 - (iv) electromagnetic flowmeter
- (g) A nozzle flowmeter has a pressure drop of 200 mm of water for a flow rate of 100 L/min. For a pressure drop of 400 mm of water, the flow rate is
 - (i) 141 L/min
 - (ii) 165 L/min
 - (iii) 200 L/min
 - (iv) 362 L/min
- (h) An example of a variable area device for measuring flow is
 - (i) flow nozzle
 - (ii) orifice meter
 - (iii) venturi meter
 - (iv) rotameter

-
- (i) A Pitot static tube is used to measure the airflow rate in a square tube. If p_d is the differential pressure, the flow rate is proportional to
- (i) $\sqrt{p_d}$
 - (ii) p_d
 - (iii) $\frac{1}{\sqrt{p_d}}$
 - (iv) p_d^2
- (j) A rotameter with a heavy float for measuring gas flow is calibrated with a gas of density 1.2 kg/m^3 . It measures the flow rate of a different gas having the density of 2 kg/m^3 and indicates a flow rate of $2.2 \text{ m}^3/\text{min}$. The actual flow rate in m^3/min is
- (i) 0.79
 - (ii) 1.32
 - (iii) 1.70
 - (iv) 2.20
- (k) In an electromagnetic blood flowmeter, the induced voltage is directly proportional to the
- (i) blood flow rate
 - (ii) square root of the blood flow rate
 - (iii) square of the blood flow rate
 - (iv) logarithm of the blood flow rate
- (l) Which of the following restrictors has the highest discharge coefficient?
- (i) orifice plate
 - (ii) flow nozzle
 - (iii) venturi tube
- (m) For flow measurement, a rotameter can be installed in a pipe line
- (i) horizontally with flow inlet in a specific direction
 - (ii) horizontally with flow rate in any direction
 - (iii) vertically with flow inlet at the bottom and the outlet at the top
 - (iv) vertically with flow inlet at the top and outlet at the bottom
- (n) An example of a positive displacement flowmeter is
- (i) orifice meter
 - (ii) rotary vane type meter
 - (iii) turbine type meter
 - (iv) ultrasonic flowmeter

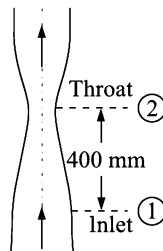
- (o) A Pitot static tube is used for measuring velocity of a gas flowing in a duct. The velocity is proportional to
- square root of the total pressure measured by the tube
 - the total pressure measured by the tube
 - difference between total and static pressures
 - square root of the difference between total and static pressures
- (p) In a rotameter, used for measuring flow rate of a fluid, if p_a is the pressure above the float, p_b is the pressure below the float, A is the area of the float, V is the volume of the float, d_1 is the density of the float material, d_2 is the density of the fluid and g is the acceleration due to gravity, then the following equation describes the equilibrium of the float:
- $(p_b - p_a)A = Vg(d_2 - d_1)$
 - $(p_a + p_b)A = Vg(d_2 + d_1)$
 - $(p_b - p_a)A = Vg(d_2 - d_1)$
 - $(p_b - p_a)A = Vg(d_2 + d_1)$
- (q) Which of the following flowmeters has the highest pressure drop for a given range of flow?
- orifice meter
 - venturi meter
 - flow nozzle
 - rotameter
- (r) For a steady flow of liquid, the float of the rotameter will remain in a particular position when
- drag force is balanced by the weight of the float and buoyancy force
 - weight of the float is balanced by the venturi meter drag force and buoyancy force
 - buoyancy force is balanced by the drag force and weight of the float
 - drag force and buoyancy together is slightly greater than the weight of the float
- (s) 'Vena contracta' is the cross-section where the flow area is a minimum for a restriction type flowmeter. For an orifice meter, if d is the diameter of the orifice opening, then the area of vena contracta is approximately
- $\frac{\pi d^2}{4}$
 - $\frac{0.99\pi d^2}{4}$
 - $\frac{0.8\pi d^2}{4}$
 - $\frac{0.6\pi d^2}{4}$

-
- 11.12 (a) Define the terms 'turn-down' and 'rangeability' in case of flowmeter.
(b) What is mass flow rate?
(c) Explain the terms 'discharge coefficient' and ' β -ratio' in case of a flowmeter.
- 11.13 What is Reynolds number? What is its role in flow calculations? How are laminar and turbulent flows distinguished by the Reynolds number?
- 11.14 State the working principle of the electromagnetic flowmeter. Describe briefly the excitation schemes of such flowmeters.
- 11.15 (a) What is the principle of operation of a hot-wire anemometer? Explain two types of hot-wire anemometer.
(b) Describe with neat sketches the working principle of a laser Doppler anemometer (LDA).
- 11.16 (a) Explain the principle of working of a transit time ultrasonic flowmeter.
(b) What is Doppler effect? Explain its use in flow measurement.
(c) In an ultrasonic flowmeter, the beat frequency is 805 cps, the angle θ between the transmitters and receivers is 45° , and the sound path is 120 mm. Calculate the fluid velocity in m/s.
- 11.17 (a) Explain the working principle of a vortex flowmeter. How is the k -factor or calibration factor of a vortex flowmeter defined? What do you understand by the 'signature curve' of such a flowmeter? What are the important advantages of using a vortex flowmeter over other types?
(b) The k -factor of a 4-bladed turbine flowmeter is 30 per gallon. What shall be the rpm of the rotor when the volumetric flow rate is 1000 gallons per min?
- 11.18 An electromagnetic flowmeter is to measure the flow rate in a pipe of diameter 5 cm. The voltage profile is symmetrical and is assumed to be uniform. The operating flux density in the liquid has a peak value of 0.1 T.
(a) Calculate the fluid velocity if the open circuit voltage between the electrodes has a peak to peak value of 0.2 mV.
(b) Calculate also the indicated fluid velocity for an output voltage of 9.4 volt peak to peak when the effluent has an impedance of 250 k Ω at the operating frequency and the output of the electrode is taken to an amplifier with gain 1000 and input impedance of 2.5 M Ω .
- 11.19 A certain pressure transducer measures the stagnation pressure (the total pressure). The density of the fluid is 1.03 g/cm³ and the velocity of flow is 100 cm/s.
(a) Calculate the dynamic pressure (pressure due to velocity of flow) in N/m².
(b) If the total pressure measured by the transducer is 10000 N/m², find the static pressure in mm of Hg.
- 11.20 An electromagnetic flowmeter is used to measure the average flow rate of a liquid in a pipe of diameter 50 mm. The flux density in the liquid has a peak value of 0.1 T. The output from the flowmeter is taken to an amplifier of gain 1000 and input impedance of

1 M Ω . If the impedance of the liquid between the electrodes is 200 k Ω , then the peak value of the amplifier output voltage for an average flow velocity of 20 mm/s is

- (a) 1.365 V (b) 1 V
(c) 0.1 V (d) 0.083 V

11.21 The accompanying figure shows a *vertical* venturi meter with upward water flow. When the measured static pressure difference ($p_1 - p_2$) between the inlet and the throat is 30 kPa, the flow rate is found to be 50 litres per second. Assume that the coefficient of discharge remains the same.



When ($p_1 - p_2$) = 20 kPa, the flow rate, in litres per second, is

- (a) 33.3 (b) 39.3
(c) 40.8 (d) 54.2

Level Measurement

Liquid level measurement constitutes an important aspect in many process industries. Depending on their operating principles, level indicators can be grouped into a few categories as listed in Table 12.1.

Table 12.1 Level indicators and their categories

| <i>Category</i> | <i>Level indicator</i> |
|--|---|
| Mechanical | Gauge glass |
| | Float type |
| | Displacer type |
| | Diaphragm type |
| | Differential pressure gauge |
| | Air bubbler |
| Optical | Laser devices |
| | Infrared and visible light source devices |
| Electrical | Resistive sensor |
| | Inductive sensor |
| | Capacitive sensor |
| Radiation (other than optical methods) | Ultrasonic type |
| | γ -ray based |
| | Radar (microwave) type |

12.1 Mechanical Level Indicators

Gauge Glass

A gauge glass (*aka sight glass*) is a simple device to determine the liquid level by attaching a transparent glass tube parallel to the liquid container.

If it is an open tank, the arrangement shown in Fig. 12.1(a) serves the purpose. In case the level of a closed tank has to be determined, the arrangement shown in Fig. 12.1(b) is suitable.

The glass tube should have a small bore and a thick wall so that it can withstand pressure. To protect it further, it needs to be encased in a metal tube having a slit opening. Valves are placed at appropriate places for the convenience of replacing a broken gauge glass without process disruption.

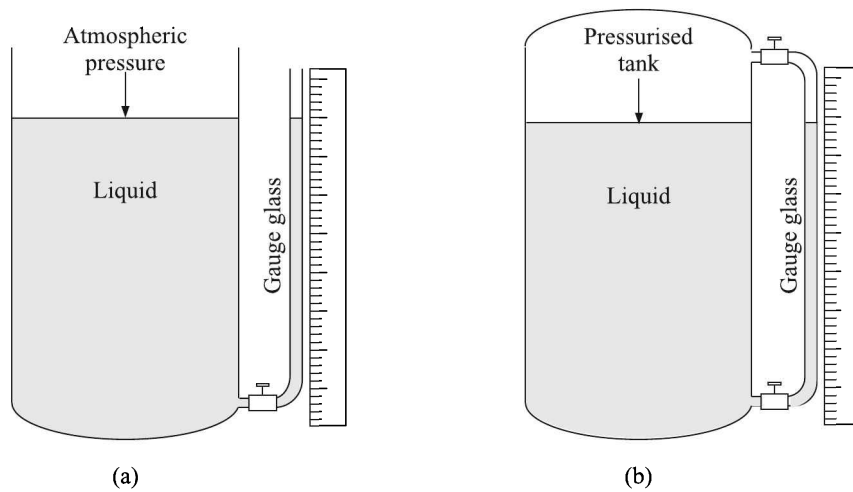


Fig. 12.1 Gauge glass: (a) tank open to atmosphere, and (b) pressurised tank.

Generally, gauge glasses are not used for level heights of more than 90 cm. For taller tanks, if necessary, two or more gauge glasses may be fixed at different heights.

The glass tubes are normally so chosen that they can withstand a steam pressure of 147 MPa (150 kg/cm², at 250°C or water pressure of 441 MPa (450 kg/cm²).

Bi-colour glass level gauge

A bi-colour glass level gauge, normally installed in boilers, is basically a gauge glass which shows red coloured portion occupied by steam and green coloured portion occupied by water [Fig. 12.2(a)].

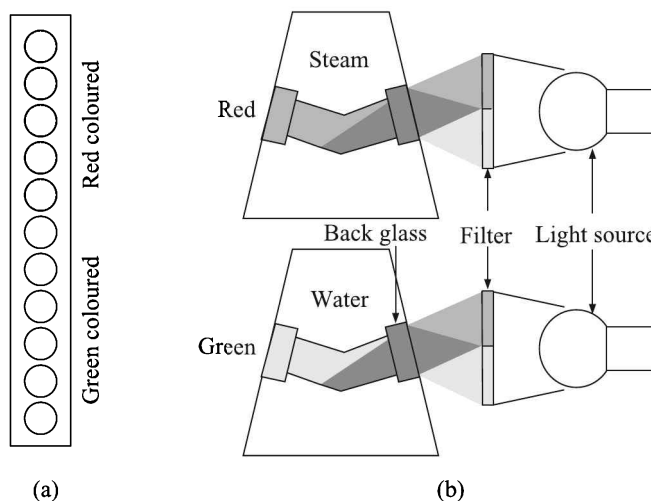


Fig. 12.2 Bi-colour glass level gauge: (a) the gauge as it appears, and (b) its working principle.

This result is obtained by exploiting the optical principle that the index of refraction is different for different colours when they pass through media like glass, water and steam.

To achieve this end, the gauge body is made to have a trapezoid section with back glasses placed on the non-parallel faces. A bi-colour light (red and green)—LED lamps, or a standard dichroic lamp with special red and green filter—is fitted on the gauge at the opposite sides of the trapezoid. This special illumination sends light obliquely through the back glasses of the level gauge that reaches the media inside [Fig. 12.2(b)].

When the gauge contains steam, green rays are considerably deviated and are prevented from emerging at the observer's side. Then only red light, which is smoothly deviated by steam, passes through the internal hole reaching the observer. Conversely, when the path of rays contains water, red rays are considerably deviated and lost inside the level gauge and green rays can reach the front glass.

Floats

A substance experiencing more buoyancy than its weight, when dipped in a liquid, floats on the surface of the liquid. So, the principle demands that the float volume, which displaces liquid, should be such that the weight of the displaced liquid is greater than the weight of the float.

Standard floats

Standard floats are spherical or cylindrical. For low density liquids the float diameter should be bigger and vice versa. Spherical float diameters vary between 75 mm and 175 mm.

Floats may be top mounted or side mounted. Movement of the float can be tracked electromechanically by fixing a potentiometer or LVDT to it [Fig. 12.3(a)] or mechanically by attaching a pointer and a scale to the float connector [Fig. 12.3(b)].

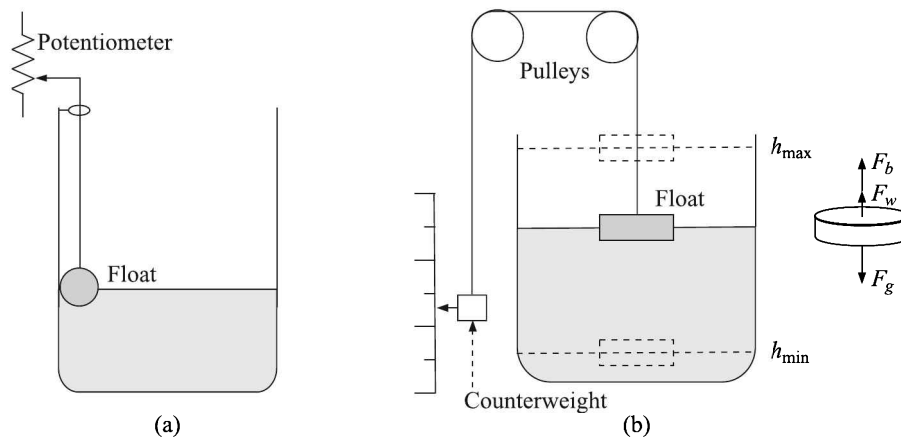


Fig. 12.3 Tracking of float movement: (a) potentiometric, and (b) mechanical.

Error due to change in density. For these devices, it is assumed that the float is immersed in the liquid by the middle of its height. Therefore, an error creeps in when the density of the liquid changes. We may calculate the error as follows.

The condition of balance of the float gives us

$$F_g - F_w - F_b = 0 \quad (12.1)$$

where, F_g is the gravitational force acting on the float

F_w is the force exerted by the counterweight, rope and friction in the pulleys

F_b is the buoyancy acting on the float

Equation (12.1) can be rewritten as

$$F_g - F_w - (\rho_l g \alpha h_l + \rho_v g \alpha h_v) = 0 \quad (12.2)$$

where, ρ_l, ρ_v are densities of the liquid and vapour

h_l, h_v are heights of the float parts in liquid and vapour

α is the area of cross-section of the float

g is the acceleration due to gravity

Now suppose, the density of the liquid has changed to $\rho_l + \Delta\rho_l$ with a consequent change $h_l - \Delta h_l$ of the depth of immersion of the float in the liquid and $h_v + \Delta h_v$, that in the vapour.

Then, Eq. (12.2) can be written as

$$F_g - F_w - (\rho_l + \Delta\rho_l)g\alpha(h_l - \Delta h_l) + \rho_v g\alpha(h_v + \Delta h_v) = 0 \quad (12.3)$$

wherein we have assumed that $\rho_v = \text{constant}$ and $(F_g - F_w) = \text{constant}$. If we now neglect the variation of the buoyancy of the float in the vapour because $\rho_v \ll \rho_l$, we get

$$\rho_v g\alpha h_v \approx \rho_v g\alpha(h_v + \Delta h_v)$$

With this condition, subtracting Eq. (12.3) from Eq. (12.2) we get

$$\begin{aligned} \rho_l g\alpha h_l &= (\rho_l + \Delta\rho_l)g\alpha(h_l - \Delta h_l) \\ \Rightarrow \rho_l h_l &= (\rho_l + \Delta\rho_l)(h_l - \Delta h_l) \end{aligned} \quad (12.4)$$

Equation (12.4) gives us the value of the error in the level measurement as

$$\Delta h_l = h_l \cdot \frac{\Delta\rho_l}{\rho_l + \Delta\rho_l}$$

Standard floats offer the advantages of

1. Simple design
2. High accuracy
3. Wide range of measuring levels
4. Possibility of level measurement in corrosive and viscous liquids

However, their disadvantage is that they cannot be used in tanks under pressure.

Magnetically coupled floats are used to indicate liquid level as well. Four such designs are shown schematically in Figs. 12.4–12.6.

Float with reed switches

An array of resistors and reed switches may be connected as shown in Fig. 12.4(a). These are generally placed about 5 mm apart in a column. The float with a permanent magnet slides along the reed switch column. Depending upon the position of the float, one of the reed switches is shorted. This sends a current through the ammeter. Obviously, according to the circuit, the higher is the position of the float, the more is the current through the ammeter. This type of level indicator indicates levels at 5 mm accuracy.

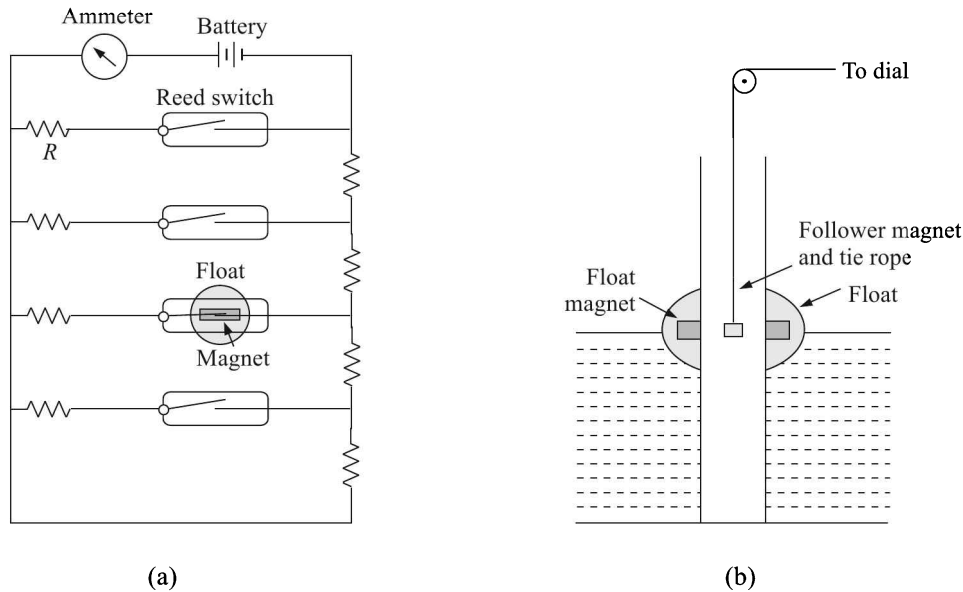


Fig. 12.4 (a) Level indication using reed switches, and (b) magnetically-coupled level indicator.

Float lifting a plunger

In another design [Fig. 12.4(b)], the magnet within the float lifts a plunger, with a magnetic tip, up or down in a guide tube according to the level of the liquid. The plunger is kept in position by a pulley arrangement.

Float rotating a magnet

Figure 12.5 shows a level indicator in which the position of the float rotates a magnet which, in turn, rotates a magnetic pointer on a dial. Here the vertical translational motion of the float is converted to rotary motion by using suitable gears and pivots. Magnetic drive, rather than mechanical drive of the pointer is resorted to in order to avoid leakage in the tank, because the gauge needs to be positioned at the middle height of the tank to cover its entire length [Fig. 12.5(b)].

Dial-type fuel level indicators are installed in motor cars. They work on the simple principle that a float arm turns a variable resistor which, in turn, changes the current through an ammeter. This ammeter is installed on the dashboard to indicate the fuel level.

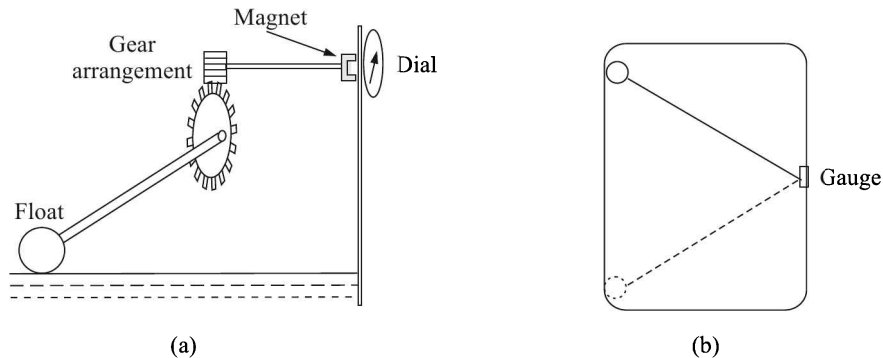


Fig. 12.5 (a) Magnetically driven level indicator, and (b) positioning the gauge.

Magnetostrictive method

The magnetostrictive method is the most elegant of all float level indicators. We have already discussed in Section 6.5 at page 215 how magnetostrictive transducers are used in measurement of displacements. Level indication is nothing but determination of the position of a substance floating on the liquid. In the magnetostrictive method, this float is a concentric circular piece of a permanent magnet.

To determine the position of the magnetic float on a liquid surface, Wiedemann and Villari effects have been utilised successfully only towards the end of the last century. A diagrammatic presentation of the process is given in Fig. 12.6.

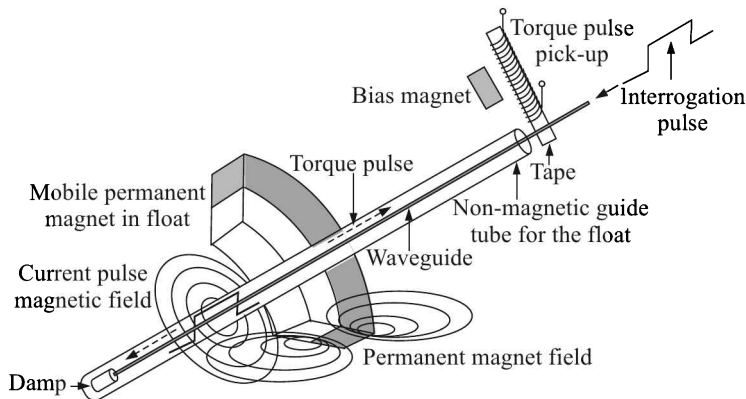


Fig. 12.6 Magnetostrictive level indication process.

To generate magnetostriction, the waveguide material has to be ferromagnetic. Naturally, therefore, a force of attraction between the waveguide and the float magnet should give rise to a frictional force, inhibiting the continuous movement of the float. This is minimised by using a waveguide of < 0.5 mm diameter and covering it with a non-magnetic stainless steel tube which guides the float movement.

An accuracy of about 0.1 mm can be achieved in level measurement by this method. It is used primarily by the pharmaceutical, food and beverage, specialty chemical, and liquid petroleum gas industries. Because many older storage tanks must be retrofitted with level

sensing devices, minimally-invasive sensors are desirable. And, of course, the fewer the process connections required, the less is the chance of leakage. There are several methods that can be used for this purpose. The comparison among them (see Table 12.2 at page 514) reveals the superiority of the magnetostriction method.

Displacers

To measure the level of a liquid in a tank, a *displacer* or buoy is partially immersed in the liquid [Fig. 12.7(a)]. In equilibrium, the weight $W - \Delta W$ measured by the spring balance is

$$W - \Delta W = F_g - F_b \quad (12.5)$$

where W is the weight of the buoy when not dipped in the liquid
 ΔW is the loss of weight of the buoy when dipped in the liquid
 F_g is the gravitational force acting downwards on the buoy
 F_b is the buoyancy acting on the buoy

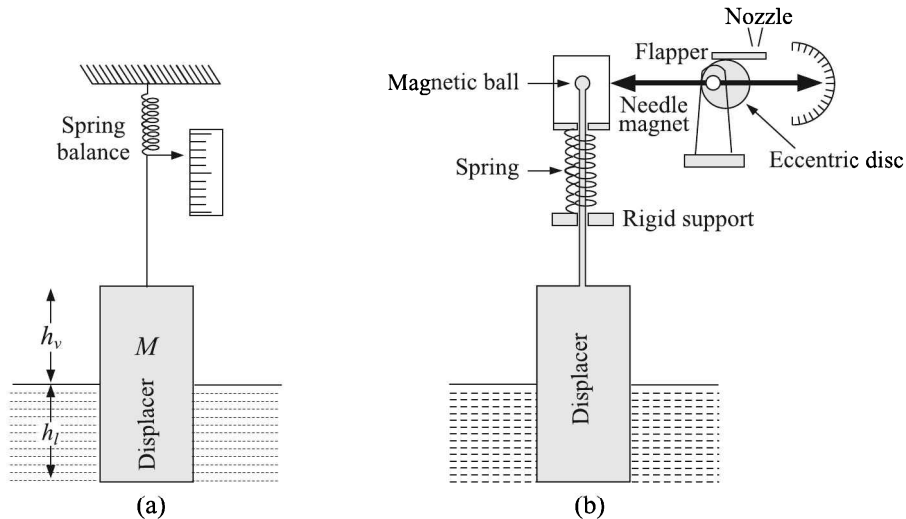


Fig. 12.7 (a) Displacer principle, and (b) spring-balance displacer.

We know from Archimedes' principle that the buoyancy equals the weight of the displaced liquid and vapour by the displacer or buoy. So,

$$F_b = \rho_l g \alpha h_l + \rho_v g \alpha h_v$$

where, ρ_l, ρ_v are densities of liquid and vapour
 h_l, h_v are parts of displacer length in the liquid and in the vapour above it
 g is the acceleration due to gravity
 α is the area of cross-section of the displacer

Therefore, Eq. (12.5) can be written as

$$W - \Delta W = Mg - (\rho_l g \alpha h_l + \rho_v g \alpha h_v) \quad (12.6)$$

Had it not been dipped in the liquid, the resultant force measured by the spring balance were

$$W = Mg - \rho_v g \alpha (h_l + h_v) \quad (12.7)$$

Subtracting Eq. (12.6) from Eq. (12.7), we get for the loss of weight of the displacer owing to its partial immersion in the liquid is

$$\Delta W = g \alpha h_l (\rho_l + \rho_v) \quad (12.8)$$

Since, g , α , ρ_l , ρ_v are constants for a set-up, we get from Eq. (12.8) that

$$\Delta W \propto h_l$$

This principle is utilised in the construction of displacer-type liquid level indicators. The displacer element, buoy, is a cylinder of constant area of cross-section and its density should be greater than that of the liquid.

Spring-balance displacer

In this instrument [Fig 12.7(b)], the up and down displacer movement, owing its origin to the variation of liquid level, causes the attached spring to contract or expand. The displacer tie rod ends in a magnetic ball. The resulting up and down movement of the magnetic ball is sensed by a needle magnet which is fixed on a pivot outside the ball housing.

The movement of the magnetic ball is around 25 mm. This can be pneumatically magnified by attaching the flapper of a nozzle-flapper transducer to a disc which is attached eccentrically to the pivot of the magnetic needle. Alternatively, the movement may be converted to an electric signal by a potentiometric arrangement.

In industrial situation, the practical problem is to seal the process from the spring-balance or other detection systems. We will consider one such mechanism which can do that.

Torque-tube displacer

In this method, shown schematically in Fig. 12.8, the displacer movement applies torsion to a tube, called the torque-tube. The hollow torque-tube consists of an inner torque-rod which is welded to the torque-tube at one end and free at the other end, supported by a frictionless bearing. The torque-tube ends in a knife-edge at one side and supports the displacer via the torque arm which ends in a block. The other end of the torque-tube ends in a flange which is anchored at the tank wall.

With this arrangement, when the displacer moves up or down, it applies torsion to the torque-tube via its knife-edge. This torsion is transmitted to the inner torsion-rod which carries it outside the tank. The angular displacement of the rod is about 5° to 6° and it is linearly related to displacer's apparent weight which, in turn, is related to the liquid level.

The little angular displacement of the torque-rod can be pneumatically amplified to a large pressure differential by driving the flapper of a nozzle-flapper transducer. The pressure differential can be calibrated to indicate liquid level.

Usually 0.3 to 1.5 m long displacers are used though the length can be as big as 18 m. Nickel, inconel, monel, hastelloy, etc. are used to construct torque-tubes.

Displacers are suitable for level measurement of clean liquids. Slurries are likely to upset their calibration by depositing solid particles on them.

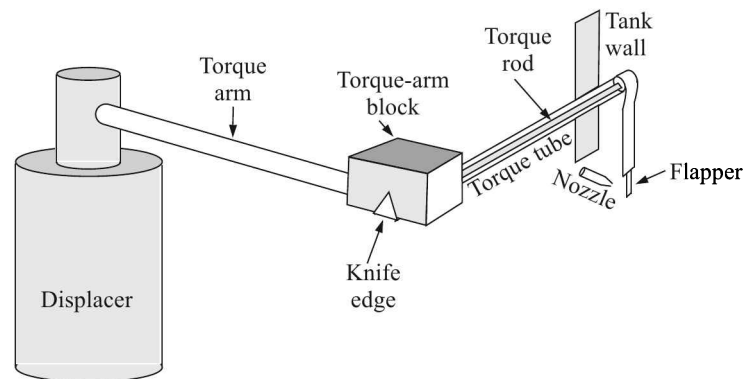


Fig. 12.8 Schematic diagram of torque-tube displacer.

Diaphragm Level Indicators

Diaphragm level indicators consist of a box closed on all sides except one where a flexible diaphragm is fixed [Fig. 12.9(a)]. The box contains captive air which is connected to a pressure detector via a capillary tubing. The diaphragm is normally made of neoprene or silicone rubber, Teflon or similar plastic material.

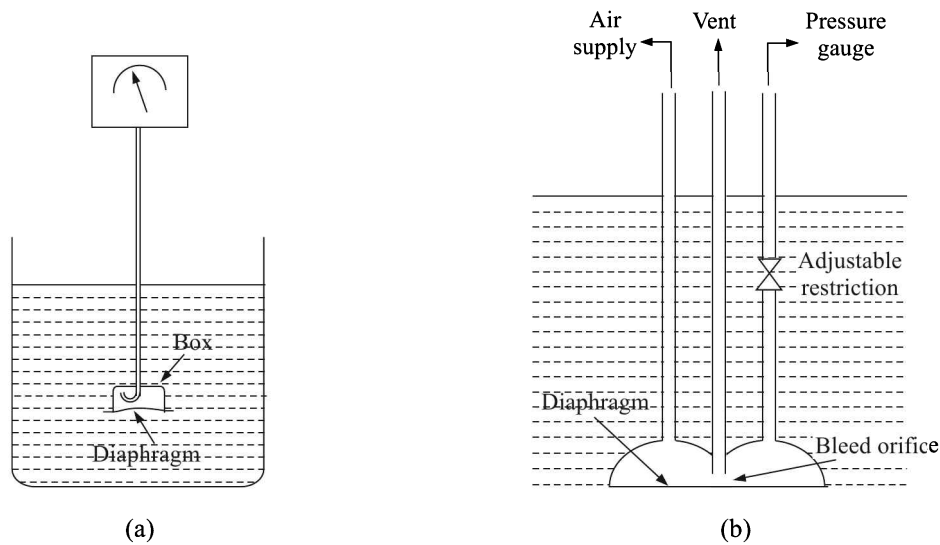


Fig. 12.9 Diaphragm level indicator: (a) simple, and (b) continuous air-supply type.

The diaphragm box is kept submerged in the liquid. As the liquid level rises, the static head of the liquid exerts an upward force on the diaphragm which, in turn, compresses the captive air. The captive air pressure is thus proportional to the liquid level. This type of level indicator, which can be used in open to atmosphere type vessels, is less costly, though its accuracy is limited.

In an improved version, the air inside the diaphragm is not captive but a continuous supply is maintained through a pipe [see Fig. 12.9(b)].

A vent pipe allows the air to bleed to the atmosphere through a bleed orifice existing between the vent pipe and the diaphragm. Another pipe connects the diaphragm to a suitable level indicator which is basically a pressure indicator. Air supply to the unit is regulated to about 3 to 5 psig (0.2 to 0.3 bar) in excess of the maximum hydraulic head to be measured. Stainless steel diaphragms are suitable for this type of level detectors. As the liquid level rises, increased pressure acting on the diaphragm makes it move upward making the bleed orifice smaller. Consequently, less air leaks through the vent tube causing the air pressure to build up. The built-up air pressure then pushes the diaphragm down increasing the air leakage and so on till equilibrium is reached. The air pressure within the diaphragm enclosure is a measure of the liquid level. These indicators are accurate to within 5 psig (0.3 bar) of the air-supply pressure. They can operate up to 160 psig (11 bar). The adjustable restriction can be suitably manipulated to increase the speed of response.

Differential Pressure Level Indicators

A liquid level exerts pressure caused by the weight of the liquid column. This pressure can be measured to figure out the liquid level, provided the liquid is at atmospheric pressure. This method, known as *hydrostatic tank gauging* (HTG), is often resorted to in industries. But if the liquid is in a pressurised tank, then one needs to measure the differential pressure¹ between the top and bottom of the liquid column to figure out the liquid level.

If, p_1 is the pressure at the bottom of the tank [Fig. 12.10(a)]

p_2 is the pressure at an intermediate point

p_3 is the pressure at the top of the tank

h is the difference in height between p_1 and p_2 tapping points, and

l is the height of the liquid level in the tank

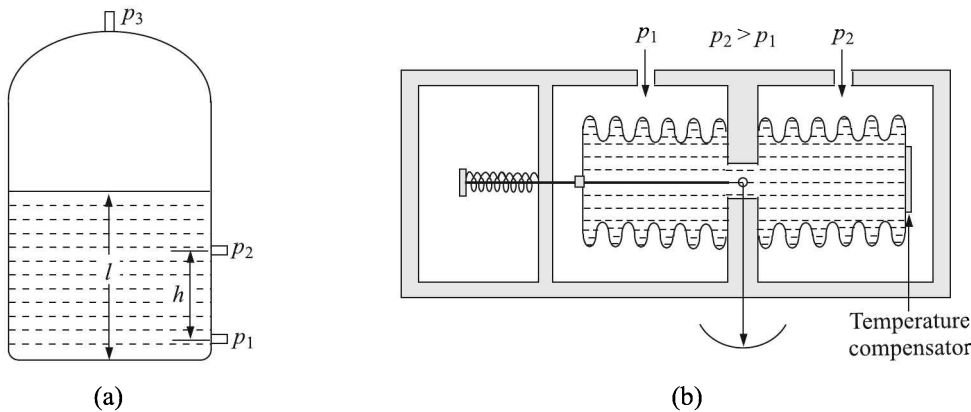


Fig. 12.10 (a) Hydrostatic tank gauging, and (b) differential pressure level indicator.

then

$$\rho = \frac{p_1 - p_2}{hg}$$

and

$$l = \frac{p_1 - p_3}{\rho g}$$

¹Often written as d/p or DP.

The pressure difference between the top and bottom of the liquid level can be measured by individually measuring the two pressures. But, in that case, the error in the two measurements, which may not be the same, may add up to give an erroneous result. Therefore, a single measurement which gives the differential pressure value is always preferred.

The differential pressure can be measured mechanically by using bellows in an enclosure where the higher pressure side may be connected to the bellows and the lower pressure side to the enclosure (see Fig. 8.10). The consequent displacement of the tip of the bellows, measured by a variable reluctance pick-up, may be calibrated to give the height of the liquid level.

Another arrangement [Fig. 12.10(b)] with liquid-filled bellows is used to measure d/p. In this arrangement, the higher pressure side pushes the liquid inside the bellows to the lower pressure side. This expands the bellows there. As a result the spring-loaded pointer lever is pushed back, moving the pointer to the right. Since d/p becomes sensitive to the temperature of the liquid, a bimetallic temperature compensator may be attached to the bellows to offset the pressure differential produced by the temperature differential.

Diaphragms with variable reluctance-type sensing (see Fig. 8.17) or capacitive sensing (see Fig. 8.18) are also used for precise d/p measurements.

In the context of differential pressure measurement or transmission, it is relevant to discuss what are called *zero suppression* and *zero elevation* of signals.

Zero suppression

Sometimes it may be necessary to mount the level transmitter at a distance x below the base of the open tank as shown in Fig. 12.11(a).

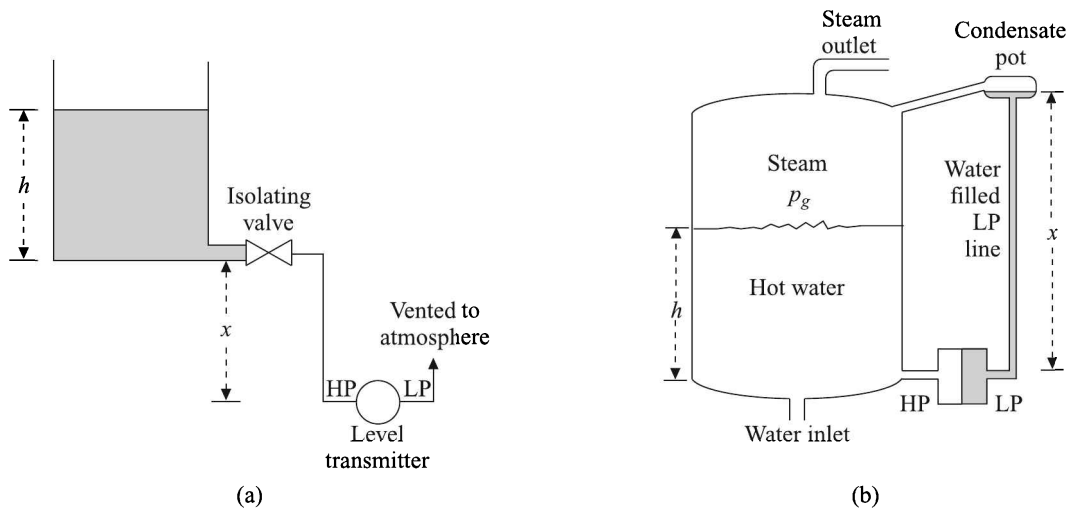


Fig. 12.11 Level transmitter: (a) where zero suppression is necessary, and (b) where zero elevation is necessary. HP and LP indicate high pressure and low pressure sides respectively.

Then, the liquid on the high pressure side exerts an excess pressure of xs where $s (= \rho g)$ is the specific weight of the liquid. If the liquid level is at a height of h , the situation will be as given by the following equations:

$$p_{\text{high}} = sh + sx + p_{\text{atm}}$$

$$p_{\text{low}} = p_{\text{atm}}$$

$$\therefore \Delta p = p_{\text{high}} - p_{\text{low}} = sh + sx \quad (12.9)$$

It is clear from Eq. (12.9) that the transmitter has to be biased $-sx$ so as to make the transmitter reading proportional to h . This negative biasing is called the *zero suppression* of the signal.

Zero elevation

Suppose, we want to measure the water level in a boiler. For this purpose, a d/p cell with a diaphragm is installed and it is connected from the bottom of the boiler to the top through a condensate pot. The arrangement is referred to as *wet leg installation*. Figure 12.11(b) shows a simplified wet leg installation.

If the height of water level in the boiler is h and $s (= \rho g)$ is the specific weight of water, we have the following equations:

$$p_{\text{high}} = p_g + hs$$

$$p_{\text{low}} = p_g + xs$$

$$\therefore \Delta p = p_{\text{high}} - p_{\text{low}} = -s(x - h) \quad (12.10)$$

Since $x > h$, the differential pressure sensed by the transmitter is always a negative quantity. To properly calibrate the transmitter, a positive bias sx is needed. This positive biasing technique is called the *zero elevation*.

Air Bubblers

A rather simple and age-old method of measuring liquid level is immersing a dip tube inside the tank, bubbling air (or any other suitable gas) through it at a constant flow rate and then measuring the pressure of the purge gas (Fig. 12.12).

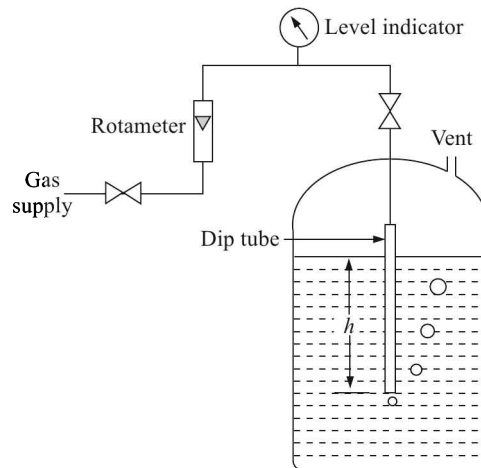


Fig. 12.12 Air bubbler level indicator.

Clearly, the pressure readout p of the purge gas is given by

$$p = h\rho g - p' \quad (12.11)$$

where h is the height of the liquid level
 ρ is the density of the liquid
 g is the acceleration due to gravity
 p' is the dynamic pressure drop due to the flow of the gas (Bernoulli's theorem).

If p' is kept constant by maintaining a steady flow rate, the readout pressure will be linearly proportional to the height of the liquid level. For this purpose, a rotameter is included in the gas supply line.

The most commonly used purge gases are air and nitrogen. Depending on the situation and considering the chemical reactivity of the liquid, other non-hazardous gases can also be used. The pressure of the purge gas supply should be at least 10 psi (68.9 kPa) higher than the maximum liquid head. The flow rate of the purge gas is kept at around 1 ft³/h (~ 500 cm³/min) to minimise the dynamic pressure loss.

Because of its simplicity, the air bubbler level indicator is not only inexpensive but also easy to maintain. It can indicate liquid level correctly even if there is foam formation in the liquid. But density variation of the liquid owing to temperature variation will upset the calibration as is evident from Eq. (12.11).

12.2 Optical Level Indicators

Laser Sensors

Laser devices of level sensing can furnish accurate data under difficult conditions. Which is why they are finding more and more applications in industries in recent times. Generally, pulsed lasers are used in level sensing.

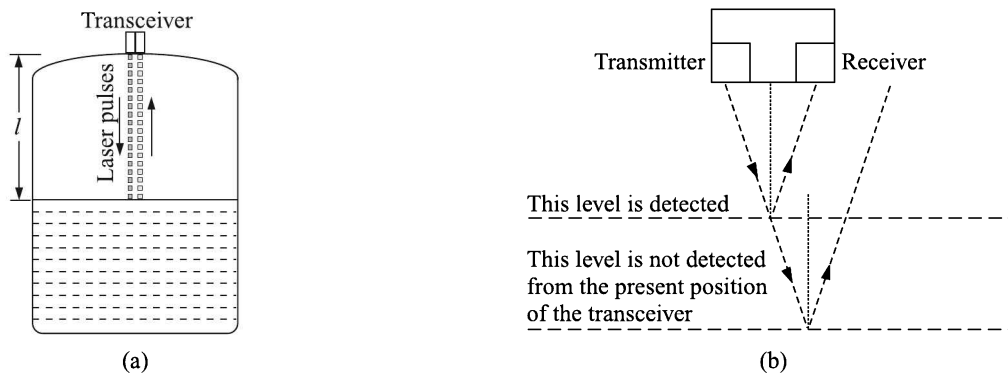


Fig. 12.13 (a) Laser level indicator, and (b) reflection-based level indicator.

A pulsed laser device sends a pulse of light which gets reflected on the liquid surface and returns [Fig. 12.13(a)], with a diminished intensity, after a certain time Δt , say. Then, if l is the liquid level measured from the top of the tank, we have

$$2l = c\Delta t$$

$$\Rightarrow l = \frac{c\Delta t}{2}$$

where, c = velocity of light = 3×10^8 m/s. Obviously Δt will be very small. But its accurate determination has been made possible by modern electronics. Suitable frequency laser that can penetrate dust and liquid vapour, may be chosen for a particular process.

The advantages of pulsed laser sensors are many. They are non-intrusive instruments that can be mounted externally on sight glasses. Therefore, their repair and maintenance can be done without disrupting the process. Secondly, since they are truly remote sensors, they can be used in corrosives and acids for level monitoring.

Infrared and Visible Light Sensors

These sensors can be of three types—reflection-based, refraction-based and optical fibre-based.

Reflection-based level indicator

Light from a source gets reflected on the surface of the liquid (or solid) level and the reflected light is detected by a sensor. This detection can be based on the time-of-flight measurement as discussed in the previous section. Alternatively, the measurement of the angle between the incident and the reflected beams will yield the desired distance through simple trigonometry. In the third method [Fig. 12.13(b)], the source emits light at a fixed angle of incidence and the sensor detects it only when the angle of reflection equals the angle of incidence.

In this method, the probe is moved up or down over measured distances to detect the liquid level. Available commercially, such sensor switches can be adjusted to detect liquid level between 0.25 in (6 mm) and 1 ft (305 mm) within a temperature range of -40°C to 60°C .

Refraction-based level indicator

In these devices, IR or visible light is sent to a tiny prism at the tip of the probe. As long as the tip is in the air, the prism reflects the light back [Fig. 12.14(a)] by total internal reflection. The moment the tip touches the liquid surface, the reflected light intensity goes down owing to refraction through the liquid [Fig. 12.14(b)].

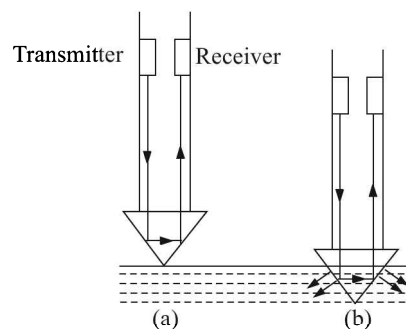


Fig. 12.14 Refraction-based level indicator: (a) when the tip does not touch the liquid, and (b) when the tip touches the level.

So, the measurement of the vertical displacement of the probe gives the value of the liquid level. It is relatively simple to construct a switch, based on this principle, for controlling the liquid level. Commercially available switches can control liquid levels within 1/16 inch at a temperature range of -90°C to 120°C . However, the sensor is prone to errors if the prism gets wet by the liquid and drops of liquid cling to it even when it is not in contact with the liquid.

Optical fibre-based level indicator

In this method (Fig. 12.15), light of fixed intensity is sent through an unclad optical fibre to a sensor.

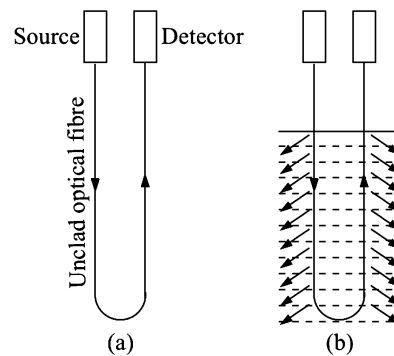


Fig. 12.15 Optical fibre-based level indicator: (a) when the fibre is in air, and (b) light dissipation when the fibre is in contact with the liquid.

The fibre is kept immersed in the liquid tank. If there is no liquid in the tank, the intensity of light at the source and the detector are almost the same. As the liquid level rises, some light dissipates through the liquid, bringing the intensity down at the detector. So, the difference between intensities of light at the source and the detector can be calibrated against the liquid level. Of course, this calibration will not be the same for all liquids. Secondly, the measurement system is prone to error owing to the wetting of the fibre as discussed in the case of refraction-based indicator.

12.3 Electrical Level Indicators

Electrical level indicators are generally of three kinds—(i) resistive, (ii) inductive, and (iii) capacitive.

Resistive Level Indicators

Resistive level indicators can be of discrete step-type or continuous type.

Discrete step-type level indicator

Individual wire-wound or carbon-type resistors are placed inside the tank containing the liquid. Each resistor forms one arm of a bridge circuit containing a voltage source and a voltmeter [Fig. 12.16(a)].

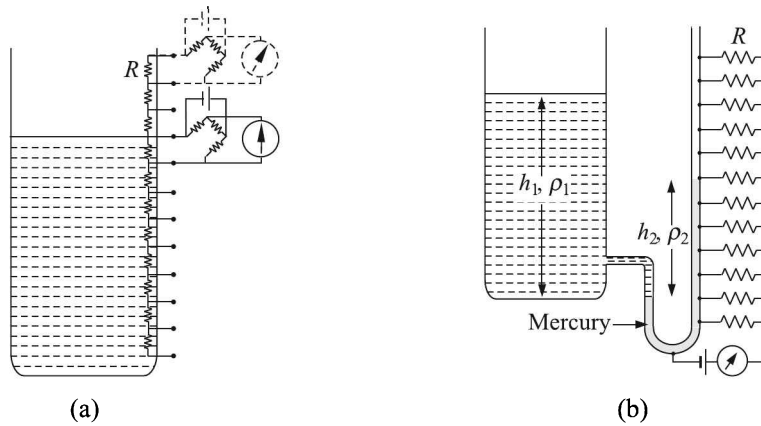


Fig. 12.16 Discrete step-type resistive level indicator: (a) one design, and (b) another design with a mercury manometer.

A small current passes through the resistor which is immersed in the liquid. The resistor is raised to a certain temperature and therefore acquires a certain value of the resistance. The bridge is balanced at that value, showing a null voltage in the voltmeter. The moment liquid level falls below the resistor, its temperature changes because of a change in its heat transfer rate. The resulting change of resistance offsets the bridge balance, showing a voltage in the corresponding voltmeter. In this way, level, in discrete steps may be detected. But since this method depends on heat transfer rates, ambient temperature variation may cause error in level indication.

Another discrete-type level detector utilises a mercury manometer attached to the tank. Resistors of fixed value are connected to the manometer tube at certain intervals [Fig. 12.16(b)].

These resistors, connected in parallel, form part of an electrical circuit having a voltage source and an ammeter. If the mercury level rises to the n th resistor, the effective resistance of the electrical circuit will be

$$R_{\text{eff}} = \frac{R}{n}$$

Consequently, the current through the ammeter will be

$$I = \frac{Vn}{R}$$

As the liquid level in the tank goes up, the mercury level in the U-tube also goes up, causing the ammeter to show more current value. The current indicated by the ammeter is thus related to the mercury level as

$$h_2 = KI$$

where K is a constant. h_2 is, however, related to the liquid level h_1 in the tank as

$$h_1 \rho_1 = h_2 \rho_2$$

where ρ_1 and ρ_2 are densities of the liquid and mercury respectively. Thus,

$$h_1 = \frac{h_2 \rho_2}{\rho_1} = \frac{\rho_2}{\rho_1} KI = K' I \quad (12.12)$$

Eq. (12.12) shows that the calibration of liquid level vs. current is liquid-specific.

Continuous-type level indicator

Monitoring liquid level almost without steps by the resistive method can be achieved with the help of resistance tapes. These specially constructed tapes incorporate a helically wound gold-plated nichrome resistance wire and a base plate covered by an elastic sheath [Fig. 12.17(a)].

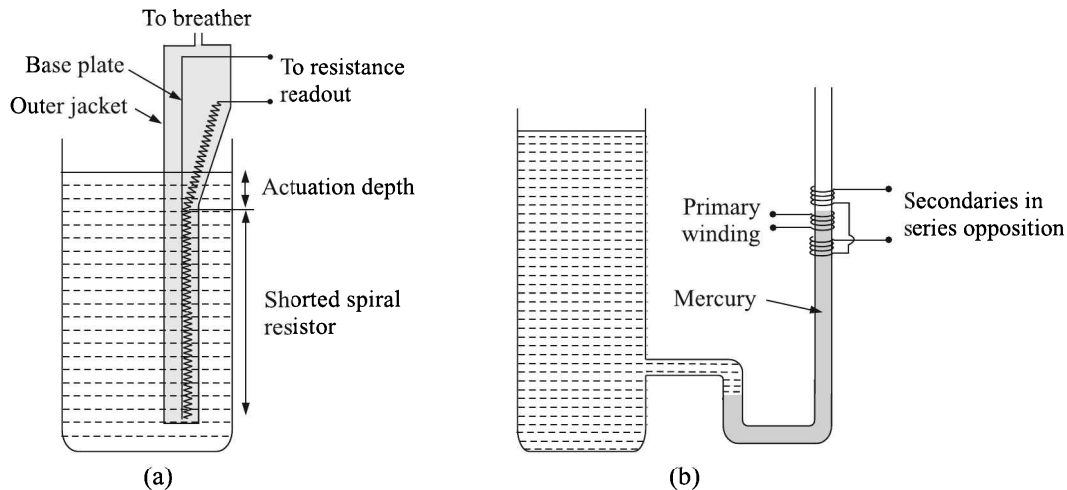


Fig. 12.17 (a) Continuous-type resistive level indicator, and (b) inductive level detector.

As the tape is dipped in a liquid, a portion of the helical resistor gets shorted to the base plate owing to the liquid pressure acting on it. As a result, the resistance of the tape becomes proportional to the liquid level, measured from the top.

However, it is to be noted that a threshold pressure is necessary to cause the shorting. This threshold pressure is attained below the *actuation depth* which will depend on the density of the liquid. For commercially available tapes, which may be 1 to 30 m long, the actuation depth for water is about 10 cm. Obviously, this detector cannot detect levels less than the actuation depth.

It should be noted that an arrangement has to be made to let air in or out (i.e. breathing) of the elastic sheath as the liquid level varies. During this breathing process, moist air may enter the sheath, cause condensation of water inside and thus make the tape ineffective. To avoid this possibility, a suitable moisture trap has to be provided to the breathing hole of the tape.

Resistance tapes can measure levels up to 3 mm precision for liquids including slurries but generally not solids. In addition to level, they are capable of measuring temperature between -30°C and 100°C .

Inductive Level Indicator

If primary and secondary coils are wound on a non-magnetic side-tube of the tank [Fig. 12.17(b)] which contains a liquid having good magnetic permeability (e.g. liquid metals), the arrangement resembles an LVDT² in appearance.

²See Section 6.2 at page 173.

Consequently, by feeding a sinusoidal voltage to the primary and through phase-sensitive detection of the voltage across the secondary coils, connected in series opposition, the arrangement works like an LVDT, offering an accurate measurement of the liquid level.

Capacitive Level Indicator

Capacitance variation with the variation of ϵ is utilised in various ways for level sensing. In the arrangement shown in Fig. 12.18(a), the terminal A is connected to a cylindrical electrode of diameter $2r$ and B is connected to the wall of an outer hollow cylindrical container which forms the second electrode. The inner cylinder is fixed to the cylindrical container by an insulator stopper. The cell thus formed may be dipped in the tank, the liquid level of which is to be measured. The liquid of the tank enters the cell through pores on the walls of the cell.

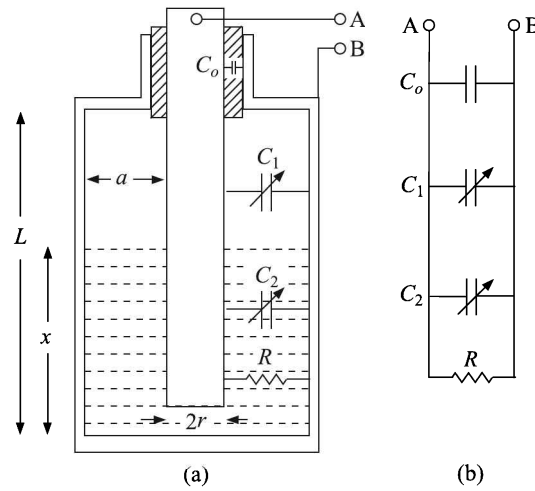


Fig. 12.18 Capacitor transducer for liquid level sensing: (a) the cell, and (b) the equivalent electrical circuit.

Electrically, the arrangement is like that in Fig. 12.18(b), with three parallel capacitors and a resistor. Of the capacitors, C_1 and C_2 are variable because of the variation of the liquid level.

We assume (i) the liquid is non-conducting so that $R \rightarrow \infty$, and (ii) the stopper insulator is a very poor dielectric, so that $C_0 \rightarrow 0$. Then the capacitance between A and B can be written as

$$\begin{aligned} C = C_1 + C_2 &= \frac{2\pi}{\ln[(r+a)/r]} \epsilon_0 (L-x) + \frac{2\pi}{\ln[(r+a)/r]} \epsilon_0 \epsilon_r x \\ &= \frac{2\pi}{\ln[(r+a)/r]} [\epsilon_0 L + x \epsilon_0 (\epsilon_r - 1)] \end{aligned}$$

which gives

$$\begin{aligned} x &= C \frac{\ln[1 + (a/r)]}{2\pi \epsilon_0 (\epsilon_r - 1)} - \frac{L}{\epsilon_r - 1} \\ &\equiv K_1 C + K_2 \end{aligned} \tag{12.13}$$

where K_1 and K_2 are constants of the cell. Thus, the transfer characteristic between the liquid level and the measured capacitance is linear.

This kind of liquid level sensing has not only been used for common liquids but also for powdered or granular solids, liquid metals, liquefied gases, corrosive liquids such as hydrofluoric acid and in very-high-pressure processes with suitable modifications of the cell structure.

Conducting liquids. However, for conducting liquids the entire length of the probe has to be insulated in order to prevent short-circuiting of the capacitor by the liquid (see Fig. 12.19).

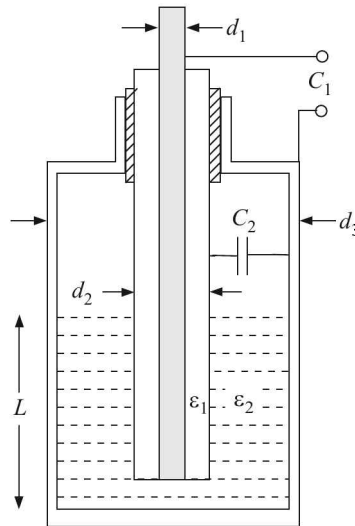


Fig. 12.19 Capacitive level indicator for conducting liquids.

For such a rotationally symmetrical configuration, the capacitance C of an insulated electrode changes with level L according to:

$$C = \frac{2\pi\epsilon_0 L}{\frac{1}{\epsilon_1} \ln \frac{d_2}{d_1} + \frac{1}{\epsilon_2} \ln \frac{d_3}{d_2}}$$

$$\Rightarrow L = C \cdot \frac{\frac{1}{\epsilon_1} \ln \frac{d_2}{d_1} + \frac{1}{\epsilon_2} \ln \frac{d_3}{d_2}}{2\pi\epsilon_0} \quad (12.14)$$

where ϵ_1 and ϵ_2 are the relative permittivities of the insulation material and the liquid, respectively.

In case the liquid is highly conducting, $\epsilon_2 \gg \epsilon_1$ and so, the second term on the numerator of Eq. (12.14) becomes much smaller than the first term. Neglecting the second term, the relation yields:

$$C = \frac{2\pi\epsilon_0\epsilon_1}{\ln(d_2/d_1)} \cdot L \quad (12.15)$$

Equation (12.15) shows that even in the case of the insulated probe, the capacitance linearly varies with the level height.

When arranged horizontally, a capacitive sensor can act as a level switch as well.

12.4 Radiative, Other Than Optical, Methods

Ultrasonic Level Indicator

Generally sound waves of frequency 20 kHz and above are termed ultrasonic sound. Continuous ultrasonic sound may be generated through inverse piezoelectric effect, impinged upon a liquid surface either from the top or bottom of the tank and the reflected beam may be detected by piezoelectric transducer. Measurement of the angles of incidence and reflection together with knowledge of the distance between the transmitter and the detector may yield the desired value of the level. But sonic reflection being diffuse in nature, this method is not suitable.

The time-of-flight method offers a better alternative (Fig. 12.20). An ultrasonic pulse is emitted from the transmitter and the time its echo on the liquid surface takes to return is measured. In case of top mounting, the pulse travels through air with a velocity of 331 m/s at 0°C. But, in case the transmitter is mounted at the bottom, the velocity of travel becomes liquid-specific. For water at 25°C, it is 1505 m/s.

Instead of packaging transmitter and detector separately, both can be housed together and the detector may be electronically blanked off while sending the pulse.

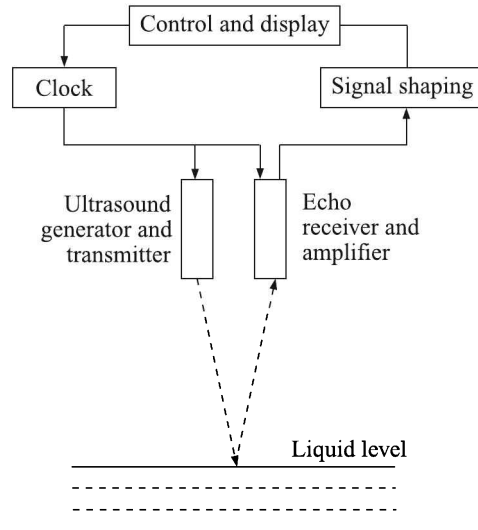


Fig. 12.20 Time of flight method of level detection with ultrasound.

The advantages of mounting the equipment on the top of the tank are:

- 1 The instrument does not come in contact with the process fluid.
- 2 The velocity of transmission is not fluid-specific.

But its disadvantage is that some sound energy gets lost owing to scattering by the liquid vapour that is invariably present on top of the liquid.

The bottom-mounted transceiver gives the best value of the liquid level if the liquid is free from suspended particles. In the bottom-mounted design, the transceiver can be placed on the outer wall of the tank with the advantage of having no contact with the process fluid. But then, some sound energy will be lost by conduction through the tank material.

Commercially available ultrasonic level detectors have a standard range of about 15 m offering a measurement accuracy of 0.1% of the full scale. They can also be used to measure solid levels. However, it is important, whether solid or liquid, the level should be flat so that the reflectivity is good. Fluffy surfaces containing foam or dirt have poor reflectivity and therefore this method is prone to yield erroneous results.

Gamma-ray Level Indicator

Most of the elements that are heavier than lead, are unstable. They disintegrate to form lighter elements, accompanied by α -, β - or γ -radiations. As is well known, α - and β -radiations consist of positively and negatively charged particles respectively while γ -radiation is electromagnetic wave of very high frequency. α -rays cannot penetrate even human skin and β -rays are stopped by aluminium sheet of about 6 mm thickness. γ -rays, on the other hand, have great penetrating power and therefore, can be utilised in level gauging. Apart from its penetrating power, γ -rays are not deflected by stray electric or magnetic fields while α - and β -rays are.

Two most commonly used sources of γ -rays are Cobalt-60 and Cesium-137, having half-lives of 5.3 years and 30 years respectively. The principle involved in the measurement is that when γ -ray travels through a material, its intensity is reduced according to the relation

$$I = I_0 \exp(-\mu\rho d) \quad (12.16)$$

where I_0 is the intensity of the incident radiation

I is the intensity of the emitted radiation after passing through the material

μ is the mass absorption coefficient of the material

ρ is the density of the material

d is the distance travelled through the material.

Detectors are usually Geiger-Muller (GM) counters, ionisation chambers or scintillation counters.

Equation (12.16) is linear only for small level changes, making the arrangement in Fig. 12.21(a) unsuitable for rather large level measurements. For such cases, a long strip of radioactive material is used and cascaded detectors are used to detect radiations [Fig. 12.21(b)]. The higher the level of the fluid, the less will be the value of the detected current (for ionisation chambers) or count (for GM or scintillation counters). But the level height vs. detected value will be strictly linear.

The γ -ray level indicators are elegant and non-intrusive. But they are somewhat hazardous for workers who need to monitor their radiation exposure carefully. Range of measurement by these level indicators can be from 25 mm to 7 m with a precision of about 1% of the span.

Radar (or Microwave) Level Indicator

To find liquid levels, radar or microwave level indicators utilise either of two methods, namely

1. Time of flight method
2. Frequency modulated continuous wave (FMCW) method

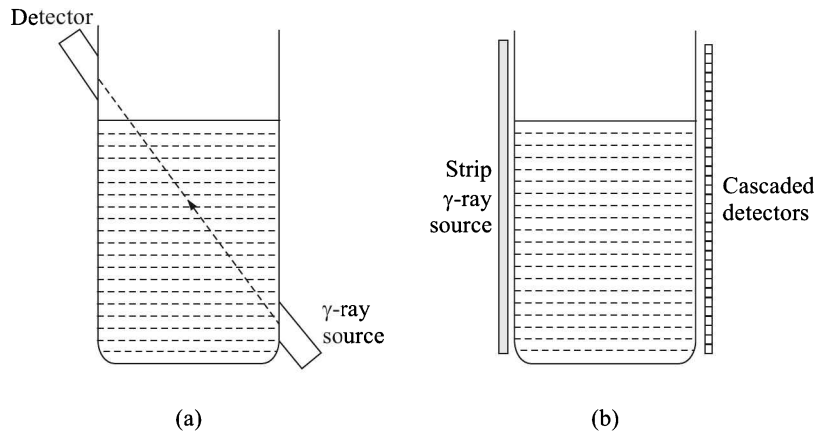


Fig. 12.21 (a) Simple arrangement for small level height, and (b) linearisation for large level heights.

Time of flight method

The time of flight method is similar to the one used in laser or ultrasonic level detectors which employ the time-of-flight measurement of a short duration pulse to determine the height of the liquid level.

Electromagnetic waves in the microwave X- and K- bands (~ 10 GHz and ~ 24 GHz respectively) are transmitted through antennae of different shapes. The reflected pulse (*echo*) is also received there. The time difference between the *send* and *echo* pulses gives the distance between the transceiver and the liquid level (see Fig. 12.20).

Frequency modulated continuous wave (FMCW) method

In this method a continuous wave radar system is used to measure the level by frequency modulation i.e., a systematic variation of the transmitted frequency [Fig. 12.22(a)]. The sweep duration of the transmitted signal is much larger than the time required for measuring the installed maximum range of the radar. The frequency vs. time sweep, which looks like a wave, gets reflected at the target and returns with a different phase. This is shown in Fig. 12.22(b).

The instantaneous frequency difference Δf between the transmitted signal and its echo is measured. This difference is directly proportional to the time delay Δt between the transmitted signal and the reception of its echo from the liquid level. The time delay can be found from the relation

$$\Delta t = \frac{T}{f_2 - f_1} \cdot \Delta f$$

where, f_1 is the minimum frequency

f_2 is the maximum frequency

T is the period of sweep from f_1 to f_2

Δf is the difference between the transmitted frequency and the echo frequency at any instant

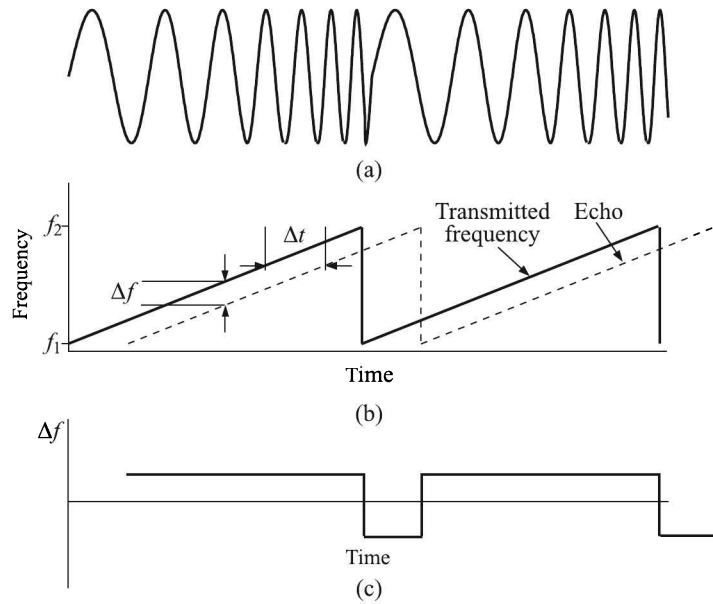


Fig. 12.22 FMCW method: (a) frequency modulation of the transmitted signal, (b) frequency vs. time sweep of transmitted and echo signals, and (c) Δf vs. time plot.

The distance d between the transmitter and the liquid level is given by

$$d = \frac{c\Delta t}{2}$$

where c is the velocity of light at which the signal propagates.

It may be noted that when the sweep resets the frequency, Δf becomes negative as may be seen from the Δf vs. time plot [Fig. 12.22(c)]. This difficulty is obviated by using a discriminator that clips off the negative signal as shown in Fig. 12.23.

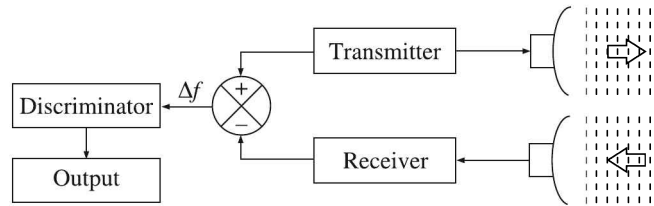


Fig. 12.23 FMCW level measuring system.

Another way to build an FMCW system is based on the comparison of the phase difference between the transmitted and received signals after they have been demodulated. This arrangement has the advantage that no discriminator is necessary to clip off negative Δf values.

Liquid level measurement involves a typical 30 m range. Hence the power requirement of the microwave generator is $\sim 0.015 \text{ mW/cm}^2$. Consequently, low power solid state devices can be used to generate the required wave pattern. Either FMCW or pulse generators can be used.

The former gives a little more accurate values. The FM characteristic of the signal, which offers more fidelity than an AM one, gives the method an edge over the pulse method. The pulse method is somewhat affected by the tank noise, most of which are in the AM domain.

Microwave velocity, being about a million times higher than the sound velocity, is little affected by changes in the ambient temperature or presence of foam, dust, mists, etc. in the wave path, unlike ultrasonic method of level indication. This is one of the reasons why the radar level gauge is finding more and more applications replacing ultrasonic or γ -ray level gauges.

However, for all error-free operation of a radar level gauge, the dielectric constant of the liquid has to be substantial, else signals are absorbed by the liquid yielding a low echo signal. Also, glass-lined tanks or tank nozzles functioning like waveguides produce spurious echo effects.

A comparison of the performance of minimally-invasive level-gauging methods is presented in Table 12.2.

Table 12.2 Comparison of minimally-invasive level-gauging methods

| <i>Method</i> | <i>Reproducibility</i> | <i>Nonlinearity^a</i> | <i>Ruggedness</i> | <i>Comments</i> |
|-------------------------------|------------------------|---------------------------------|-------------------|--|
| Magnetostriction ^b | High | Low | High | |
| Radar ^c | High | Medium | High | The dielectric constant of the liquid has to be substantial. |
| Ultrasonic ^d | Medium | Medium | Medium | Output signal can be scattered by turbulence and foam. Other liquids, steam, dense vapours and dust interfere. |
| Capacitive ^e | Medium | Medium | Low | Changes in dielectric constant of the stored material, if it is nonconductive, may upset the calibration. |

^a Lower nonlinearity is better.

^b see Section 12.1 at page 496.

^c see Section 12.4 at page 511.

^d see Section 12.4 at page 510.

^e see Section 12.3 at page 508.

12.5 Level Switches

Level switches are used to detect liquid levels or interfaces between liquids. They can be of the following types:

1. Magnetic
2. Thermal
3. Vibrating

Magnetic Switch

Switching action for maintaining a particular level may be initiated magnetically as shown in Figs. 12.24 and 12.25.

Top-mounted float

In the top-mounted arrangement (Fig. 12.24), the float is attached to a tie rod with a magnetic sleeve. When the liquid moves to the desired level, the sleeve moves into the field of a permanent magnet attached to a crank and spring arrangement. As the liquid level falls below the desired level lowering the sleeve, the magnet is drawn away by the spring. This movement of the crank changes the electrical state of a mercury or any other switch which actuates a suitable motor to pump in more liquid.

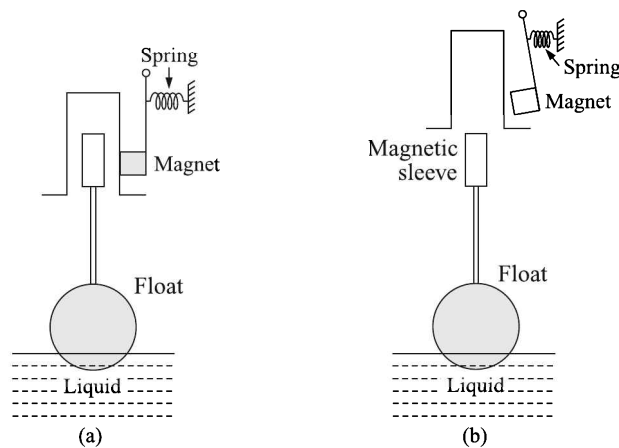


Fig. 12.24 Top-mounted float: (a) The switching element, having a magnetic tip, is attached to the float connector, and (b) the switching element is drawn away by the spring as the level falls, changing its electrical state.

Side-mounted float

The same action for a side-mounted float is shown in Fig. 12.25. Here, instead of a magnet, a variable reluctance type magnetic pick-up has been shown.

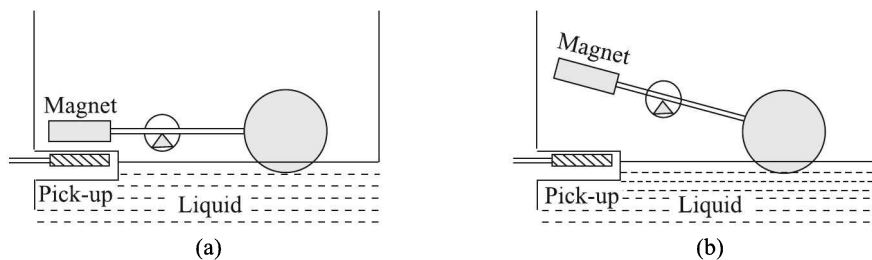


Fig. 12.25 Side-mounted float: (a) The float connector is close to the magnetic pick-up, and (b) the float connector is away from the magnetic pick-up and this changes the electrical state of the switching element.

Permanent magnets are also utilised to actuate hermetically sealed reed switches. A top-mounted and a side-mounted device are shown in Figs. 12.26(a) and (b) respectively. These rather inexpensive level switches are used in vending machines for liquid drinks and beverages.

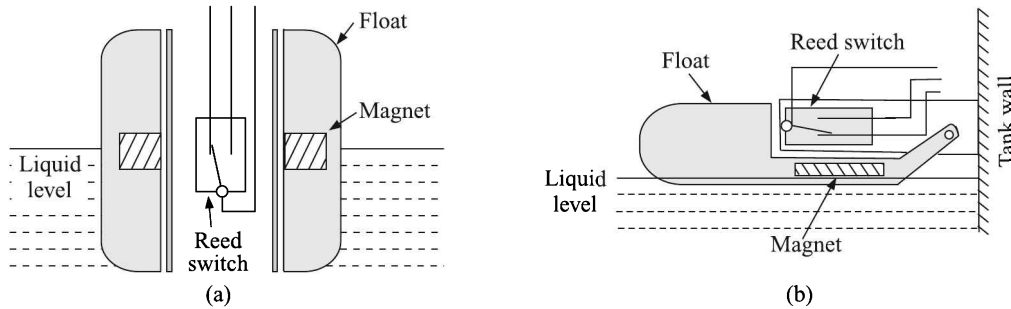


Fig. 12.26 Permanent magnets along with reed switches for level switching action: (a) top-mounted, and (b) side-mounted float.

Thermal Switch

Simplest thermal switch

The simplest thermal level switch design consists of a temperature sensor heated with a constant current. As long as the probe is in the vapour region of the tank, the probe remains at a high temperature, because vapours are poor conductors of heat. When the probe is submerged, the liquid conducts more heat with a consequent drop in the probe temperature. The switch is actuated when this change in temperature occurs.

Heated and unheated RTD type

In another design, two resistance temperature detectors (RTDs), both mounted at the same elevation, are used. One of them is heated and the other provides an unheated reference [Fig. 12.27(a)]. Both the sensors are connected to a Wheatstone bridge [Fig. 12.27(b)]. When both probes are submerged in the process liquid, their temperatures will approach that of the liquid because of good thermal conduction by the liquid. As a result, their resistances will be nearly equal and the bridge will remain balanced.

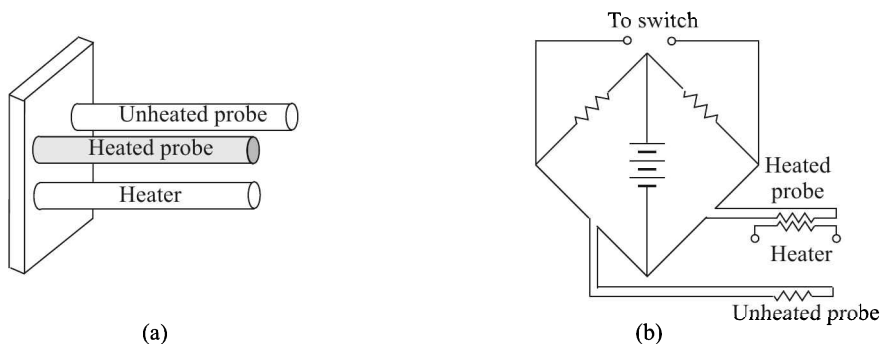


Fig. 12.27 Thermal switch: (a) RTD and heater arrangement, and (b) bridge configuration.

However, if the level goes down and the sensor is in the vapour phase, the heated probe will be warmer than the reference probe because of poor thermal conduction of the vapour, and the bridge will be thrown out of balance. The level switch is actuated when a change in bridge balance occurs.

Capsule type

In a third type of thermal switch, two resistive sensors inside the same vertical probe are used. One of them is mounted above the other and both are connected to a voltage source (Fig. 12.28). The current flow through the two sensors is the same when both of them are in the vapour or in the liquid phase. If, however, the lower one is in liquid and the upper in vapour, more current will flow through the lower sensor. This happens because the liquid is a better conductor and therefore, it will keep the lower sensor at lower temperature with a consequent reduction of its resistance. This difference in current through the two sensors is detected by a current comparator which signals that the sensor has reached the vapour/liquid interface.

One interesting feature of this design is that the sensor capsule can be suspended by a cable into a tank or well, and the sensor output can be used to drive the motor that moves the cable up and down. In this way, the level switch can be used as a continuous detector of the location of the vapour/liquid interface.

Since all process materials have characteristic heat transfer coefficients, thermal level switches can be calibrated to detect the presence or absence of *any* fluid. Therefore, these switches can be used in difficult services, such as interfaces, slurry, and sludge applications. They can also detect thermally conductive foams if they are spray-cleaned after each operation.

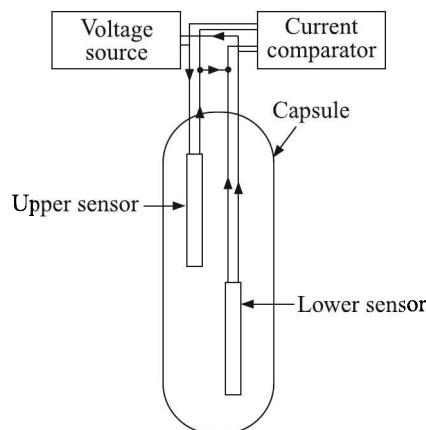


Fig. 12.28 Capsule type thermal level switch.

The advantage of thermal level and interface switches is that they have no mechanical moving parts. They are rated for pressures up to 3,000 psig and process temperatures from -75 to 175°C . When detecting water level, the response time is typically 0.5 second and accuracy is within 2 mm. In general, thermal level switches work best with non-coating liquids and with slurries having 0.4–1.2 specific gravity and 1–300 cP viscosity.

Vibrating Switch

Vibrating level switches detect the damping that occurs when a vibrating probe is immersed in a process medium.

The effect of viscosity, which is significantly higher for liquids than for vapours, dampens the movement of a body. These level switches measure the degree of damping of a vibrating fork when dipped in a liquid. Normally, it is only used as a point level switch. Figure 12.29 shows such a *tuning fork*, named according to the typical form with two or three vibrating paddles.

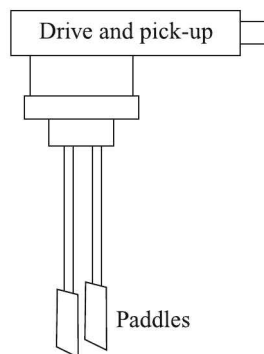


Fig. 12.29 Vibrating level switch.

The driver coil induces a 120 Hz vibration in the paddles that are damped out when the paddles get covered by a process material. The integrated electronics evaluate the power loss or the frequency shift of the mechanical resonance system. The switch can detect both rising and falling levels, and only its actuation depth (the material depth over the paddle) increases as the density of the process fluid drops. The variation in actuation depth is usually less than an inch.

For solids, a sensor with a rotating paddle that stops when contacting the product is useful.

Apart from the three kinds described above, the switching action can be triggered with the help of displacers, differential pressure indicators and all electrical, optical and radiative level indicators.

Review Questions

- 12.1 (a) Explain with diagram how a differential pressure transmitter can be used to measure level of liquid in an enclosure (i.e. not open to atmosphere).
 (b) Explain the working principle of a non-contact type level sensor.
- 12.2 (a) Consider a high pressure switch for giving alarm if pressure goes high with calibrated set-point at 10 kg/cm^2 and differential gap of 0.8 kg/cm^2 . Explain what will happen to the state of pressure switch
 (i) if the applied pressure increases from 0 kg/cm^2 to 10.2 kg/cm^2
 (ii) then the pressure comes down from 11 kg/cm^2 to 9.5 kg/cm^2
 (iii) then the pressure comes down from 9.5 kg/cm^2 to 0 kg/cm^2

- (b) Explain with a suitable diagram, the working principle of a displacer-type level sensor.
- 12.3 (a) How will you measure the level of a liquid in a pressurised tank containing vapours which are likely to condense, using a differential pressure transmitter?
- (b) A d/p transmitter is being used for level measurement in an open to atmosphere tank. The specific gravity of the liquid in the tank is 1.1. The transmitter is connected to the tank 200 mm below the minimum level. The maximum level is 2 metres above the minimum level (refer to Fig. 12.30 for dimensions)

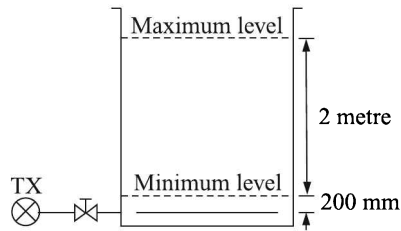


Fig. 12.30

- (i) Find the calibrated range and span of the transmitter in mm of water column (mmwc).
- (ii) What can you say about the zero of the transmitter?
- 12.4 (a) A cylindrical metal rod (probe) is used to measure the level of non-conducting liquid in a metal tank, based on capacitive principles. Show that the effective capacitance varies linearly with the level in the tank.
- (b) Explain the FMCW method of radar level measurement.
- 12.5 Write a short note on torque tube.
- 12.6 How can the capacitive transducer be used to measure the level of a non-conducting liquid? What is the special arrangement required to measure the level of a conducting liquid?
- 12.7 Write the principle of operation of a level measurement system using capacitive transducer for a non-conducting liquid.
- 12.8 What is level gauge? Explain the principle of resistance switching type level gauge.
- 12.9 (a) Show that in the capacitive level measurement using insulated probes, the level is directly proportional to the capacitance.
- (b) What are the advantages of using insulated probes?
- (c) Explain the method of level measurement in an open-to-atmosphere tank using differential pressure transmitter. Draw diagrams whenever needed.
- 12.10 (a) Describe with a neat sketch, the working principle of float type level measurement method.
- (b) What are different types of thermal level sensors? Where are they used? Explain the method of level measurement using one of them.

- 12.11 Explain with a neat sketch, the method of level measurement in an underground water tank using float operated device. Distinguish between the operation of float and displacer type level gauge.
- 12.12 Two concentric tubes of length 8 m and diametric ratio of 2.0 are used as a capacitive level transducer to measure the depth h of liquid in a tank. The liquid depth varies between 0 and 7 m. The dielectric constant of the liquid is 2.4 and the permittivity of the free space is 8.85 pF/m . The transducer (C_2) is incorporated in a bridge as shown in Fig. 12.31.
- (a) Calculate the value of the adjustable capacitance C_1 to set the open circuit bridge voltage to zero when the tank is empty.
- (b) Calculate the output voltage V_o when the tank is full.

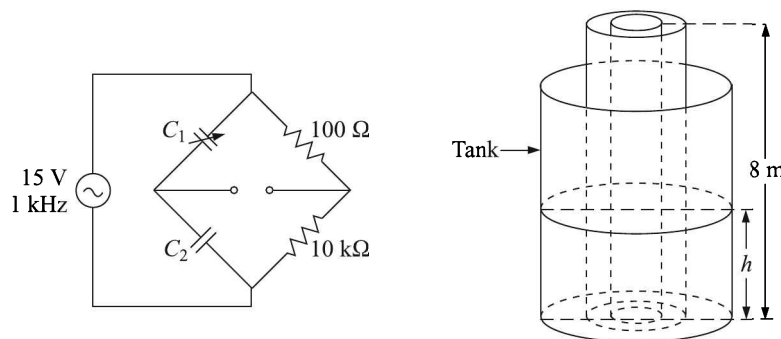


Fig. 12.31

- 12.13 Indicate the correct choice:
- (a) When it is desired to measure liquid level with liquid under pressure or vacuum, the sight glass must be connected to the tank
- at the top
 - at the bottom
 - at the top as well as bottom
 - none of these
- (b) In the radiation level detector, when the tank is full with liquid, the amount of radiation received at the detector is
- directly proportional to the amount of liquid between the radiation source and the detector
 - inversely proportional to the amount of liquid between the radiation source and the detector
 - independent of the amount of liquid
 - none of these

-
- (c) The performance of a capacitive level indicator is severely affected by dirt, because it changes the
- (i) area of the plate
 - (ii) distance between two plates
 - (iii) dielectric constant
 - (iv) none of these
- (d) The basic principle of float type level sensor is
- (i) force balance
 - (ii) motion balance
 - (iii) energy balance
 - (iv) none of these

Miscellaneous Measurements

13.1 Humidity and Moisture Measurement

Humidity and moisture measurements are necessary in weather monitoring or heating-ventilation-airconditioning (HVAC) and process industries where moisture present in gases or liquids or solids constitutes an important parameter for monitoring. Methods employed in these kinds of measurement vary widely depending upon the requirement of accuracy of measurement. We will consider here only a few of such methods.

Before going into the methods of measurement of humidity and moisture, let us define the relevant parameters.

Dew-point. Suppose, we have a volume of air in a room at the ambient temperature T_0 °C. It contains a certain amount of water vapour such that the air is not saturated with it. The air pressure in the room is the sum of the partial (or fractional) pressures of dry air and water vapour. Now, if we go on cooling the room, a temperature will be reached when dew will form on a clean solid surface such as a mirror. This temperature is called the *dew-point*. Clearly, at the dew-point the air is saturated with the same quantity of water vapour that was present at T_0 °C.

Relative humidity. Relative humidity (RH) is defined as:

$$\text{RH} = \frac{\text{Mass of water vapour actually present in } V \text{ volume of air}}{\text{Mass of water vapour necessary to saturate } V \text{ volume of air}} \Bigg|_{T_0 \text{ } ^\circ\text{C}} \quad (13.1)$$

$$\begin{aligned} &= \frac{\text{Pressure of water vapour actually present in } V \text{ volume of air}}{\text{Pressure of water vapour necessary to saturate } V \text{ volume of air}} \Bigg|_{T_0 \text{ } ^\circ\text{C}} \\ &= \frac{\text{Saturation vapour pressure of air at the dew-point}}{\text{Saturation vapour pressure of air at } T_0 \text{ } ^\circ\text{C}} \quad (13.2) \end{aligned}$$

Obviously, whatever we talked about *air* is true for any gas.

Hygrometers, which are used generally to measure humidity, can broadly be divided into two categories:

1. RH hygrometers
2. Dew-point hygrometers

The former uses Eq. (13.1) while the latter uses Eq. (13.2) to determine RH. RH, normally expressed in its percentage, is the most commonly required information.

Gravimetric method. Moisture determination is intimately related to humidity measurement. Therefore, we will not deal with it separately. The only indisputable method for measuring moisture content (MC) is to weigh the sample, oven dry (OD) it in accordance with published standards, and then find the MC based on OD weight (the amount of glue solids may alter this a bit, but not enough to make a practical difference). We will, however, omit details of this obvious method.

Wet and Dry Bulb Hygrometer

This ubiquitous hygrometer consists of two mercury-in-glass thermometers, the bulb of one of which is covered with a wick or muslin. The wick or muslin, in turn, is always kept moist by dipping its one end into water contained in a small vessel. Continuous evaporation of water from the surface of the wet bulb keeps its temperature lower than that of the dry bulb. Temperatures indicated by the two thermometers are related to the RH of the sample atmosphere.

For a reliable measurement, the sample velocity should be well over 3 m/s. The hygrometer, therefore, should be mounted at a place where the circulation is adequate either naturally or caused by a circulating fan.

Sling psychrometer. A variant of this hygrometer, called the *sling psychrometer*¹, has a sling attached to it (Fig. 13.1) so that the unit can be whirled manually before a reading is taken.

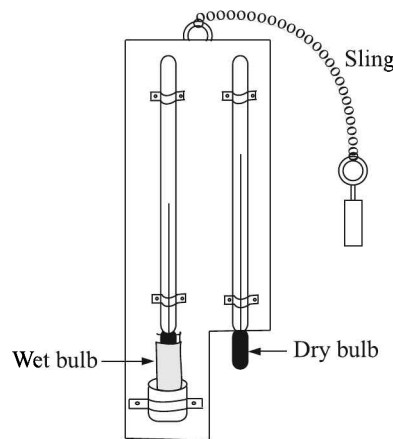


Fig. 13.1 Sling psychrometer.

From a knowledge of the dry-bulb temperature T_d and wet-bulb temperature T_w , RH can be calculated using psychrometric Table² or from theoretical formulae. At any temperature T_d , vapour pressure can be expressed by a linear relation of the type

$$p_v = A + B\Delta T$$

where A and B are empirical constants and $\Delta T = T_d - T_w$. A is not constant with temperature because it corresponds to the saturation vapour pressure when $\Delta T = 0$. B is nearly constant,

¹The prefix *psychro* is a Greek word meaning *cold*.

²See Appendix E.

showing a small variation with T_d . In fact, A and B can be expressed as quadratic functions of T_d . The following relation holds good for $T_d = 10^\circ\text{C}$ to 20°C and $\Delta T = 0^\circ\text{C}$ to 6°C .

$$p_v|_{\text{mm of Hg}} = A + B\Delta T \quad (13.3)$$

where

$$A = 5.73927 + 0.10369T_d + 0.02361T_d^2$$

$$B = -0.82788 - 0.00863T_d - 0.00105T_d^2$$

Better thermodynamic relations involving enthalpies of evaporation, saturated vapour, etc. cover the entire range of dry and wet bulb temperatures. But for all practical purposes, the psychrometric table is adequate.

Example 13.1

The readings of dry and wet bulbs are 18°C and 16°C at normal atmospheric pressure. What is the value of the RH?

Solution

From Eq. (13.3),

$$p_v|_{18^\circ\text{C}} = 15.255 \text{ mm for } \Delta T = 0^\circ\text{C}$$

$$p_v|_{16^\circ\text{C}} = 12.608 \text{ mm for } \Delta T = 2^\circ\text{C}$$

Therefore,

$$\text{RH} = \frac{12.608}{15.255} \times 100 = 82.6\%$$

The psychrometric Table gives the value as 82%.

Hair Hygrometer

Certain organic and synthetic fibres elongate at humid atmospheres and contract in dry ambience. Hair is one such material. Hygrometers which utilise this property are collectively called *hair hygrometer* though the sensing fibre may not necessarily be hair (originally horse hair). Figure 13.2 shows a hair hygrometer in which several strands of fibre are mounted together to form a band.

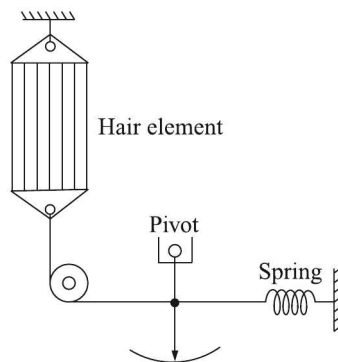


Fig. 13.2 Hair hygrometer.

One end of this band is kept taut by a spring while the free end is linked to a circular readout device. Transmitters can be pneumatic (such as nozzle-flapper) or electrical (such as LVDT) to generate suitable signals.

Like wet and dry bulb thermometers, hair hygrometers require good circulation of the air/gas in order that a reliable RH value can be obtained.

Dunmore Cells

Dunmore cells consist of either a wire grid or a dual winding on an insulated substrate which is coated with a hygroscopic substance such as lithium chloride (Fig. 13.3). The resistance between terminals A and B will depend on the amount of moisture absorbed by the intervening medium, i.e. LiCl.

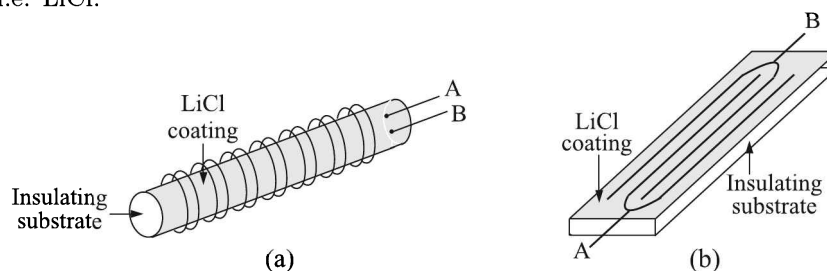


Fig. 13.3 Dunmore cell: (a) dual winding type, and (b) wire grid type.

To avoid polarisation of the medium, the cell is excited by an ac source rather than dc and the resistance is measured by a bridge. Non-reacting metals like platinum are used to construct the wire grid/winding. Typically, a cell is about 5 cm long and 2.5 cm in diameter. The resistance vs. RH curve is nearly linear over a rather small range covering about 10% in RH. So, to cover the entire range of 0 to 100%, about 10 cells of different thickness of LiCl coating may be necessary. Typical parameters of Dunmore cells are:

| <i>Accuracy</i> | <i>Resolution</i> | <i>Working temperature</i> | <i>Speed of response</i> |
|-----------------|-------------------|----------------------------|--------------------------|
| ~1.5% of RH | ~0.15% of RH | -40 to 65°C | ~3 s |

Pope Cells

Pope³ cells use polystyrenes, treated with sulphuric acid, as substrates. They are similar to Dunmore cells otherwise. The sulphuric acid treatment leaves a thin hygroscopic layer on the surface of the substrate. Change in RH of the ambience can cause a huge change in resistance from 1 MΩ to 1 kΩ at 0% and 100% RH respectively. Thus, a single sensor can cover the entire RH range, though the characteristic curve is nonlinear.

Solution-conductivity Cell

We know that LiCl absorbs moisture from the atmosphere or the surrounding gas. How long will it continue to do so? The answer is, until the vapour pressure of moisture absorbed

³Named after its inventor Martin Pope (1918) who is a physical chemist and professor emeritus at New York University.

by the desiccant equals the partial pressure of water vapour present in the atmosphere or gas⁴. Therefore, by estimating the vapour pressure of moisture in the desiccant at equilibrium condition, the moisture content of air or gas can be estimated. Based on this principle, the solution-conductivity cell is constructed (Fig. 13.4).

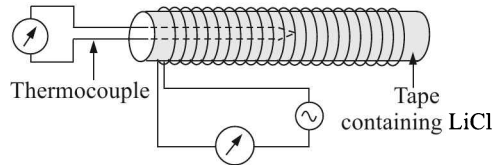


Fig. 13.4 Solution-conductivity cell.

An insulating tape, soaked with minute crystals of lithium chloride and wrapped over a thin-walled hollow tube constitutes the sensor. Two wires are wound over the tape and connected to a steady ac voltage source. When the sensor is exposed to humid atmosphere, water condensing on the hygroscopic LiCl forms an ionic solution completing the electrical circuit between the two wires. Then a current starts flowing through the circuit. As in Dunmore cells, an ac voltage source is used to prevent electrolysis of the ionic solution.

The way the cell functions is interesting. The resistivity of LiCl falls sharply when the RH is about 11%. Then current conduction through the cell begins. But the flowing current heats it up and some moisture is driven off from it as a result. This results in increasing its resistance and lowering the current. That, in turn, lowers its temperature. The moment its temperature falls, it absorbs more moisture. The cycle is repeated until it reaches an equilibrium temperature when the RH near the sensor element stays at around 11%, irrespective of the RH of the atmosphere.

This equilibrium temperature of LiCl sensor is measured by a suitable sensor, like thermocouple, placed inside the hollow metal tube. It has been established that the equilibrium temperature of the LiCl sensor is directly related to dew-point. Once the dew-point is known, RH can be calculated from standard tables of saturation vapour pressures.

These cells help measure dew-points between -45°C and 70°C with an accuracy of $\pm 0.5^{\circ}\text{C}$.

Condensation on a Chilled Surface

The dew-point is the temperature when a molecular layer of condensation takes place on a clean, polished, chilled surface. The condensation can be tracked optically or by measuring the electrical conductivity of the surface.

Optical method

A polished surface, such as that of mirror, can be chilled by any standard method and the temperature at which first fog appears on the surface can be observed visually. However, the measurement is prone to human errors.

⁴This is related to *Henry's law* which states that at constant temperature, the mass of water vapour dissolved in a given volume of liquid is proportional to the partial pressure of water vapour in the sample. William Henry (1774–1836) was an English chemist.

A completely automated method utilises cooling through Peltier effect and the current sent through the junction is controlled by a feedback from an optical sensing arrangement (Fig. 13.5).

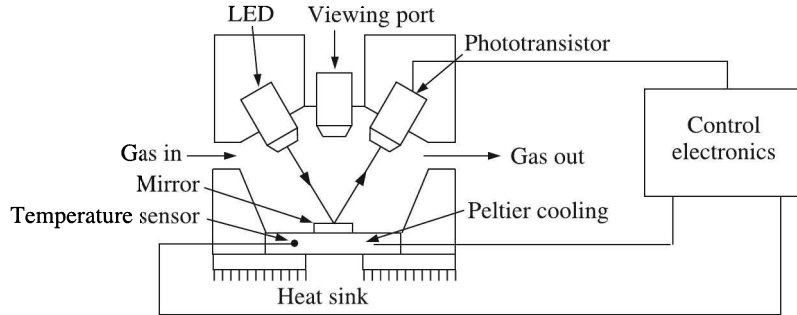


Fig. 13.5 Automated optical sensing of dew-point.

Light from the LED falls on the mirror at an angle. Its reflection is received by a phototransistor. The mirror substrate is chilled by an electrical cooler which utilises Peltier effect. The substrate temperature is tracked by a thermistor sensor. The phototransistor output, compared to a standard voltage, is the feedback element. No sooner this output goes down owing to the fogging of the mirror, than the current to the Peltier cooler is stopped and the thermistor reading is noted. In fact, through this feedback system, the mirror temperature can always be maintained at the dew-point and a programmed microprocessor can give the RH readout.

Surface conductivity method

In this method, the measuring element is made of gold grid laid on a highly polished inert surface. A thermocouple is fixed to the surface (Fig. 13.6).

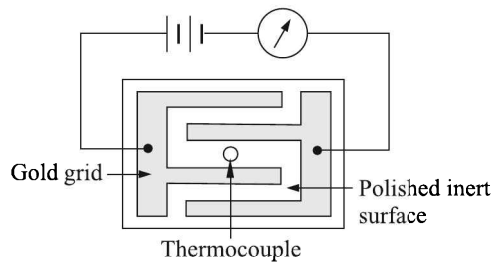


Fig. 13.6 Dew-point measurement by surface conductivity method.

At temperatures above the dew-point, a low current flows through the circuit because of presence of water vapour in the air (though not visible) and therefore, on the surface of the substrate. If the substrate temperature is lowered, the current will increase logarithmically near the dew-point. Below the dew-point, the logarithmic increase will again disappear. Thus by monitoring the surface electrical conductivity while chilling the substrate, the dew-point can be determined.

Electrolytic Hygrometer

Two helical windings of a noble metal are embedded inside an insulator, like glass or Teflon tube. These two windings form two parallel electrodes. On top of the windings, a thin layer of a desiccant, like P_2O_5 , is deposited (Fig. 13.7).

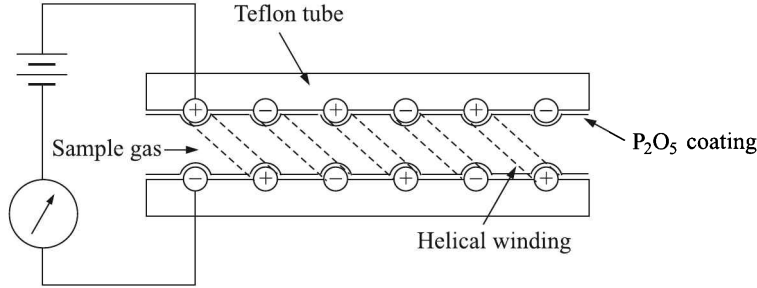


Fig. 13.7 Electrolytic hygrometer.

A fixed voltage, with a current measuring arrangement and a constant sample flow rate (typically $100 \text{ cm}^3/\text{min}$) arrangement, form the electrolytic cell.

Water vapour, present in the sample gas, is absorbed by the desiccant. Water molecules suffer electrolysis and dissociate as a result to form hydrogen and hydroxyl ions which are collected at the cathode and anode respectively. This results in a current flow in the circuit. The more the water vapour in the sample gas, the more is the current flow. We show here that the relation between water vapour content in the air and the electrolytic current is linear.

If I is the electrolytic current in amperes (= coulomb/s), then

$$\text{number of electrons flowing/s, } n_e = (6.25 \times 10^{18})I \quad (13.4)$$

because, 1 coulomb = 6.25×10^{18} electrons. Each water molecule, on electrolysis, releases 2 electrons. Hence,

$$\text{number of water molecules flowing/s, } n_w = \frac{(6.25 \times 10^{18})I}{2} \quad (13.5)$$

From Avogadro's law it follows that a gram-molecule of gas, which contains N (= Avogadro's number = 6.02×10^{23}) atoms/mole, occupies 22.4 litres at NTP. So, the volume occupied by n_w water molecules at NTP is given by

$$V_1 = \frac{22.4 \times 10^3}{N} n_w \text{ cm}^3 \quad (13.6)$$

At any other temperature T_2 and pressure P_2 , the occupied volume can be obtained from the perfect gas law as

$$V_2 = V_1 \frac{p_1}{p_2} \cdot \frac{T_2}{T_1}$$

where p_1 and T_1 are 1 atm (= 101.3 kPa) and 273 K respectively. So, from Eqs. (13.4), (13.5) and (13.6), we get

⁵Lorenzo Romano Amedeo Carlo Avogadro di Quaregna e di Cerreto (1776–1856) was an Italian savant. He is most noted for his contributions to molecular theory, including what is known as *Avogadro's law*.

$$\begin{aligned}
 V_2 &= \frac{22.4 \times 10^3 n_w}{N} \cdot \frac{p_1}{p_2} \cdot \frac{T_2}{T_1} \\
 &= \frac{22.4 \times 10^3}{6.02 \times 10^{23}} \cdot \frac{6.25 \times 10^{18}}{2} I \cdot \frac{p_1}{p_2} \cdot \frac{T_2}{T_1} \quad \text{in cm}^3/\text{s} \quad (13.7)
 \end{aligned}$$

Equation (13.7) shows that the volume of water vapour present in the atmosphere at a particular temperature and pressure is proportional to the electrolytic current. In fact, by measuring the flow rate of the sample gas, the electrolytic current, ambient temperature and pressure, the absolute humidity of the sample can be found by this method. Example 13.2 will make it clear.

Electrolytic hygrometers have a typical range of 0 to 2000 ppm with a resolution of < 1 ppm. The time constant is typically 30 s and the precision, 5% of full scale. The main source of error arises out of recombination of dissociated sample molecules which again suffer dissociation and contribute to current. Hydrogen and oxygen samples are mostly prone to recombination which, it has been found, can be minimised by using rhodium electrodes. Among other gases, alcohol vapours, amines, ammonia, hydrogen fluoride and Freon should not be used as sample gas in this hygrometer. It is suitable for most elemental gases and other gases that do not react with P_2O_5 .

Example 13.2

Determine the water concentration in a gas sample when the sample flow rate is $100 \text{ cm}^3/\text{min}$ at 40°C and 70 kPa gauge pressure. The electrolytic current is measured at $300 \text{ }\mu\text{A}$.

Solution

Given,

$$T_2 = 40^\circ\text{C} = 313 \text{ K} \quad p_2 = (70 + 101.3) = 171.3 \text{ kPa} \quad I = 300 \text{ }\mu\text{A} = 300 \times 10^{-6} \text{ C/s}$$

We know,

$$\begin{aligned}
 N &= \text{Avogadro number} = 6.02 \times 10^{23} & p_1 &= \text{normal pressure} = 101.3 \text{ kPa} \\
 T_1 &= \text{normal temperature} = 273 \text{ K}
 \end{aligned}$$

Therefore, from Eq. (13.7)

$$\begin{aligned}
 V_2 &= \frac{22.4 \times 10^3}{6.02 \times 10^{23}} \cdot \frac{6.25 \times 10^{18}}{2} \cdot 300 \times 10^{-6} \cdot \frac{101.3}{171.3} \cdot \frac{313}{273} \\
 &= 2365 \times 10^{-8} \text{ cm}^3/\text{s} = 141908 \times 10^{-8} \text{ cm}^3/\text{min}
 \end{aligned}$$

Since the flow rate is $100 \text{ cm}^3/\text{min}$, the volume concentration of moisture is

$$(141908 \times 10^{-8})/100 = 14.1908 \times 10^{-6}$$

Expressed in ppm (parts per million), it becomes $\cong 14.2 \text{ ppm}$.

Although the example shows the absolute value of moisture in the sample, the RH can be calculated using Eq. (13.1) and standard Tables.

Capacitive Hygrometer

We have seen in Section 6.2 that the capacitance of a parallel plate capacitor changes with the change in the intervening dielectric material. This property is utilised in constructing capacitive hygrometers which can be used to measure RH of air or moisture content in a sample gas.

Two variants are used—one uses alumina (aluminium oxide) desiccant as thick dielectric between electrodes while the other uses a thin layer of the same desiccant.

In the former application [Fig. 13.8(a)], two concentric metal cylinders form capacitor plates and the annulus between the electrodes is filled with alumina. Two spring-loaded metal discs are used to support the alumina and insulated electrode cylinders.

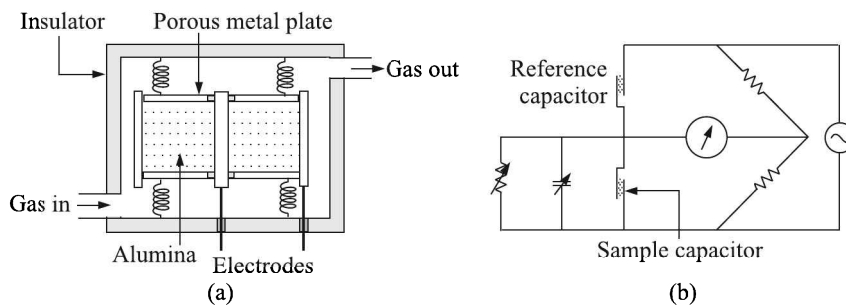


Fig. 13.8 (a) Thick-layer desiccant capacitive transducer (sectional view), and (b) measurement bridge.

Two identical capacitors are used—the sample gas or liquid is passed through one and the other is used as reference. The two capacitors form two arms of a bridge [Fig. 13.8(b)] which is excited by a 15 kHz sinusoidal voltage.

Before the sample gas is passed, the bridge may be balanced with the help of a variable resistor and capacitor. Once the sample gas is passed through the sample capacitor, the bridge is thrown out of balance, the offset voltage being a measure of the moisture content of the sample.

The thin-layer-desiccant-type capacitive hygrometer is constructed by depositing a layer of alumina on ultra-high purity aluminium substrate and vacuum coating the alumina with a thin film of gold (Fig. 13.9).

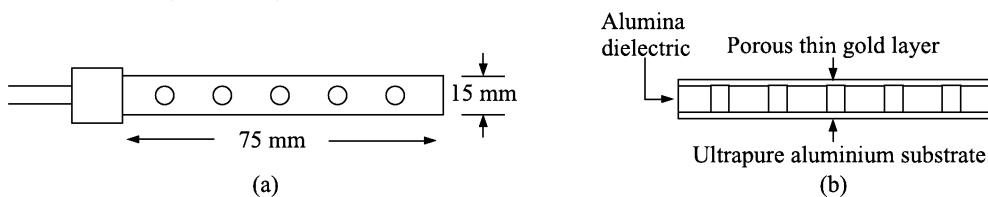


Fig. 13.9 Thin desiccant layer capacitive transducer: (a) probe with holes, and (b) sectional view of the probe.

Aluminium and gold form two electrodes of the capacitive transducer. The thin film of gold allows the moisture to reach alumina which absorbs it, changing the dielectric constant of the capacitor.

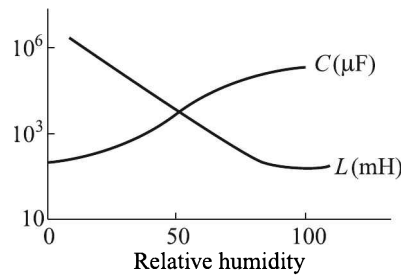
Advantages and disadvantages. The advantages and disadvantages of a capacitive hygrometer are listed in Table 13.1

Table 13.1 Advantages and disadvantages of a capacitive hygrometer

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|--|
| 1. Small size | 1. Calibration curve nonlinear |
| 2. Probe-type packaging | 2. Needs periodic re-calibration to compensate for ageing and contamination |
| 3. Wide measurement range | 3. Low measurement accuracy |
| 4. Suitable for measuring moisture content in many gases including hydrocarbons | 4. Each sensor needs to be calibrated individually |
| 5. Being a capacitor, it can be easily incorporated in microprocessor-based electronic instrumentation | 5. Cannot be used to measure moisture content in polar materials such as alcohols, because then it starts conducting at its exciting frequency of 15 kHz |

Impedance Hygrometer

The construction of impedance hygrometer is very similar to that of the thin-layer capacitive hygrometer. In this case, ac impedance, rather than capacitance of the cell is measured by a suitable bridge. While the capacitance vs. RH curve is nonlinear, the inductance vs. RH curve is linear over a good range (Fig. 13.10).

**Fig. 13.10** Characteristic curves of capacitive and inductive hygrometers.

As in the case of thin-layer capacitive transducer, the impedance hygrometer can be used to measure moisture content in most of the gases and liquids except polar ones like alcohol.

Piezoelectric Hygrometer

A piezoelectric crystal, when excited by an ac signal, starts vibrating at a frequency which is the natural frequency of vibration of the crystal. This natural frequency of vibration is related to the mass of the crystal. Thus, if a piezoelectric crystal, can be somehow loaded with the moisture or water vapour of the atmosphere, its frequency of vibration will change. This property is utilised in the construction of piezoelectric hygrometer.

A suitable piezoelectric crystal, such as quartz, is coated with a hygroscopic material and exposed to the sample gas. The hygroscopic material absorbs the sample moisture, increases the mass of the vibrating piezoelectric crystal and thereby decreases its frequency of vibration. This decrease in the frequency of vibration is a measure of the moisture content of the sample.

To increase the accuracy of measurement, two identical piezoelectric transducers may be used—one exposed to the moist sample and the other to dry gas—at the same flow rate. A little later, the gas flow between transducers may be interchanged by energising and de-energising the solenoid valves that are suitably placed in the flow path of gases. This ensures a better accuracy and maintenance of transducers.

Infrared Absorption Hygrometer

Near infrared radiations (NIR) of wavelengths $1.45\ \mu\text{m}$, $1.93\ \mu\text{m}$ and $2.95\ \mu\text{m}$ are strongly absorbed by water. So, an infrared radiation of either of these wavelengths may be sent through the sample gas [Fig. 13.11(a)] or reflected from the solid sample [Fig. 13.11(b)] and the resultant attenuation may be measured to determine the moisture content of the sample.

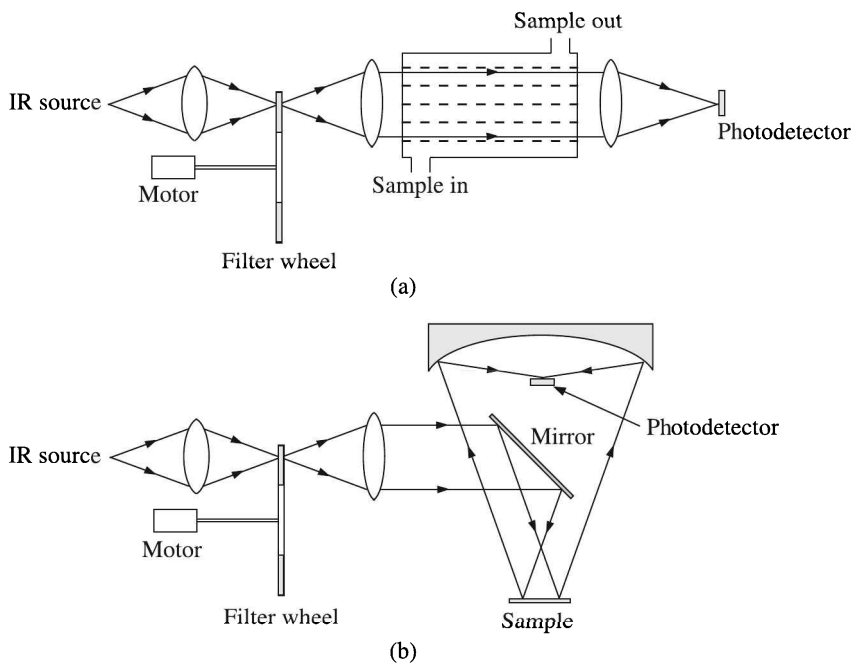


Fig. 13.11 Infrared absorption hygrometer: (a) for gas or liquid sample, and (b) for solid sample.

The attenuation, however, may be caused by not only the water vapour, but also other factors such as reflection and dispersion by other molecules. To eliminate this factor, light of the same intensity but of another wavelength which is not absorbed by water molecules is made incident on the sample by a rotating filter-wheel arrangement and the difference in attenuation of the two radiations is measured with the help of a photodetector which outputs an electric current.

The method yields dependable results for gases. But for solids, its disadvantages are:

1. Its dependence on surface roughness
2. Indication of surface moisture only
3. Nonlinear effects due to water-filled pores of the sample

Microwave Absorption Hygrometer

Microwave radiation of wavelength about 13 mm to 15 mm, corresponding to 20 to 22 GHz (*K*-band) is absorbed by water molecules. So, this radiation can be utilised to determine moisture content of samples much in the same way as done in the IR absorption hygrometer. The microwave radiation can be propagated by waveguides or transmitted from the source to the detector through the sample.

This method can be applied to gases only.

Neutron Backscatter Moisture Analyser

Neutrons are chargeless particles. When a neutron collides with an atom at rest, its trajectory not being affected by surrounding electrons, it hits the atomic nucleus resulting in its lower velocity. The event occurs obeying the laws of conservation of momentum and energy.

If m_n is the mass of neutron

v_1 is the initial velocity of neutron

v_2 is the final velocity of neutron after impact

m_a is the mass of the atomic nucleus

v_a is the initial velocity of the atomic nucleus

then equations for momentum and energy conservations are respectively

$$m_n v_1 = m_n v_2 + m_a v_a \quad (13.8)$$

$$\frac{1}{2} m_n v_1^2 = \frac{1}{2} m_n v_2^2 + \frac{1}{2} m_a v_a^2 \quad (13.9)$$

Equations (13.8) and (13.9) can be rearranged to yield

$$m_n (v_1 - v_2) = m_a v_a \quad (13.10)$$

$$m_n (v_1^2 - v_2^2) = m_a v_a^2 \quad (13.11)$$

Dividing Eq. (13.11) by Eq. (13.10), we get

$$v_a = v_1 + v_2 \quad (13.12)$$

Substituting the value of v_a from Eq. (13.12) in Eq. (13.8), we get on rearranging

$$v_2 = -v_1 \frac{m_a - m_n}{m_a + m_n} \quad (13.13)$$

If $m_a \cong m_n$, the neutron velocity becomes zero which means that the neutron gets absorbed at the nucleus. The hydrogen nucleus contains only one neutron. Hence, it is the most efficient absorber of neutrons. For the same reason, water vapour is also an efficient absorber of neutrons. So, by knowing the attenuation of a neutron beam scattered back by the sample, its moisture content can be determined. However, a simultaneous measurement of the density of the substance is also necessary because the neutron backscatter attenuation also depends on it. The density determination is done by γ -ray attenuation of the sample (Fig. 13.12).

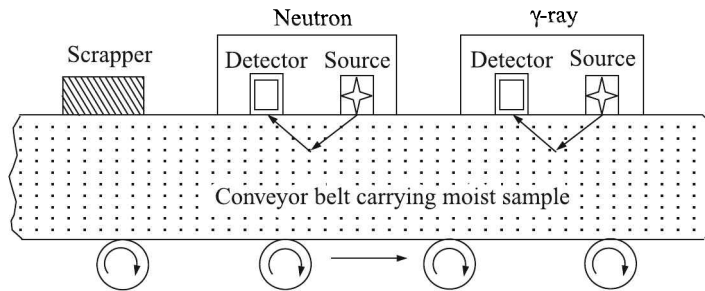
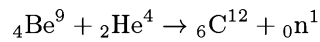
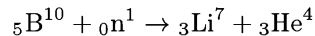


Fig. 13.12 Moisture measurement by the neutron backscatter moisture analyser.

Source and detection. A high energy plutonium-beryllium or americium-beryllium neutron source is used to generate the neutron beam. The light element beryllium, bombarded by α -particles emitted by plutonium or americium, generates neutrons according to the nuclear reaction:



These fast neutrons, when they come in contact with water vapour, get slowed down (called *thermalised*) by elastic collisions and form a cloud. The density of the cloud depends on the equilibrium between the rate of collision of fast neutrons and their thermalisation by nuclei. Thermal backscattered neutrons are detected by a GM counter lined with boron or filled with BF_3 gas. Boron absorbs slow neutrons to emit α -rays according to the nuclear reaction



GM counter detects the emitted α -radiation. But, as already discussed, a simultaneous determination of the density of the sample is also necessary. While discussing γ -ray level detection (Section 12.4) we have seen how the study of attenuation of γ -rays gives information about the density of a material.

Measurement by this method is usually resorted to for materials carried by a conveyor belt. Sample volume should be large—about 40 cm wide \times 5 cm thick. To make the thickness of the sample more or less uniform, a scraper plough may be used.

Advantages and disadvantages. Advantages and disadvantages of moisture determination in solid samples by this method are as follows:

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|---|
| 1. Accuracy of 0.1% or higher can be achieved with some commercial instruments | 1. Since this method involves determination of concentration of H atoms in the sample, it is not suitable if the sample contains hydrocarbons or other hydrogenous material |
| 2. Moisture of a sample can be determined regardless of its physical state | 2. Care must be taken to minimise health risks |
| 3. The instrument can form part of an automated measurement system | |

Nuclear Magnetic Resonance Moisture Analyser

Nuclear magnetic resonance (NMR) is described in Section 14.7 at page 684. The CW-type NMR may be used for moisture measurement. Careful selection of the RF in relation to the magnetic field strength can make the device specific for hydrogen. For example, suppose the RF is set at 21.2 MHz. Then, if magnetic field is maintained at 0.5 tesla by suitably adjusting the magnet current, and the auxiliary coils generate a variation of $\pm 10^{-4}$ tesla about this B , it will produce a proton resonance. Making the device specific to hydrogen allows NMR spectrum to be directly related to moisture content of the sample.

NMR offers a rapid, non-contact method of measurement of moisture content of a sample. But since the method basically determines the concentration of hydrogen nuclei in the sample, the presence of organic matter in the sample may interfere with the measurement. The instrument is rather costly.

Time-domain Reflectometry

Time-domain reflectometry (TDR) is one of the recent advances in determination of moisture in loose solid samples, like soil.

In this method, a pulse of RF energy is sent through a transmission line which is embedded in the sample. The pulse travels through the line to its end and then reflects back. The TOF between the transmitted and reflected pulse is measured to calculate the velocity of pulses. The velocity depends on the dielectric constant of the sample and the loss in transmission line. If losses other than the dielectric constant are small, the velocity v is given by

$$v = \frac{c}{\sqrt{\epsilon}}$$

where, c is the velocity of light and ϵ is the dielectric constant. Measurement of ϵ through TDR helps determine the moisture content of the sample that varies with ϵ . An accuracy of 2% can be achieved by this method.

13.2 Density Measurement

Density determination of a process stream leads to the determination of its composition, concentration and calorific value (for fuels). It also helps to convert a volumetric flow to a mass flow.

Density and specific gravity. As it is well known, density is defined as the mass per unit volume, its SI unit being kg/m^3 . But often it is expressed in g/cm^3 or lb/ft^3 units. Specific gravity (SG) on the other hand is the ratio of the mass of a certain volume of the sample to the same volume of water at 4.4°C . Hence it is a number.

SG and density numbers are the same in CGS units because the mass of 1 cm^3 of water at 4.4°C is 1 g by definition. In all other units, the two numerical values are different. The two quantities, however, are temperature dependent.

In various industries different units for density, other than what we stated above, are in use. Table 13.2 gives an idea about their variety.

Density is measured by various methods in industries. We describe a few of them here.

Table 13.2 Industry-specific units of density

| Industry | Unit in vogue | Definition |
|--------------------|--|---|
| Alcohol | °S (sikes), °R (richter), °T (tralles) | % of ethyl alcohol by volume |
| Acids/syrups | °Be (baume) | $\begin{cases} 145 - \frac{145}{\text{SG at } 17.5^\circ\text{C}} & \text{SG} \geq 1 \\ \frac{140}{\text{SG at } 17.5^\circ\text{C}} & \text{SG} < 1 \end{cases}$ |
| Brewing/sugar | °Ba (balling) | % of dissolved solids by weight |
| Dairy | °Q (quevenue) | $(\text{SG} - 1) \times 1000$ |
| Sugar | °Br (brix) | ° % of sucrose by weight |
| Tanning | °Bk (barkometer) | $(\text{SG} - 1) \times 1000$ |
| Acid/sugar/tanning | °Tw (twaddell) | $(\text{SG} - 1) \times 200$ |

Hydrometers

Hydrometers utilise Archimedes' principle which implies that a partially-floating body will sink to that extent till the weight of the displaced liquid equals the weight of the body. At that state

$$mg = V\rho g$$

where m is the mass of the hydrometer
 V is the volume of the displaced liquid
 ρ is the density of the liquid
 g is the acceleration due to gravity.

Since m is a constant for the hydrometer, V varies with ρ . That means, the hydrometer sinks to different depths in liquids of different densities.

The most common hydrometer consists of a weighted cylindrical float with a narrow 15 to 40 cm long stem, graduated in any unit. The float and stem are made of plastic or glass. Figure 13.13(a) shows a hand-held hydrometer while Fig. 13.13(b) shows an in-line one.

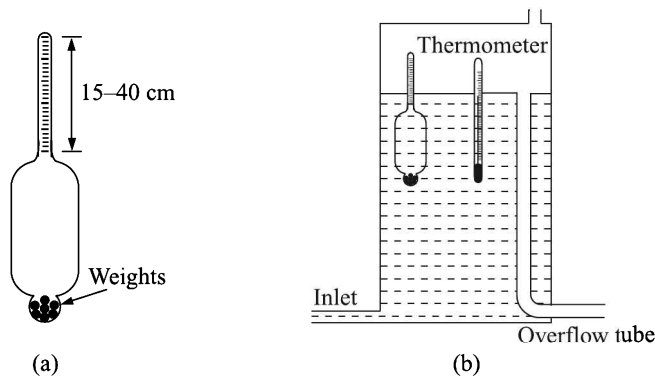


Fig. 13.13 (a) Hand-held hydrometer, and (b) hydrometer fixed for continuous monitoring.

Hydrometer spans vary from 0.05 to 0.5 SG. Spans can be selected from an SG range of 0.6 to 2.1. Minimum scale divisions can be 0.0005 SG. Inaccuracy is generally 1% of the span. Normally, they can work in temperatures up to 95°C.

Differential Bubblers

We have seen in Section 12.1 at page 502 that the pressure readout of a constant flow air-bubbler is directly proportional to the density of the liquid, provided the liquid level is maintained constant. So, an arrangement, as shown in Fig. 13.14(a), can be used to measure the process liquid density.

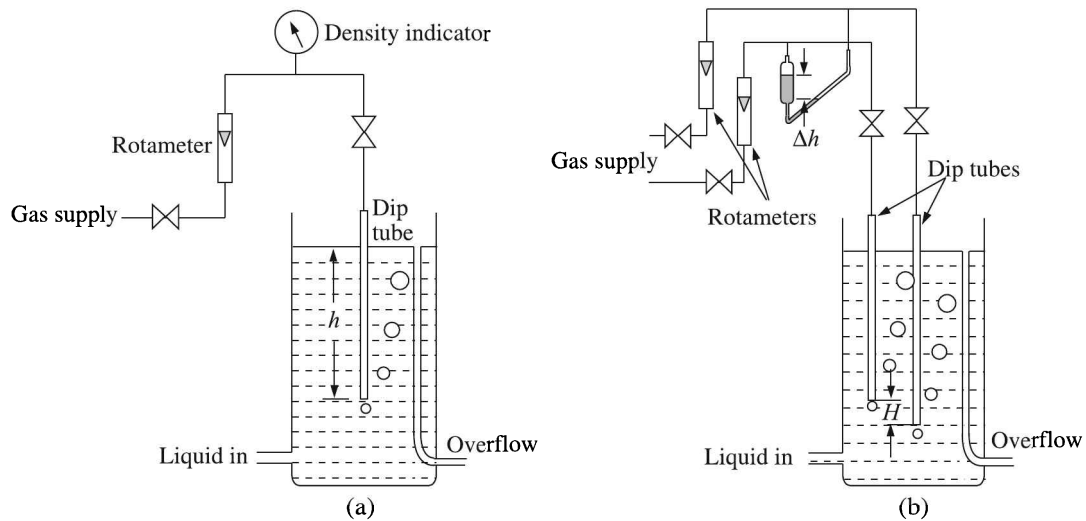


Fig. 13.14 Bubbler method of measuring density: (a) single-bubbler, and (b) double-bubbler.

Obviously, a single bubbler demands a steady superincumbent atmospheric pressure. Else, the density measurement becomes susceptible to errors. Differential bubblers are free from this error.

The differential pressure measurement between two bubblers of different lengths is one of the oldest methods of density measurement [Fig. 13.14(b)].

The tubes are immersed in the liquid to different depths. An inclined manometer is used to measure the differential pressure. In this case,

$$\Delta h \rho_m g = H \rho g$$

$$\Rightarrow \rho = \frac{\Delta h}{H} \rho_m$$

where Δh is the height differential of the manometer
 H is the height differential of bubbler tubes in the liquid
 ρ_m is the density of manometric fluid
 ρ is the density of the process fluid
 g is the acceleration due to gravity.

Differential Pressure Cells

In this method, the liquid level is held constant and the pressure difference between two heights in the liquid tank is measured by a differential pressure cell [see Fig. 12.10 at page 500]. Here,

$$\rho = \frac{\Delta p}{hg} \quad (13.14)$$

where the symbols have their usual significance. A bellows-type indicator, as is used in level measurement [see Fig. 12.10(b) at 500], may be used to measure the differential pressure.

Coriolis Densitometer

We have discussed in Section 11.4 at page 473 the construction and principle of operation of the Coriolis mass flow meter (CMF). There we mentioned that the resonant frequency of the tube depends on the density of the fluid flowing through it. Therefore, the CMF can measure density of the fluid as well. Here we analyse the method in detail.

If, k is the spring constant of the material of the tube
 m is the mass of the system, and
 ω_n is the natural frequency of vibration of the system

then,

$$\omega_n = \frac{2\pi}{T} = \sqrt{\frac{k}{m}} \quad (13.15)$$

Now, the mass of the system is a combination of the mass of the tube and that of the fluid. Therefore,

$$m = \rho_f \alpha_f l + \rho_t \alpha_t l \quad (13.16)$$

where ρ_f is the density of the fluid
 ρ_t is the density of the tube material
 α_f is the area of cross-section of the fluid
 α_t is the area of cross-section of the tube
 l is the length of the tube.

From Eqs. (13.15) and (13.16), we get

$$\rho_f \alpha_f + \rho_t \alpha_t = \frac{kT^2}{4\pi^2 l}$$

This yields

$$\rho_f = \frac{kT^2}{4\pi^2 l \alpha_f} - \frac{\rho_t \alpha_t}{\alpha_f} \equiv K_1 T^2 - K_2$$

where K_1 and K_2 are constants for the densitometer. These constants can be determined experimentally by using two fluids of known densities in the tube and measuring the corresponding time periods of oscillation of the Coriolis tube.

Advantage. The advantage of the Coriolis densitometer is that it is amenable to electronic microprocessor-based instrumentation. Coriolis electronics can support up to 4 detector assemblies and can detect density within an error of $\pm 0.0005 \text{ g/cm}^3$ over a range of 0.001 to 1.8 g/cm^3 at an ambient temperature variation between 5 and 80°C . This high precision, coupled with its ability to measure densities of gas, liquid and slurries, makes it a favourite choice in industries.

Displacer- and Float-type Densitometers

Torque-tube displacer

The construction of torque-tube displacer, acting as a densitometer, is much similar to the one used for level measurement (see Section 12.1 at page 498). But in level measurement, the displacer remains partially immersed in the liquid, while in density measurement it is always totally submerged in the liquid. The reason is as follows.

We know that the apparent loss of weight of a displacer, displacing V volume of liquid, amounts to $V\rho g$. In the case of level measurement, the displacer was partially immersed to make V variable which, in turn, was proportional to the displacement x . In density measurement, we need to make ρ variable. Hence, V is made constant by making the displacer fully submersible.

In a torque-tube displacer, the force range is normally 0 to 0.33 kg-wt. For a thin-wall tube, it is half as much. The volume of the displacer is thus an important parameter which needs to be carefully chosen so that the buoyant force generated by the displacer owing to the maximum density variation of the fluid lies in the force range of the torque-tube. The required volume can be calculated as follows.

Let V be the volume of the displacer, in m^3
 ρ_w be the density of water, in kg/m^3
 ρ_{\max} be the maximum density of the process fluid
 ρ_{\min} be the minimum density of the process fluid
 S_{\max} be the maximum specific gravity of the process fluid
 S_{\min} be the minimum specific gravity of the process fluid
 F_{\max} be the maximum torque-tube force, in kg-wt
 g be the acceleration due to gravity.

Then, the buoyant force is

$$V(\rho_{\max} - \rho_{\min})g = F_{\max}$$

Therefore,

$$\begin{aligned} V &= \frac{F_{\max}}{(\rho_{\max} - \rho_{\min})g} = \frac{F_{\max}}{(S_{\max} - S_{\min})\rho_w g} \\ &= \frac{F_{\max}}{(S_{\max} - S_{\min})(\text{wt. of } 1 \text{ m}^3 \text{ of water})} \end{aligned} \quad (13.17)$$

Displacer-type density sensing may be used for clean and non-viscous fluids. Slurry materials may stick to the surface of the displacer changing its volume and thus making its calibration ineffective.

Example 13.3

The density range of the process fluid is 0.99 SG and 1.00 SG. A thin-walled torque-tube of force range 0.17 kg-wt is to be used. Determine the volume of the torque-tube displacer.

Solution

From Eq. (13.17),

$$V = \frac{0.17}{(1.00 - 0.99)(1000)} = 0.017 \text{ m}^3 = 17000 \text{ cm}^3$$

This volume roughly corresponds to a 96.2 cm long cylinder of diameter 15 cm.

Angular position densitometer

In an angular position densitometer, the chamber contains three displacers as shown in Fig. 13.15. Displacers are made of materials of different densities and their volumes are also different from each other's. Attached to a common shaft, the displacers are fixed at angles $> 90^\circ$ between themselves. The whole assembly is free to rotate about the hub.

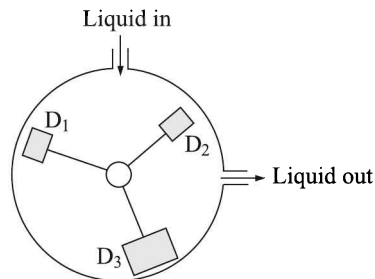


Fig. 13.15 Angular position densitometer.

A steady flow rate, ~ 2 L/min, of the sample liquid is maintained through the chamber. The angular position of the displacer assembly depends upon the balancing of moments acting on each displacer. The corresponding forces are resultants of their weights and buoyancies acting on each. Buoyancies, in turn, depend on the density of the liquid. So, if the density of the process liquid changes, the angular position of the displacer assembly will change.

Electromagnetic suspension densitometer

In an electromagnetic suspension densitometer, the float, fully immersed in process fluid, is kept suspended by a solenoid situated directly over the float (Fig. 13.16). Two search coils, symmetrically located near the float, keep track of the float position and give feedback to the solenoid current such that the float is always kept in position.

Density variation of the process fluid tends to alter the float position. This results in current variation in the solenoid as discussed earlier. So, the solenoid current can be calibrated in terms of density of the process fluid.

Full spans of this instrument range from 0.01 to 0.4 specific gravity units within the limits of 0.4 SG to 2.0 SG. Accuracy varies between 0.5% and 1% of full-span. Pressure and temperature limitations are 140 kPa and 170°C respectively.

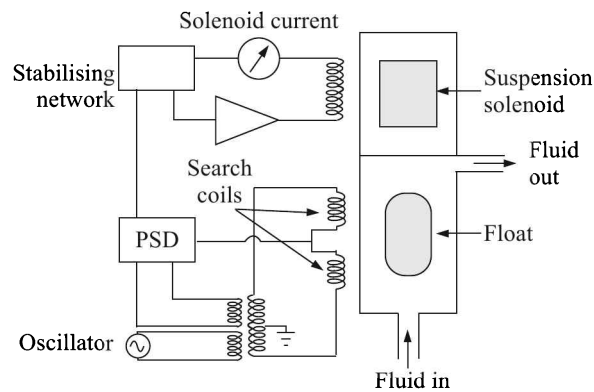


Fig. 13.16 Electromagnetic suspension densitometer.

Vibrating U-tube Densitometer

If a mass m is suspended from a spring of spring-constant k , then we know, when the mass is disturbed from its equilibrium position, it vibrates with a frequency f so that

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (13.18)$$

If the mass m comprises a solid shell of mass M with liquid of volume V and density ρ at the core as shown in Fig. 13.17, then

$$m = M + \rho V \quad (13.19)$$

Substituting the value of m from Eq. (13.19) in Eq. (13.18), we get on rearrangement of terms

$$\begin{aligned} \rho &= \frac{k}{4\pi^2 V f^2} - \frac{M}{V} = \frac{k}{4\pi^2 V} T^2 - \frac{M}{V} \\ &\equiv AT^2 - B \end{aligned} \quad (13.20)$$

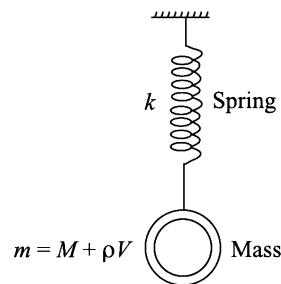


Fig. 13.17 Spring-suspended mass.

where T is the time period of oscillation, and A , B are constants. This principle is utilised in the measurement of fluid density by measuring the time period of natural vibration of a U-tube through which the fluid is allowed to flow. The schematic arrangement is shown in Fig. 13.18.

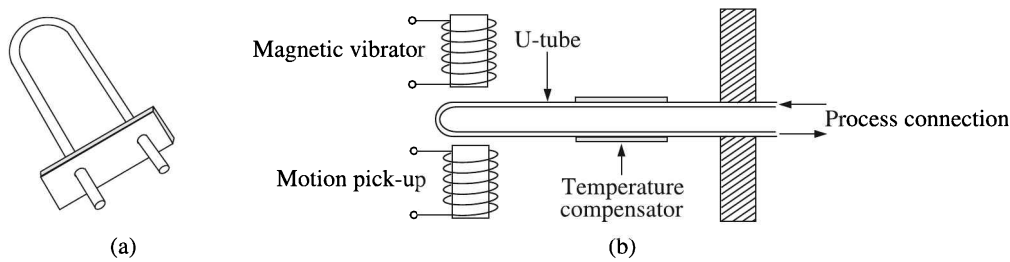


Fig. 13.18 Vibrating U-tube densitometer: (a) the U-tube, and (b) the set up.

The oscillation of the U-tube is maintained by an electromagnet fed with an alternating current. This current is fed back through a motion pick-up and amplifier so that the oscillation is maintained at the resonant frequency of the U-tube assembly. An RTD is attached to the U-tube for automatic temperature compensation.

The resonant frequency and temperature are constantly monitored and the density of the process fluid at the temperature is found out. Equation (13.20) suggests that the system can be calibrated by using two different fluids of known densities.

This densitometer is mainly used in brewing, soft drink, pharmaceutical and chemical industries where online high precision density measurements are called for. Densities of homogeneous liquids, light slurries and gases can also be measured by this method.

Weight-based U-tube Densitometer

The weight-based densitometer consists of a U-tube on bearing or flexures about the horizontal axis. Process fluid flowing through the tube is continuously weighed by a spring and nozzle-flapper displacement transducer. The arrangement is shown in Fig. 13.19.

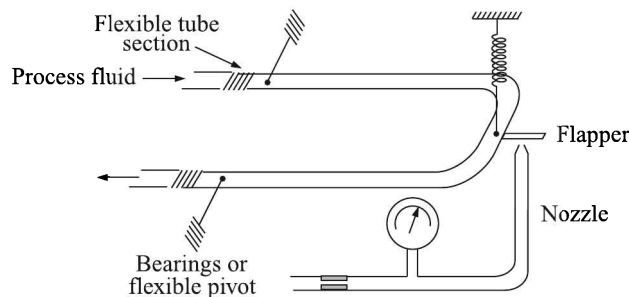


Fig. 13.19 Weight-based densitometer set up.

An increase in fluid density generates a proportional additional force in the U-tube which displaces the flapper of the nozzle-flapper transducer. So, the response of this transducer can be calibrated in terms of the fluid density.

Stainless steel, monel, nickel or even glass is used to construct the U-tube and the flexible coupling is made of neoprene, Teflon or silicone rubber. Full spans of the densitometer range from 0.02 SG to 0.05 SG within the limit of 3.5 SG. Measurement accuracy is 1% to 2% of the full-span.

Ultrasonic Densitometer

Ultrasonic densitometers are mainly used to measure density of sludge and slurry. The attenuation of an ultrasound pulse depends on the amount of suspended particles in its path as well as on the length of the path. An ultrasonic pulse is directed across a pipe section. Solid particles in the slurry scatter the sound beam and a weaker attenuated signal is received back by the crystal which acts both as the transmitter and receiver (Fig. 13.20).

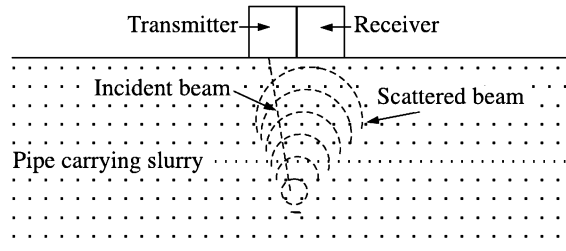


Fig. 13.20 Schematic diagram of the ultrasonic densitometer action.

The ratio of the emitted and received energy is related to the sludge density.

Ultrasound pulse cannot pass through air-bubbles or heavy solids. So, the success of the densitometer depends on their absence from the process fluid. These densitometers can measure densities with a precision of about 0.1% of the span in ranges between 0.5 SG and 1.5 SG.

Gamma-ray Densitometers

We know from Eq. (12.16) at page 511 that the strength or intensity of γ -radiation depends on the density of the material through which it passes. So, it can be used to determine density of a process fluid as shown in Fig. 13.21.

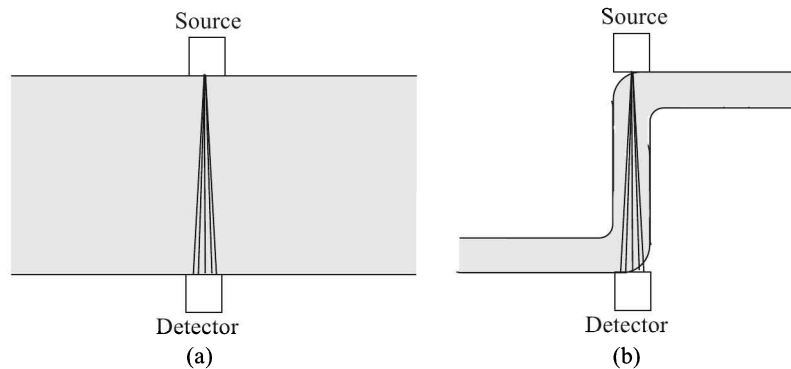


Fig. 13.21 γ -ray densitometer installation: (a) for pipe diameter > 150 mm, and (b) for lesser pipe diameter.

For pipes of diameter 150 mm or more, the source and the detector are mounted as shown in Fig. 13.21(a). For thinner pipes of lesser diameter, the mounting shown in Fig. 13.21(b) is adopted.

Source and safe dose. The commonly used source is ^{137}Cs of size between 200 and 2000 mCi⁶ depending on the pipe diameter and SG of the fluid. The design consideration is that the radiation field should not be more than 5 mR/h⁷ at a distance of 300 mm from the gauge. This dose is considered safe from health considerations.

Detection. The detector commonly used is the ionisation chamber⁸ which yields dependable results if it is provided with a constant power supply.

Limitations. γ -ray densitometers are not reliable if the change in radiation intensity is $< 5\%$. That eventually means that for a pipe of 50 mm diameter, the minimum span is 0.2 SG while for a 150 mm pipe it is 0.05 SG.

13.3 Conductivity Measurement

Conductivity measurement is a useful requirement for

1. Quality control of boiler feed-water
2. Quality control of drinking or process water
3. Estimation of total number of ions in a solution

Definitions

Conductivity is the ability of a material—gas, liquid or solid—to pass electric current through it. In that sense, it is the reciprocal of resistivity. The unit of resistance is ohms. Therefore, conductance unit is ohm^{-1} or mho or siemens (S). However, in some applications it is expressed in terms of *total dissolved solids* (TDS), which is related to conductivity by a factor that depends upon the level and type of ions present.

The conductivity of a solution depends upon the following factors:

1. Concentration of the solution
2. Mobility of ions
3. Valence of ions
4. Temperature of the solution

In general, we will be concerned with conductivity of liquids where the current carriers are positive and negative ions. All liquids possess some degree of conductivity. The range varies from 10^{-7} S/m (pure water) to greater than 1 S/m (concentrated chemical samples).

If two electrodes are dipped in a solution and a current is passed through them, some liquid molecules break to form ions that move to cathode (negative electrode) and anode (positive electrode) depending on the sign of their charge. This results in a current which means that the solution behaves as a conductor. Depending upon the formation of number of ions, solutions are classified as:

⁶Units used to specify radioactivity are curie (Ci) and becquerel (Bq). 1 g of $^{226}\text{radium}$ produces 3.7×10^{10} disintegrations per second. This rate of activity is by definition 1 Ci or 3.7×10^{10} Bq whether it is produced by radium or some other source.

⁷roentgen/h is the unit of radiation dose. A 1 Ci source will produce a dose of 1 R/h at a receiver placed 1 m away.

⁸See Section 14.10 at page 714.

Strong electrolytes. Strong acids, such as HCl, are strong electrolytes.

Weak electrolytes. Weak acids, such as acetic acid which only partially dissociates into acetate and hydrogen ions, are weak electrolytes.

Be it mentioned here that the mobility of dissimilar ions is different. Hence, the contribution to the resulting current may be different for cations and anions in the solution. However, if the mobility of one is much higher than the others, the latter's contribution may be neglected.

The conductance of a solution may be expressed as

$$L = \sigma \frac{A}{d}$$

where L is the conductance, in ohm^{-1} or mho or S

A is the area of the electrode, in cm^2

d is the distance between the electrodes, in cm

σ is the conductivity of the solution

In these units, the unit of conductivity is S/cm or $\text{ohm}^{-1}/\text{cm}$ or mho/cm. The ratio d/A is generally termed as *cell constant* θ and conductivity is expressed as

$$\sigma = \theta L$$

The cell constant θ is determined experimentally by measuring σ of known concentrations of KCl for which accurate ASTM⁹ data are available. A simplified conductivity measurement arrangement is shown in Fig. 13.22.

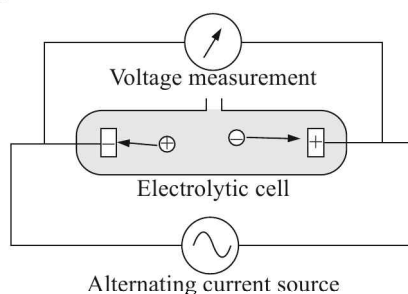


Fig. 13.22 Simplified conductivity measurement arrangement.

Conductivity Cells

Conductivity cells may be of two types:

1. Two-pole
2. Four-pole

Two-pole cell

In a two-pole cell, current feeding and voltage measurement are both done through the same terminals [Fig. 13.23(a)]. The aim is to measure the solution resistance R_s only. But polarisation¹⁰ of the solution and formation of ion clouds around electrodes cause electrode resistance R_e to build up. So, in this method R_e and R_s are measured together and they cannot be separated.

⁹American Society for Testing and Materials (www.astm.com).

¹⁰See *Sources of error in conductivity measurement* at page 546.

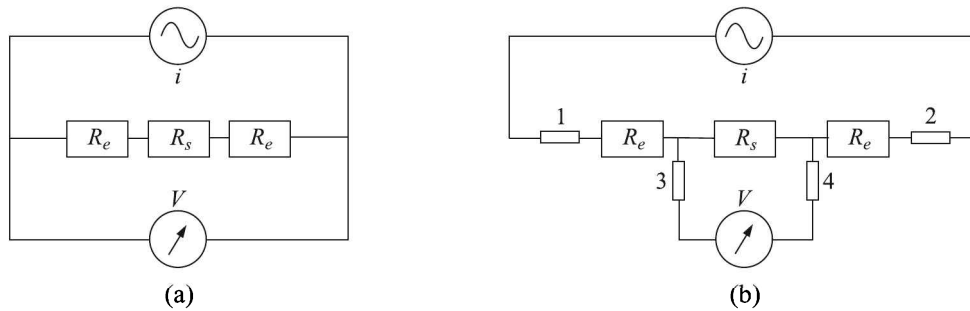


Fig. 13.23 Measurement of conductivity: (a) two-pole, and (b) four-pole.

Four-pole cell

In a four-pole cell [Fig. 13.23(b)], the current is applied to the outer electrodes (1, 2) such that a constant voltage is maintained between the inner electrodes (3, 4). The voltage measurement entails a negligible current passing through the voltmeter which has high input impedance. Therefore, there is negligible polarisation effect on the voltage measuring electrodes. As a result, the R_s value can be measured without any interference.

A simplified typical four-pole measurement arrangement is shown in Fig. 13.24. Current is adjusted so that the voltage remains around 200 mV. The rest of the arrangement is self-explanatory.

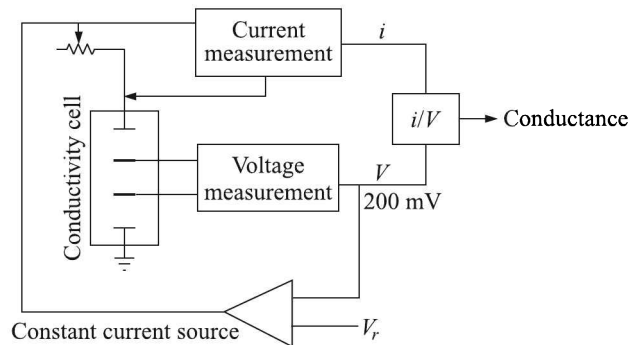


Fig. 13.24 Four-pole measurement arrangement.

Four-pole cells with an outer tube minimise the beaker wall effect because the measurement volume is well defined within the tube. Relative advantages of two- and four-pole cells are listed in Table 13.3.

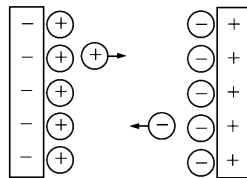
Sources of error in conductivity measurement

Polarisation. As an electric current is applied to an electrode dipped in a solution, oppositely charged ions start accumulating on its surface and have a tendency to form an ion cloud around it (Fig. 13.25).

This ion cloud, having the same polarity as the approaching ions, repels them to some extent. This phenomenon is called the *polarisation* and it gives rise to an electrode resistance R_e which interferes with the measurement of the solution resistance R_s .

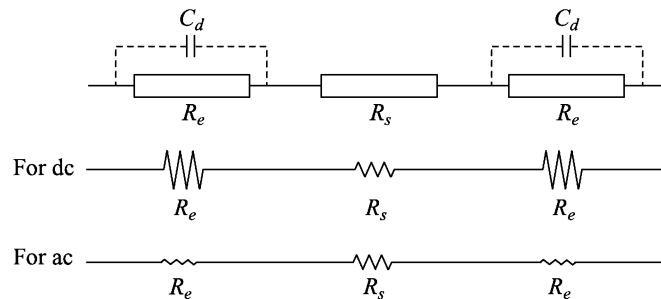
Table 13.3 Advantages and disadvantages of two- and four-pole cells

| Cell | Advantages | Disadvantages |
|-----------|--|--|
| Two-pole | <ol style="list-style-type: none"> 1. Easy to maintain 2. Cheaper 3. Recommended for viscous fluids or fluids with suspended matter | <ol style="list-style-type: none"> 1. Affected by polarisation 2. Affected by field effect 3. Offers limited accuracy because of nonlinearity of characteristic curve |
| Four-pole | <ol style="list-style-type: none"> 1. Ideal for high conductivity measurements 2. Linear over a large range 3. Polarisation and field effects are minimal | <ol style="list-style-type: none"> 1. Small sample measurement is not possible |

**Fig. 13.25** Ions collected around electrodes repel other incoming ions.

Polarisation effect can be minimised or eliminated by:

1. *Applying an alternating current:* As shown in Fig. 13.26, the measuring current will then flow through the double layer capacitance C_d of the electrode rather than being obstructed by the electrode resistance R_e .

**Fig. 13.26** Difference between application of dc and ac.

2. *Using platinised electrodes and optimising electrode areas:* A layer of platinum black on the electrodes increases their surface areas, thus decreasing the current density. This eventually reduces the polarisation effect. Scratching or damaging the platinum black coating, however, will change the cell constant because of modification of surface area of electrodes.
3. *Using a four-pole cell:* We have already discussed that a four-pole cell is almost free from effects of polarisation.

4. *Optimising the frequency of the current:* Low conductivity solutions (for $40 \mu\text{S}/\text{cm} < \sigma < 4 \text{ mS}/\text{cm}$) will give rise to low polarisation effect. Hence a low frequency ($\sim 100 \text{ Hz}$) supply may be used here. For higher conductivities of solutions (for $4 \text{ mS}/\text{cm} < \sigma < 2 \text{ S}/\text{cm}$), higher frequency ($\sim 45 \text{ kHz}$) may be used.

Field effect. Distortion of the electric field near the end of electrodes of a two-pole cell (Fig. 13.27) may interfere with the cell wall and cause further distortion. This is often called the *beaker wall effect*.

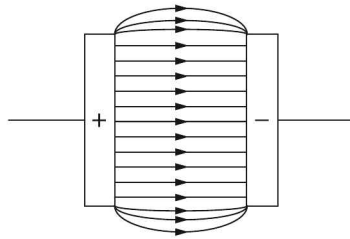


Fig. 13.27 Distortion of field near the ends of electrodes.

Such distortion will obviously affect the measurement where the potential difference between electrodes is assumed to be uniform. Four-pole cells, as discussed earlier, are not affected by the beaker wall effect. Two-pole electrodes may be placed centrally in the cell to minimise the error.

Cable resistance and capacitance. Cable resistance becomes a source of error for two-pole cells when the conductivity of the solution is high and therefore cell resistance is below 50Ω . Of course, the error may be exactly calculated by measuring the cable resistance. Four-pole cells are free from this error.

Cable capacitance, on the other hand, is a source of error in four-pole cells when measuring conductivities below $4 \mu\text{S}/\text{cm}$.

Variation with temperature. The conductivity of a solution increases with temperature. The reference temperature is usually 20°C or 25°C . Conductivity is measured at actual temperature T and then it is converted to reference temperature T_0 using the linear relation

$$\sigma_T = \sigma_0[1 + \phi(T - T_0)] \quad (13.21)$$

where σ_0 is the conductivity at T_0

σ_T is the conductivity at T

ϕ is the temperature coefficient.

The temperature coefficient ϕ is determined by measuring the conductivity σ_1 and σ_2 of the sample at two temperatures T_1 and T_2 , close to T_0 . Then

$$\sigma_1 = \sigma_0[1 + \phi(T_1 - T_0)] \quad (13.22)$$

$$\sigma_2 = \sigma_0[1 + \phi(T_2 - T_0)] \quad (13.23)$$

Subtracting Eq. (13.23) from Eq. (13.22) and on rearranging terms, we get

$$\phi = \frac{1}{T_1 - T_2} \cdot \frac{\sigma_1 - \sigma_2}{\sigma_0} \quad (13.24)$$

Substituting the expression for ϕ from Eq. (13.24) in Eq. (13.21) and on rearranging terms, we get

$$\sigma_0 \simeq \sigma_T \left[1 - \left(\frac{\sigma_1 - \sigma_2}{\sigma_0} \right) \cdot \left(\frac{T - T_0}{T_1 - T_2} \right) \right] \quad (13.25)$$

In Eq. (13.25), σ_0 occurs on both sides. The well-known iterative procedure of first substituting σ_1 on the right hand side and then converging on the true σ_0 value may be adopted for better accuracy in the measurement.

The approximate temperature coefficients of a few electrolytes are given in Table 13.4.

Table 13.4 ϕ values for a few electrolytes

| <i>Electrolyte</i> | ϕ (%/°C) |
|--------------------|---------------|
| Acids | 1.0–1.6 |
| Bases | 1.8–2.2 |
| Salts | 2.2–3.0 |
| Drinking water | 2.0 |
| Ultrapure water | 5.2 |

The conductivity values for a few liquids are given in Table 13.5.

Table 13.5 Conductivity values for a few liquids

| <i>Substance</i> | σ_0 (mS/cm) |
|-----------------------|--------------------|
| Pure water | 0.000055 |
| Deionised water | 0.001 |
| Rainwater | 0.05 |
| Drinking water | 0.005 |
| Industrial wastewater | 5.0 |
| Seawater | 50.0 |
| NaCl (1 mole/L) | 85.0 |
| HCl (1 mole/L) | 332.0 |

Toroidal (or Electrodeless) Conductivity Measurement

A toroid is a solenoid of finite length but bent into the shape of a doughnut [Fig. 13.28(a)].

The windings of the toroid uniformly encase the core in copper. This results in a natural magnetic screening effect which, in combination with the elimination of the air gap, results in a reduction of radiated magnetic field. The windings covering the solid ring core also help reduce magnetostriction—the main source of acoustic *hum* in solenoid winding.

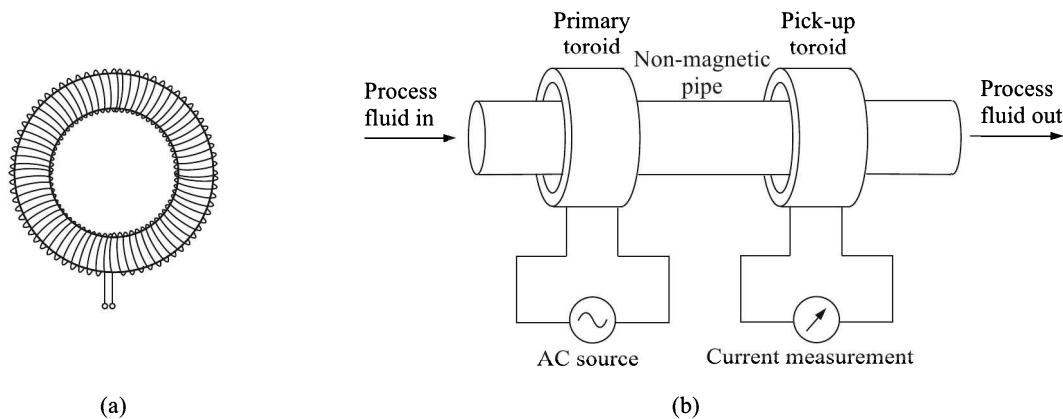


Fig. 13.28 Toroidal conductivity measurement: (a) toroid, and (b) set-up.

Electrodeless or toroidal conductivity measurement is done by passing an alternating current of about 20 kHz frequency through a primary toroidal coil, called the *driver coil*, which generates a strong magnetic field. As the liquid containing conductive ions passes through the hole of the coil, it acts as a *one turn secondary winding*. The passage of this fluid then induces a current proportional to the voltage induced by the magnetic field. The strength of current in the one turn secondary winding will depend on the concentration of ions in the liquid. This concentration of ions determines the ionic conductivity of the liquid.

The second toroid, called the *pick-up toroid* [Fig. 13.28(b)], is affected by the passage of the fluid in a similar fashion. The liquid passing through the second toroid acts as a *one turn primary winding* carrying the induced ac. This ac, in turn, creates a varying magnetic field which induces a current in the pick-up toroid. The induced current from the pick-up toroid is measured. The magnitude of current induced in the pick-up toroid is, therefore, proportional to the conductivity of the solution.

If the pipe carrying the process fluid is made of non-magnetic materials, they are placed outside of the pipe. But if the pipe material is magnetic, the toroids are placed inside, encapsulated in non-conductive, chemically resistant and temperature-stable materials, such as fluorocarbon polymers. As long as the twin-toroid sensor has a clearance of at least 3 cm, the proximity of pipe or container walls will have a negligible effect on the induced current.

Why toroid. An alternating current passed through a primary solenoid may generate current in the secondary (pick-up) coil due to

1. Electrical conduction through the liquid
2. Leakage of a varying magnetic field generated by the primary solenoid

Since we want to measure the current generated by the electrical conductivity of the solution and not the current directly induced in the pick-up by a fluctuating magnetic field of the driver coil, the toroidal, rather than solenoidal form of coil is used. A toroid has a minimum leakage of magnetic flux in the perpendicular direction of its plane.

Advantages and disadvantages

The advantages and disadvantages of toroidal conductivity measurement are as follows:

- | | |
|---|---|
| 1. Completely eliminates polarisation because no electrodes are used. | 1. Lacks the sensitivity of electrode-type measurement. |
| 2. Coils are not in contact with the solution. | 2. Toroids are bigger in size than electrodes which may cause problem in some applications. |

13.4 Oxidation-Reduction Potential (ORP)

It is well-known that when Zn or Cu electrodes are dipped in solutions of ZnSO_4 and CuSO_4 , separated by a porous membrane, an electric potential difference develops between the electrodes (Fig. 13.29). Such combinations are called *galvanic* or *voltaic* cells.

There is another kind of cells, called *electrolytic* cells, where a small voltage from an external source is applied to produce the desired electrochemical reaction.

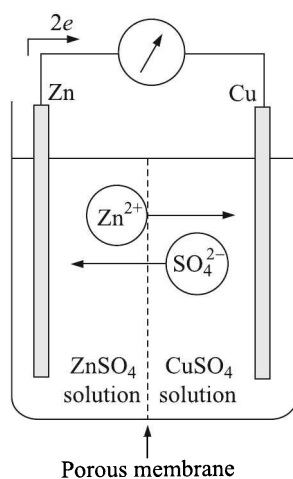
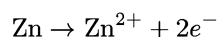


Fig. 13.29 Galvanic cell.

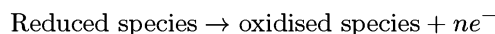
In the voltaic cell under consideration, Zn dissolves into the solution as Zn^{2+} ions and liberates 2 electrons. Written in equation form,



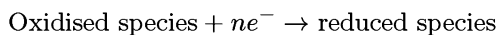
Liberated electrons travel along the external circuit to neutralise Cu^{2+} ions at the Cu electrode. The Cu^{2+} ions were produced owing to dissociation of CuSO_4 into Cu^{2+} and SO_4^{2-} ions.

Quite often the word *redox* is used to indicate the reactions involving the transfer of electrons. In a redox couple one component is oxidised by the removal of electrons and the other component is reduced by the addition of electrons.

In general, in a redox an electric potential develops between the electrodes. This potential is called the *electromotive force* (or, emf) to indicate that as if this force is driving electrons from the anode to the cathode. It is customary to visualise the cell reaction in terms of two half-reactions—an oxidation half-reaction and a reduction half-reaction—that can be symbolically written as follows:



| |
|--------------------|
| Oxidation at anode |
|--------------------|



| |
|----------------------|
| Reduction at cathode |
|----------------------|

The cell potential has a contribution from the anode which loses electrons and another contribution from the cathode which gains electrons. The former is called the *oxidation potential* and the latter the *reduction potential*.

Standard Electrode Potential

Thus, each cell can be thought of consisting of two half-cells:

1. Where electrons are evolved
2. Where electrons are used up

The potential of a half-cell is called *standard electrode potential* which is the difference of potential between a metal and the solution of its salt¹¹.

If we could measure the oxidation and reduction potentials of all available electrodes, then we could predict the emfs of voltaic cells created from the concerned pair of electrodes. In fact, tabulating one or the other is sufficient since the oxidation potential of a half-reaction is the negative of the reduction potential for the reverse of the reaction. But to do that we have to overcome two obstacles which stand in the way:

1. No electrode potential can be determined in isolation without having another electrode.
2. The electrode potential depends upon a number of factors such as the temperature, the pressure (in the case of a gas electrode) and the concentrations of substances.

The first of the obstacles is overcome by measuring potentials with respect to a *standard hydrogen electrode*¹²—the electrode potential of which is arbitrarily assigned a value of zero. The half-cell whose electrode potential is to be determined is combined with a standard hydrogen electrode to form a complete cell and the resulting emf is measured by a potentiometer.

The second obstacle is overcome by choosing the standard thermodynamic conditions for the measurement of potentials. These conditions are:

1. The solute concentration of 1 molar
2. The gas pressure of 1 atmosphere
3. The temperature of 25°

The standard electrode potential is denoted by a superscript of the degree sign. Thus,

$$E_{\text{electrode}}^{\circ} \left| \begin{array}{l} \text{is measured against the standard hydrogen electrode} \\ \text{at the concentration of 1 molar} \\ \text{at a pressure of 1 atmosphere, and} \\ \text{at the temperature of 25}^{\circ}\text{C.} \end{array} \right.$$

Table 13.6 lists values of reduction potentials of a few substances. The reduction potential is measured in volts (V), millivolts (mV), or Eh (1 Eh = 1 mV).

¹¹It is, in fact, the difference of Fermi levels of the metal and the solution.

¹²See Section 13.5 at page 563.

Table 13.6 Standard reduction potentials of a few substances

| <i>Cathode (reduction) half-reaction</i> | <i>Standard reduction potential E_r° (volt)</i> |
|--|---|
| $\text{Li}^+(\text{aq}) + e^- \rightarrow \text{Li}(\text{s})$ | -3.04 |
| $\text{K}^+(\text{aq}) + e^- \rightarrow \text{K}(\text{s})$ | -2.92 |
| $\text{Ca}^{2+}(\text{aq}) + 2e^- \rightarrow \text{Ca}(\text{s})$ | -2.76 |
| $\text{Na}^+(\text{aq}) + e^- \rightarrow \text{Na}(\text{s})$ | -2.71 |
| $\text{Zn}^{2+}(\text{aq}) + 2e^- \rightarrow \text{Zn}(\text{s})$ | -0.76 |
| $\text{Cu}^{2+}(\text{aq}) + 2e^- \rightarrow \text{Cu}(\text{s})$ | 0.34 |
| $\text{O}_3 + 2\text{H}^+(\text{aq}) + 2e^- \rightarrow \text{O}_2(\text{g}) + \text{H}_2\text{O}(\text{l})$ | 2.07 |
| $\text{F}_2(\text{g}) + 2e^- \rightarrow 2\text{F}^-(\text{aq})$ | 2.87 |

(s), (l) and (g) indicate solid, liquid and gas respectively.

(aq) indicates that the respective ion forms in aqueous solutions.

Many authors use standard oxidation potentials rather than reduction potentials in their calculations. These are simply the negative of standard reduction potentials as we have already mentioned. However, because these can also be referred to as *redox potentials*, the terms *reduction potential* and *oxidation potential* are preferred by the IUPAC¹³. The two may be explicitly distinguished in symbols as E_r° and E_o° .

Voltaic Cell Potentials

We have seen that a voltaic cell is created when an electrochemical cell is arranged with two half-reactions separated by an electrically conducting path. The maximum voltage or emf that will be produced between the electrodes of the cell is determined by their standard electrode potentials.

Let us consider, for example, the well-known Daniel cell where zinc and copper constitute the electrodes. The data for the standard electrode potentials are:

| <i>Cathode (reduction) half-reaction</i> | <i>Standard reduction potential E_r°</i> |
|--|--|
| $\text{Zn}^{2+}(\text{aq}) + 2e^- \rightarrow \text{Zn}(\text{s})$ | -0.76 |
| $\text{Cu}^{2+}(\text{aq}) + 2e^- \rightarrow \text{Cu}(\text{s})$ | 0.34 |

The standard cell potential is found from the relation

$$E_{\text{cell}}^\circ = E_{\text{cathode}}^\circ - E_{\text{anode}}^\circ \quad (13.26)$$

where E_{cathode}° is the standard reduction potential of the substance reduced, and

E_{anode}° is the standard reduction potential of the substance oxidised.

The data given above are reduction potentials. So, from Eq. (13.26) we find that the standard cell potential, or emf, for the Daniel cell is 1.10 V which is what we measure in standard conditions.

¹³International Union of Pure and Applied Chemistry www.iupac.org.

Nernst Equation

We saw how to calculate the cell potential for a voltaic cell from standard electrode potentials under standard conditions. But under real conditions, that are different from the standard conditions, the cell potential can be calculated from what is known as the Nernst¹⁴ equation.

$$E_{\text{cell}} = E_{\text{cell}}^{\circ} - \frac{RT}{nF} \ln Q \quad (13.27)$$

where E_{cell} is the actual cell potential

E_{cell}° is the standard cell potential

T is the absolute temperature (298 K at 25°C)

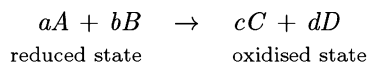
R is the gas constant (8.3144 J·mol⁻¹·K⁻¹)

F is the faraday (96485.3 coulomb·mol⁻¹)

n is the number of electrons transferred in the cell reaction

Q is the thermodynamic reaction quotient.

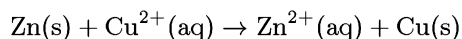
The quantity Q is like a dynamic version of the equilibrium constant in which the concentrations and gas pressures are the instantaneous values in the reaction mixture. Suppose the redox reaction is given by



Then the reaction quotient can be written as

$$Q = \frac{[C]^c [D]^d}{[A]^a [B]^b}$$

where $[C]$ indicates the molar concentration (or the partial pressure in atmospheres if gas) of the component C . Consider the example of the Daniel cell where the redox reaction is



Here,

$$Q = \frac{[\text{Zn}^{2+}]}{[\text{Cu}^{2+}]} \quad (13.28)$$

because the values of concentrations of pure metal solids are taken as 1.

Variation with concentration and temperature

From Eqs. (13.27) and (13.28) we find, since other factors are constant, the actual cell potential differs from the standard cell potential if the temperature and ion concentrations vary.

It is clear from these equations that the more the concentration of Cu^{2+} , the higher the actual cell potential while an excess concentration of Zn^{2+} will result in a lower cell potential. Generally, the concentration of Cu^{2+} is higher which implies that the second factor on the right-hand side of the Nernst equation is positive. Then, an increase in the cell temperature will increase the cell potential—a commonly observed phenomenon with dry cell batteries.

¹⁴Walther Hermann Nernst (1864 – 1941) was a German physical chemist and physicist who is known for his theories behind the calculation of chemical affinity and the third law of thermodynamics, for which he won the Nobel Prize in chemistry in 1920.

Nernst equation for half-cell potentials

A little altered form of the Nernst equation is applicable to calculate half-cell potentials as well. The equation for an oxidation reaction is given by

$$E_o = E^\circ - \frac{2.303 RT}{nF} \log[\text{ion}] \quad (13.29)$$

where E_o is the single electrode oxidation potential

E° is the standard oxidation potential On

$\log[\text{ion}]$ is the logarithm₁₀ of the concentration of the concerned ion

substituting the values of R , $T = 298$ K and F , Eq. (13.29) becomes

$$E_o = E^\circ - \frac{0.0592}{n} \log[\text{ion}] \quad (13.30)$$

For a reduction reaction, the equation becomes

$$E_r = -E^\circ + \frac{0.0592}{n} \log[\text{ion}]$$

Measurement of ORP

The measurement of ORP in aqueous solutions is straightforward through potentiometry. But its use in theoretical interpretations is fraught with problems like slow electrode kinetics, process of multiple redox couples, electrode poisoning, etc. Nevertheless, the ORP measurement has proven useful as a tool for monitoring changes in a chemical process rather than determining their absolute values.

In an ORP measurement, it is necessary to use a reference electrode, an indicator electrode and a voltage measuring arrangement having high input impedance.

Reference electrodes

Commonly used reference electrodes are (see Section 13.5 at page 564):

1. Ag/AgCl electrode
2. Calomel electrode

Indicator electrodes

Indicator electrodes fall into two classes:

1. Metallic electrodes
2. Membrane electrodes

Metallic electrodes. Depending on their response, metallic electrodes can be divided into four types as follows:

| <i>Type</i> | <i>Constitution</i> | <i>Example</i> | <i>Comments</i> |
|-------------|---|--|--|
| 1 | A metal wire, mesh or strip that responds to its own cation in solutions. | Cu/Cu ⁺ , Pb/Pb ⁺ , Hg/Hg ⁺ and Ag/Ag ⁺ . Ag and Hg are most commonly used. | Suffer from poor selectivity, responding not only to their own cation but also to any other more easily reduced cation. |
| 2 | A metal either coated with, or immersed in one of its sparingly soluble salts. The electrode responds to the anion of the salt. | A silver wire coated with AgCl which responds to changes in chlorine activity. | |
| 3 | Uses two equilibrium reactions to respond to a cation other than that of the metal electrode. | A mercury electrode in a solution containing EDTA ¹⁵ and Ca. It responds to the Ca ion activity. | |
| 4 | Noble or inert metal electrode. | Platinum, palladium, gold or other inert metals that serve to measure redox reactions for species in solution e.g. Fe ²⁺ /Fe ³⁺ , Ce ³⁺ /Ce ⁴⁺ . | Called the <i>redox indicator electrode</i> , these electrodes are often used to detect endpoint in potentiometric titrations. |

Membrane electrodes. Membrane electrodes are a class of electrodes that respond selectively to ions or molecules by the development of a potential difference across a membrane that separates the analyte solution from a reference solution. The potential difference is related to the concentration difference in the specific ion measured on either side of the membrane.

How a membrane electrode works is sometimes explained in the following way. Membrane electrodes always have a filling solution sealed inside. The filling solution formula contains ions to which the membrane is selective. If there is a difference in activity of these ions on the two sides of the membrane, ions will enter the membrane from the side where the activity is higher, and ions will exit the membrane on the other side. Because ions carry a charge, a potential difference will develop across the membrane.

The general scheme of the measurement with a membrane electrode is shown in Fig. 13.30. To make a measurement, a second unvarying potential against which the membrane potential may be compared, is required. The second reference electrode provides this function. A standard filling solution completes the electrical circuit between the sample and the internal cell of the reference electrode. The filling solution is typically potassium chloride, saturated with silver chloride, and a salt of the ion to which the ISE responds.

Membrane electrode tree looks like Fig. 13.31.

Ion selective membrane electrodes. For their selectivity of dissolved ions in the analyte, these electrodes are often referred to as *ion selective electrodes (ISE)*. Their three variations are

¹⁵E(thylene) d(iamine) t(etraacetic) a(cid)—a crystalline acid that acts as a strong chelating agent.

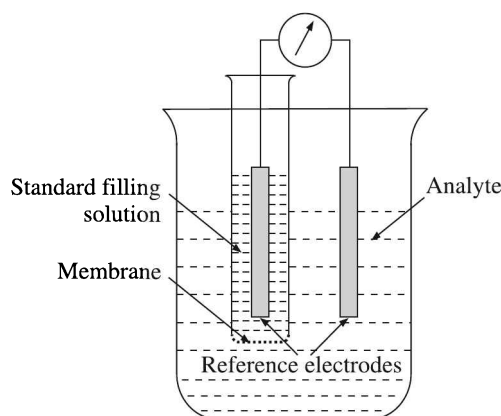


Fig. 13.30 Scheme of measurement with a membrane electrode.

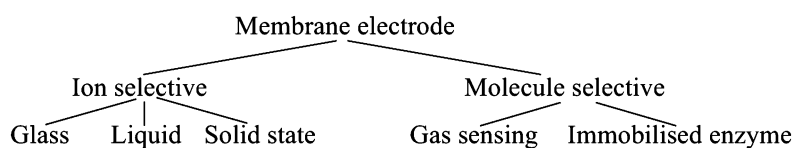


Fig. 13.31 Membrane electrode tree.

1. Glass electrode
2. Liquid membrane electrodes
3. Solid state electrodes

Glass electrodes. The glass electrode uses an ion selective glass membrane in its construction. Commercial glass electrodes respond strongly to +1 ions, including H^+ . The +1 ions that can be measure potentiometrically are Li^+ , K^+ , Na^+ , H^+ , Ag^+ and NH_4^+ .

The essential elements of a glass electrode are schematically shown in Fig. 13.35(b). The potential of this electrode is controlled by the difference between the hydrogen ion concentrations inside and outside the thin glass membrane at the bottom. The H^+ concentration inside the electrode being constant, the electrode potential varies only with the concentration of H^+ in the solution outside.

Liquid membrane electrodes. A liquid membrane electrode contains a water-soluble organic liquid that is capable of transporting some specific ion across the boundary which is made of an organophilic¹⁶ membrane. Liquid membranes are made to measure Ca^{2+} , Mg^{2+} , Cu^{2+} , Pb^{2+} , Cl^- , NO_2^- and ClO_4^- .

The essential elements in a liquid ion exchange electrode are shown schematically in Fig. 13.32(a).

The main difference between liquid membrane electrodes and glass electrodes is that in the former, reactive sites are mobile and able to travel to the ion while such sites are fixed in the latter.

¹⁶-*phile* (also *-phil*) denoting fondness for whatever specified. See *Pocket Oxford Dictionary*, Oxford University Press (1996).

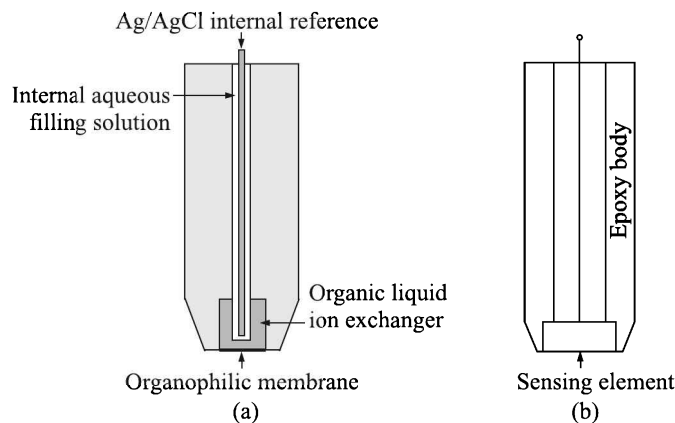


Fig. 13.32 Essential elements of (a) liquid membrane electrode, and (b) solid state membrane electrode.

Solid state membrane electrodes. The essential elements in a solid state membrane electrode are shown in Fig. 13.32(b). Here, the sensing element is a crystal or a pellet pressed from crystalline material. For example, the chloride sensing membrane consists of a membrane containing silver chloride pressed in the form of a pellet onto a substrate of silver metal and sealed into an inert body so that only the AgCl surface is exposed. The fluoride sensing membrane, however, is made from a single crystal of lanthanum fluoride.

Like the glass membrane electrodes, the solid state membrane electrodes have fixed reactive sites that cannot travel to the ions measured. The solid state devices include electrodes for chloride, bromide, iodide, sulphide, silver, lead, copper, and cadmium.

Molecule sensitive membrane electrodes. As shown in the membrane tree, these electrodes belong to two categories:

1. Gas sensing electrodes
2. Immobilised enzyme electrodes

Gas sensing electrodes. These electrodes help measure the concentration of gases like carbon monoxide, ammonia, hydrogen sulphide and nitrite (NO_2^-) salts dissolved in aqueous solutions, or the concentration of ions in solution that can be converted to a dissolved gas by a simple chemical reaction. The diagram of a typical gas sensing electrode is shown in Fig. 13.33(a).

Made of a hydrophobic¹⁷ porous plastic, the membrane prevents water from entering its pores or passing through it. The gas in the sample solution diffuses through the membrane and chemically reacts with some substance in the electrode to form ions. These ions are detected by the ion selective electrode. The emf developed between the ISE and an internal reference electrode is measured.

Immobilised enzyme membrane electrodes. Enzymes are biocatalysts that are highly selective for complex organic molecules of biochemical interest, such as glucose. Many enzymes catalyse reactions that produce ammonia, carbon dioxide and other simple species for which ISEs are available as detectors. Coating an ISE with a thin layer of an enzyme

¹⁷Meaning *something that repels water*.

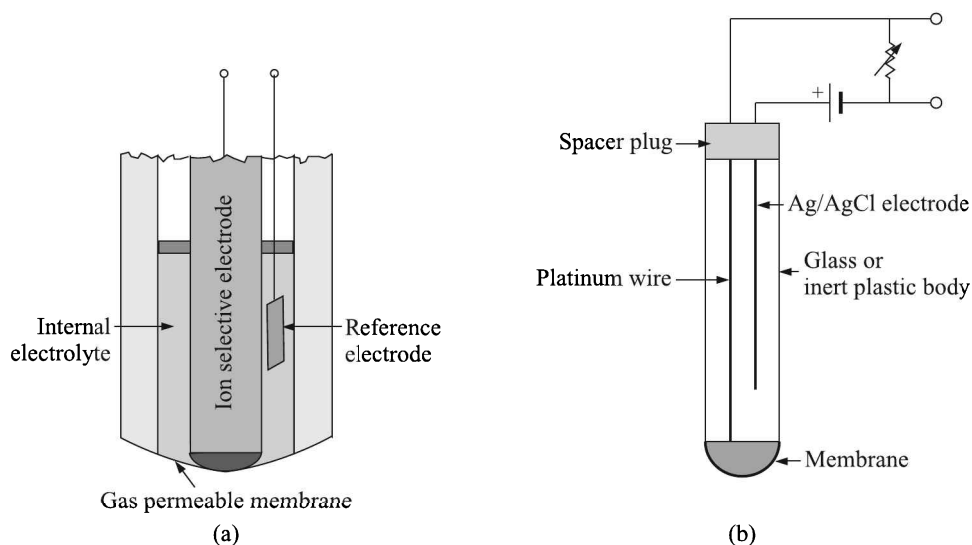


Fig. 13.33 Schematic diagrams of: (a) gas sensing membrane electrode, and (b) dissolved oxygen analysis.

holds the promise of making highly selective biosensors for complex molecules with an advantage of potentiometry, speed, low cost and simplicity. The enzyme is generally immobilised on the electrode by incorporation into a gel, or by covalent bonding to a polymer support or by direct adsorption onto the electrode surface, so that a small amount of enzyme may be used for many measurements.

However, no commercial potentiometric enzyme based electrodes are available so far.

The ORP measurement is useful in biochemical studies as well. Many enzymatic reactions are oxidation-reduction reactions in which one compound is oxidised and the other is reduced. The ability of an organism to carry out oxidation-reduction reactions depends on the oxidation-reduction state of the environment, or its reduction potential E_r . Strictly aerobic¹⁸ microorganisms can be active only at positive E_r values, whereas strict anaerobes can be active only at negative E_r values. Redox affects the solubility of nutrients, especially metal ions.

Let us now work out a few examples to gain an insight into the redox.

Example 13.4

Calculate the following half-cell potentials:

(a) $\text{Ag}^+ (1 \times 10^{-4} \text{ M}) | \text{Ag}$, given $E^\circ = 0.799$.

(b) $\text{Fe}^{3+} (1 \times 10^{-5} \text{ M}), \text{Fe}^{2+} (0.1 \text{ M}) | \text{Pt}$, given $E^\circ = 0.769$.

¹⁸An aerobic organism or aerobe is an organism that has an oxygen based metabolism. Aerobes, in a process known as cellular respiration, use oxygen to oxidise substrates (e.g. sugars and fats) in order to obtain energy.

Solution

Using Eq. (13.30) we get

$$(a) \quad E = 0.799 - (0.0592) \left(\log \frac{1}{1 \times 10^{-4}} \right) = 0.562 \text{ V}$$

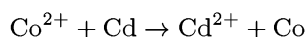
$$(b) \quad E = 0.769 - (0.0592) \left(\log \frac{0.1}{1 \times 10^{-5}} \right) = 0.532 \text{ V}$$

Example 13.5

What is the standard potential for a cell containing Cd and Co? Write the reaction. Given, the standard reduction potentials of Cd^{2+} and Co^{2+} are -0.4022 V and -0.277 V respectively.

Solution

From the given standard reduction potentials, it is apparent that Co^{2+} is reduced and Cd is oxidised. So, the reaction is



Therefore, from Eq. (13.26) we get

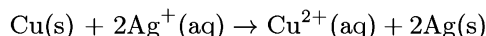
$$E_{\text{cell}} = -0.277 - (-0.4022) = 0.125 \text{ V}$$

Example 13.6

For measuring Cu^{2+} concentration in water sample, the concerned ORP cell consisted of a silver electrode dipped into 1 M AgNO_3 solution connected by a salt bridge to another half-cell consisting of copper electrode dipped into the sample water. The measured cell potential with the copper serving as anode was 0.62 V at 25°C . Find the concentration of Cu^{2+} in the sample. Given, standard reduction potentials of Ag^+ and Cu^{2+} are 0.80 V and 0.34 V respectively.

Solution

Copper was oxidised because it served as anode. Therefore, Ag^+ was reduced. So, the redox reaction was



Given,

$$E_{\text{cell}} = 0.62\text{V}, \quad E_{\text{cathode}}^\circ = 0.80\text{V}, \quad E_{\text{anode}}^\circ = 0.34 \text{ V}, \quad n = 2$$

Therefore,

$$E_{\text{cell}}^\circ = 0.80 - 0.34 \text{ V} = 0.46 \text{ V}$$

So, we get from Eq. (13.27)

$$E_{\text{cell}} = E_{\text{cell}}^\circ - \frac{0.0592}{2} \log \frac{[\text{Cu}^{2+}]}{[\text{Ag}^+]^2} = 0.46 - 0.0296 \log \frac{[\text{Cu}^{2+}]}{(1)^2}$$

or

$$\log \frac{[\text{Cu}^{2+}]}{1} = \frac{0.62 - 0.46}{-0.0296} = -5.4$$

Hence

$$[\text{Cu}^{2+}] = \text{antilog}(-5.4) = 3.98 \times 10^{-6} \text{ M}$$

Example 13.7

The following values were obtained while calibrating a redox cell for fluoride concentration measurement:

| | | | | | |
|-------------------------------|-----|-----|-----|-----|-----|
| <i>Concentration C</i> (ppm) | 0.2 | 0.5 | 0.8 | 1.0 | 2.0 |
| <i>Redox potential V</i> (mV) | 262 | 248 | 234 | 224 | 210 |

If a sample gives a reading of 250 mV, what is the corresponding fluoride concentration?

Solution

Through a least square fit of the given data, we get the linear equation

$$V = -54.4 \log C + 229$$

Substituting the measured value of 250 mV to this equation, we get

$$\begin{aligned} 250 &= -54.4 \log C + 229 \\ \therefore C &= \text{antilog} \frac{250 - 229}{-54.4} = \text{antilog}(-0.386) \\ &= 0.411 \text{ ppm} \end{aligned}$$

Dissolved Oxygen Analysis

With the development of selectively permeable membranes, it has become possible to construct specific electrodes for analysis of oxygen dissolved in liquids.

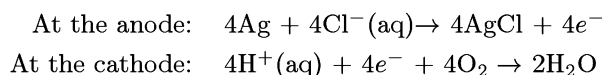
An oxygen sensing electrode operates on the basis that oxygen gas in solution reacts with a negatively charged (polarised) metal surface and forms OH^- radicals. Platinum is a strong catalyst for the covalent dissociation or recombination of water. In the oxygen sensing electrode, electrons 'boil' off the platinum electrode, combining with dissolved molecular oxygen and hydrogen ions to produce water. The rate at which electrons boil off is proportional to the concentration of oxygen that is available to 'grab' them. The movement of electrons generates an electrical current, which is then converted to a voltage. A schematic design of the arrangement is shown in Fig. 13.33(b).

The platinum cathode is formed from a 10 μm to 25 μm wire. Normally the silver electrode is silver wire with a diameter larger than that of platinum. The silver wire is actually an Ag/AgCl reference electrode and is chlorinated before it is used for measurement. Typical membranes are PVC, mylar, cellulose acetate, polyethylene, polypropylene, silicone rubber and ethyl cellulose. For blood oxygen measurements, Eastman polypropylene has been found to be the most satisfactory. A concentrated potassium chloride solution is held in place over the surfaces of the electrodes by a Teflon membrane which is attached by an O-ring that surrounds the electrodes.

The oxygen electrode is polarographic¹⁹ and in operation requires a polarising potential of 0.2 V to 0.9 V, with negative bias to the platinum electrode. Under that condition, the oxygen reaching the platinum is reduced electrolytically. The basic reaction that takes place is as follows.

¹⁹See Section 13.6 at page 571.

Current flows from the silver electrode to the platinum electrode as electrons boil off into solution from the latter. Removal of electrons from the solid silver produces silver ions. The silver ions combine with chlorine ions in solution to precipitate silver chloride on the surface of the silver electrode. This leaves potassium ions behind. However since hydrogen ions are taken out of solution by the consumption of oxygen, the charge remains balanced. Written in equation form:



The electrode response time is rather slow. It is not unusual for the initial response time to be as long as 1 min since it takes time for the oxygen to diffuse through the membrane. In general, the smaller the cathode tip, the faster the response.

13.5 pH Measurement

Apart from the academic interest of finding the chemical characteristics of a solution, pH measurement is an essential first step toward managing chemical reactions. Nearly all industries that deal with water—not merely chemical industries, but agriculture- and fishery-related industries, biological industries, public organisations—need pH measurement at some stage or other. Table 13.7 will give an idea how the pH measurement is useful in many industries.

Table 13.7 Usefulness of pH measurement in industries

| <i>Industry</i> | <i>Importance of pH measurement</i> |
|---|---|
| Textiles, dyeing, paper and pulp industries | Important in product testing. |
| Oil refining | Necessary in desulphurisation process. |
| Metallurgy | When extracting a particular material from crude ore or mixed metal, pH is controlled so as to extract only the desired metal without dissolving the slag. |
| Electrochemical industries | pH control is necessary in plating, etching of metal surfaces and the manufacture of batteries. For example, without proper control of the pH of the plating solution, the finished plating will lack the optimum colour and lustre or is likely to peel off. |

Definition of pH

The term pH^{20} was coined by Sørensen²¹ in 1909 in order to express the very small concentrations of hydrogen ions. Essentially, it expresses the degree of acidity or alkalinity of a solution. The degree, determined by the hydrogen ion concentration $[\text{H}^+]$, is expressed mathematically²² as

²⁰An abbreviation of *pondus Hydrogenii* meaning *power of hydrogen*.

²¹Søren Peder Lauritz Sørensen (1868 – 1939) was a Danish chemist, famous for the introduction of the concept of pH.

²²We will use log to indicate \log_{10} .

$$\text{pH} = \log \frac{1}{[\text{H}^+]} = -\log[\text{H}^+]$$

For example, if $[\text{H}^+] = 1 \times 10^{-4}$ g-ions/L, $\text{pH} = 4$. Pure water ionises to a small degree to form H^+ and OH^- ions as expressed by the chemical equation



The ionic product of concentrations of H^+ and OH^- ions is taken as 1×10^{-14} g-ions/L. That means, the concentration of H^+ and OH^- ions in pure water is 1×10^{-7} g-ions/L each. Therefore, the pH of pure water is 7. Solutions having $\text{pH} < 7$ are acidic while those having $\text{pH} > 7$ are alkaline. Obviously, pH value will lie between 0 and 14. pH of a few common substances are given in Table 13.8.

Table 13.8 pH of a few common substances

| <i>Substance</i> | <i>pH</i> | <i>Substance</i> | <i>pH</i> |
|------------------|-----------|------------------|-----------|
| Gastric acid | 1 | Sea water | 8 |
| Lemon juice | 2 | Baking soda | 9 |
| Orange juice | 3 | Milk of magnesia | 10 |
| Tomato juice | 4 | Ammonia solution | 11 |
| Black coffee | 5 | Soapy water | 12 |
| Human urine | 6 | Bleach | 13 |
| Distilled water | 7 | | |

Buffer Solution

Although pH of pure water is 7, even the purest form of water fails to retain the value for a long time owing to solution of aerial CO_2 or silicates from glass container into it. However, it has been observed that a solution of a salt with a strong base of a weak acid into the acid itself—such as CH_3COONa in CH_3COOH (acetic acid), or a salt with a strong acid of a weak base into the base itself—such as NH_4Cl in NH_4OH , has a definite pH value that does not alter either with time or on dilution. The pH value slightly alters if a strong acid is added to the former solution or a strong base to the latter. Such solutions which more or less preserve their pH values are called ‘buffer solutions’.

To measure the pH of a solution, a reference electrode and a measuring electrode are required. We will first consider a few reference electrodes and then move on to the actual pH measurement.

Reference Electrodes

Standard hydrogen electrode

A small strip of platinum, coated with platinum black to absorb hydrogen gas, forms the electrode. It is half-immersed in a solution of hydrogen ions at unity activity (1.228 M^{23} HCl at 25°C) and half in pure hydrogen gas at 1 atm pressure [Fig. 13.34(a)].

²³1 M = 1 mole per litre.

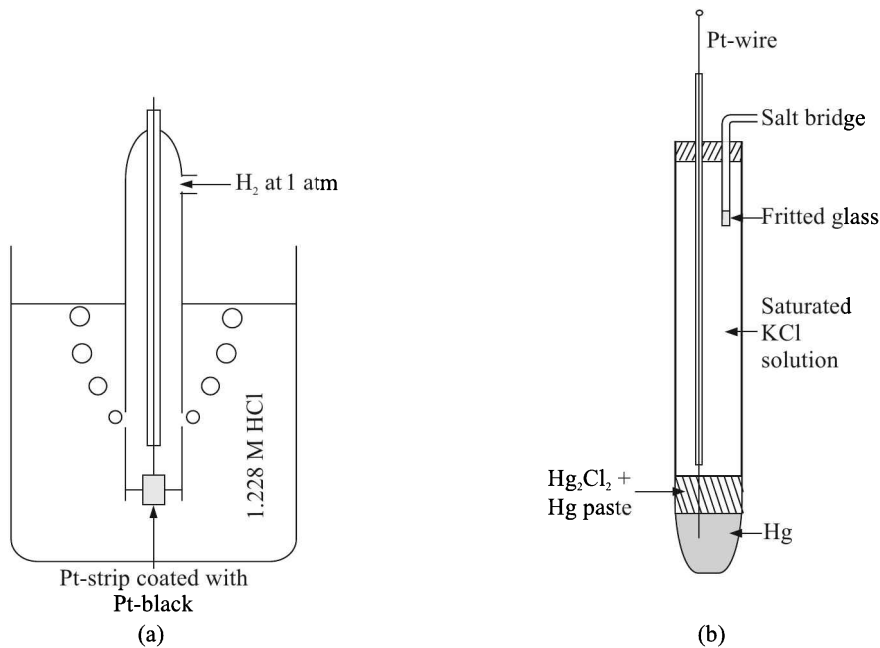


Fig. 13.34 (a) Standard hydrogen electrode, and (b) calomel electrode.

A part of the gas is adsorbed by the platinised electrode while the excess escapes through holes in the glass envelope.

Setting up a hydrogen electrode is inconvenient for all practical purposes. Hence, secondary standard electrodes such as calomel electrode and silver/silver-chloride electrode have come into existence.

Calomel electrode

Calomel electrode [Fig. 13.34(b)] contains a pool of mercury at the bottom of a glass tube. On top of mercury, there remains a paste of mercury and mercurous chloride (calomel) in potassium chloride solution. The electrolyte is also a solution of potassium chloride. A platinum wire, which may be amalgamated, maintains contact with the mercury pool.

The calomel electrode serves as a secondary standard reference electrode. The salt bridge²⁴ serves as a connector between the reference electrode and the measuring (or indicator) electrode. It consists of KCl solution of concentration 3.8 M with excess gelatine and plugged at both ends with a porous fritted glass stopper.

The oxidation potential of the Hg/Hg_2Cl_2 electrode is -0.2415 V for saturated KCl solution at $25^\circ C$.

Silver/silver chloride electrode

Silver/silver chloride electrode has the advantage that it is (i) reversible, (ii) stable, and (iii) it can be combined with cells containing chlorides without inserting salt bridges.

²⁴Connecting tube containing liquid electrolyte [see Fig. 13.36(a)]

It consists of a silver electrode coated with silver chloride and dipped in a saturated solution of potassium chloride and silver chloride [Fig. 13.35(a)]. A liquid junction²⁵ separates it from the process fluid.

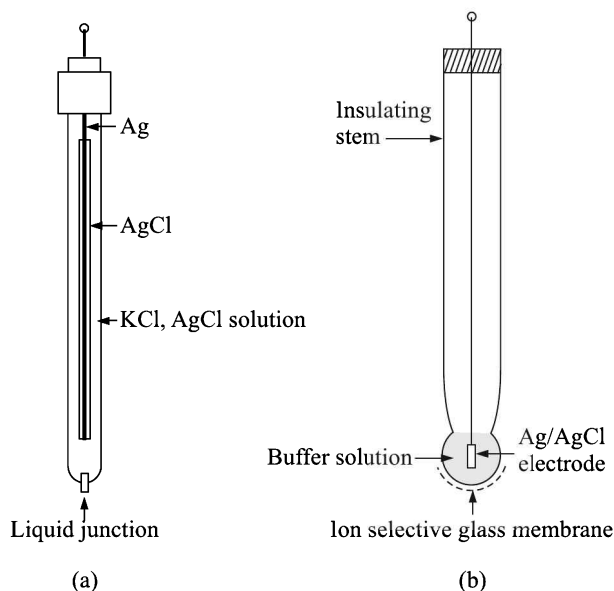


Fig. 13.35 (a) Ag/AgCl electrode, and (b) glass electrode.

The Ag/AgCl electrode acts as a secondary standard reference electrode having a standard potential of -0.2224 V at 25°C .

Measurement (or Indicator) Electrodes

Glass electrode

The glass electrode is the most commonly used measurement electrode. It consists of an electrode membrane that responds to pH, a highly insulating base material to support the unit, an internal buffer solution of known pH (like KCl at $\text{pH} = 7$), an internal electrode (either Ag/AgCl or Hg/Hg₂Cl₂ type), a lead wire and an electrode terminal [Fig. 13.35(b)].

The most critical item in the electrode is the ion-selective glass membrane. Its characteristics should be as follows:

1. To remain practically impermeable to all other species of ions except the one for which it is being used.
2. Though accurately sensitive to alkalinity/acidity, it must not get damaged by either.
3. Its electric resistance must not be very high.
4. It must not generate too large a potential difference²⁶ between the solutions inside and outside the electrode when the electrode is dipped in a solution inside the electrode (i.e. the buffer solution).

²⁵Fine capillary plugged with asbestos fibre.

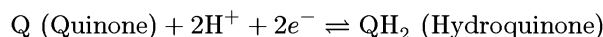
²⁶Called *asymmetry potential*.

5. It must be resistant to shock and chemical reactions.

Glass, containing lithium, which is chemically-strong and has low electric resistance, fits the bill. Glass electrodes selective to Na, K, NH₄ or Ag ions are available commercially.

Quinhydrone electrode

Quinone and hydroquinone form a reversible redox system in the presence of hydrogen ions



When an inert electrode, such as platinum, is immersed in the solution, the emf E developed is given by the Nernst equation as

$$\begin{aligned} E &= E^\circ + \frac{2.303RT}{nF} \log \frac{[QH_2]}{[Q][H^+]^2} \\ &= E^\circ + \frac{2.303RT}{nF} \left\{ \log \frac{[QH_2]}{[Q]} - 2 \log[H^+] \right\} \end{aligned} \quad (13.31)$$

The ratio $[QH_2]/[Q]$ is maintained at 1 by saturating the solution with quinhydrone which is 1:1 molar compound of quinone and hydroquinone. Then, Eq. (13.31) becomes

$$E = E^\circ + 0.0591\text{pH}$$

The standard oxidation potential of a quinhydrone electrode is -0.6994 V at 25°C . The advantages and disadvantages of quinhydrone electrode are given in Table 13.9.

Table 13.9 Advantages and disadvantages of quinhydrone electrode

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|---|
| 1. The electrode can be quickly set up by adding a pinch of quinhydrone to the solution and dipping a clean platinum wire into it. | 1. It cannot be used with Fe^{2+} , MnO_2 , aniline, etc. which react with quinone or hydroquinone. |
| 2. Indicated pH values are accurate even in the presence of interfering oxidising ions. | 2. Hydroquinone begins to oxidise at $\text{pH} = 8.5$, and therefore, it cannot be used for alkalis having $\text{pH} \geq 8.5$. |
| 3. Measurements can be made with small quantity of the sample liquid. | |

Antimony electrode

The tip of a polished antimony rod can be immersed into the test liquid to act as a measuring electrode. The electrode responds to pH by surface oxidation.

Before the advent of glass electrodes, the method was widely used because the electrode is sturdy and easy to handle. But it requires periodic cleaning because oxidant accumulation on its surface deteriorates the accuracy. The electrode's response is nonlinear, and therefore, it requires to be calibrated at narrow spans. Its measurement range is 3 to 11 and it is very temperature sensitive.

Methods of Measurement

Methods for measuring pH can be divided into two categories:

1. Indicator methods
2. Electrode methods

Indicator methods

This category can be sub-divided into two procedures:

1. An indicator solution is added to the test liquid using buffer solution. The resulting colour is compared with that of a known pH.
2. pH test papers are prepared by soaking them with indicator solution. Test paper is immersed in the sample liquid and the resulting colour of the paper is compared with standard colours of known pH.

Indicator methods, normally ranging between 3 and 11, do not give pH values to a high degree of accuracy. Sources of errors are

1. High salt concentration in the sample liquid
2. Temperature of the sample liquid
3. Presence of organic substances in the sample liquid

Electrode methods

To measure pH of a sample solution, a reference electrode and a measuring electrode are necessary. A salt bridge may be used to establish connection between them, as shown in Fig. 13.36(a). If the reference electrode is of Ag/AgCl variety and the measuring one is a glass electrode, both can be straightaway dipped into the test solution [Fig. 13.36 (b)].

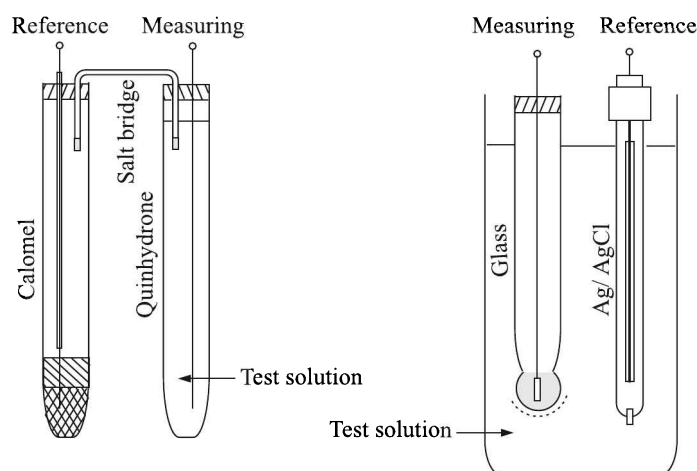


Fig. 13.36 (a) pH measurement with salt bridge between cells, and (b) pH measurement without salt bridge.

The voltage difference between reference and measuring electrodes cannot be measured by an ordinary voltmeter because a current flow through the cell will

1. Develop an Ir -drop where I is the current and r is the internal resistance of the cell
2. Result in polarisation of the test solution

The output impedance of an electrode is on the order of 10^9 ohms. Therefore, the emf can be measured with the help of a potentiometer which acts on voltage balancing basis and so allows no current through the cell, or an op-amp voltage follower which does the same because of its infinite input impedance.

The net e.m.f. E_o between the reference and measuring electrodes for the arrangement shown in Fig. 13.36(a) is given by

$$E_o = E_R + E_M + E_J$$

where E_R is the reference electrode potential

E_M is the measuring electrode potential

E_J is the junction potential between the salt bridge and the test solution.

The value of E_J is maintained within ± 2 mV by a suitable choice of the concentration of the salt solution. Thus, for a calomel-quinhydrone combination

$$E_o = 0.1984(T + 273.16)(7 - \text{pH}) \text{ mV}$$

The entire set-up of glass electrode with Ag/AgCl electrode is available as a single unit, called the *combination electrode*, a diagram of which is given in Fig. 13.37.

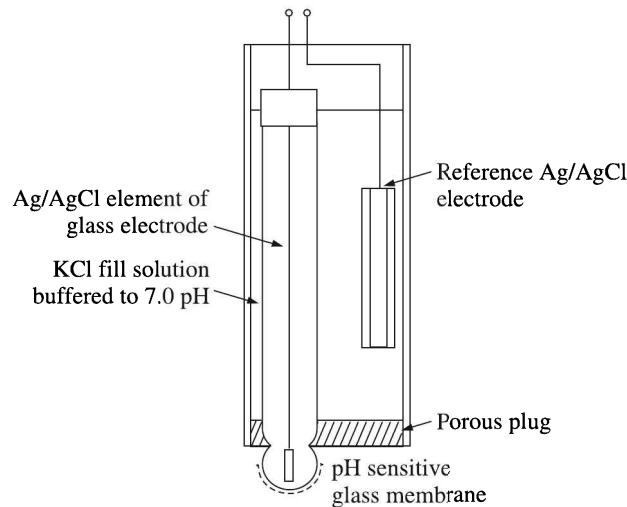


Fig. 13.37 Combination electrode.

Example 13.8

In a pH measurement, the reference and measuring electrodes were hydrogen and calomel electrodes respectively. The measured emf was 650 mV. If the oxidation potential of the saturated calomel electrode is -0.246 V at that temperature, what was the pH of the solution?

Solution

The calomel electrode accepts electrons. Hence, reduction takes place there. Therefore,

$$E_{\text{cathode}} = -(-0.246) \text{ V} = 0.246 \text{ V}$$

Since oxidation takes place at the hydrogen electrode,

$$E_{\text{anode}} = -0.0591 \log[\text{H}^+] = 0.0591 \text{ pH}$$

Now,

$$E_{\text{cell}} = E_{\text{cathode}} + E_{\text{anode}}$$

$$\Rightarrow 0.65 = 0.246 + 0.0591 \text{ pH}$$

$$\Rightarrow \text{pH} = \frac{0.65 - 0.246}{0.0591} = 6.84$$

Ion-Selective Field Effect Transistor (ISFET)

An Ion-selective Field Effect Transistor (ISFET) is a device which generates an output voltage the magnitude of which varies with the change of logarithm of the sensed ion activity or concentration in the same way (but not necessarily in sign) as the corresponding ion-selective electrode (ISE).

It consists of a *p*-type silicon substrate with source and drain diffusions separated by a channel. The channel is overlain by the solution (analyte) which is in direct contact with the gate insulator layer(s) and a reference electrode. Silicon nitride, Si_3N_4 , overlying the SiO_2 is used to provide a charge blocking interface and an improved pH response. The scheme is illustrated in Fig. 13.38. An encapsulation of all regions of the device other than the gate region which is to be exposed to the analyte solution, is mandatory.

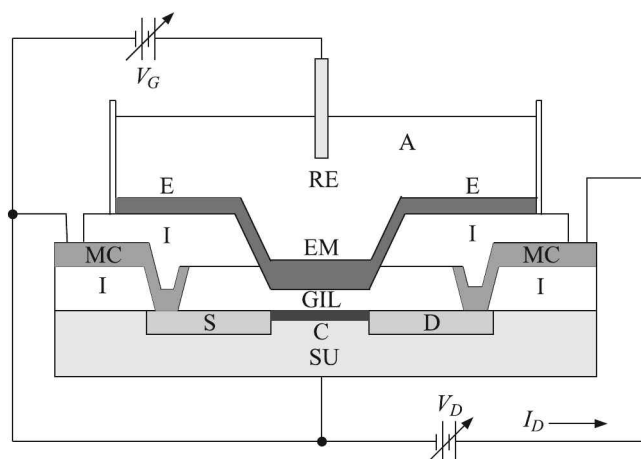


Fig. 13.38 Schematic diagram of an ISFET: RE - reference electrode, A - analyte solution, EM - electroactive membrane, E - encapsulating material, I - insulator, MC - metal contacts, GIL - gate insulating layer(s), S - source, C - channel, D - drain, SU - substrate, V_D - drain bias, V_G - gate bias.

The polarity and magnitude of the gate voltage difference (V_G) applied between the substrate and the gate are so chosen that an n -type inversion layer is formed in the channel between the source and drain regions. The magnitude of the drain current (I_D) is determined by the effective electrical resistance of the surface inversion layer and the voltage difference (V_D) between the source and the drain.

Figure 13.39 shows a diagram of the complete electrochemical system, together with the relevant electrical potentials, i.e. differences in inner potentials between the bulk phases.

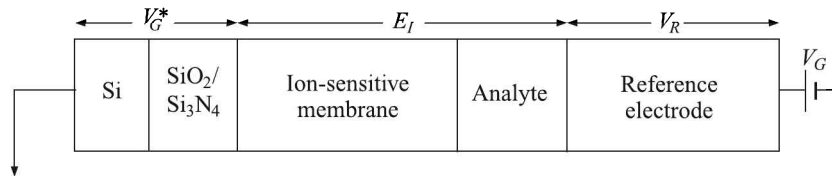


Fig. 13.39 Diagram showing ISFET potential difference contributions.

An equivalent ISFET gate voltage difference V_G^* can be defined as the electrical potential difference between the bulk phases of the semiconductor and gate material as

$$V_G^* = E_I + V_R + V_G$$

where E_I is the interfacial membrane-solution potential difference (from Nernst equation)

V_R is the reference electrode potential

V_G is the gate bias potential

If the drain current I_D is maintained at a constant value by means of a suitable feedback electronics, then the output is a potential difference which varies with change in activity of the sensed ion in accordance with the Nernst equation. The output is, therefore, effectively the same as that of an ISE.

Typical specifications. Typical specifications of a commercially available ISFET pH sensor are given below.

| | |
|-----------------------|------------------------------|
| Measurable pH range | : 0 to 14 |
| Power requirements | : 4.5 V to 5.5 V dc |
| Current consumption | : 2 mA nominal |
| Response time | : 200 ms to 500 ms |
| Sensitivity | : 37 mV to 40 mV per pH unit |
| Accuracy | : ± 0.1 pH unit |
| Operating temperature | : 0°C to 70°C |

Advantages and disadvantage. The advantages and disadvantages of the ISFET pH sensor are as follows:

| <i>Advantages</i> | <i>Disadvantage</i> |
|---|--|
| <ol style="list-style-type: none"> 1. It can be used under demanding circumstances where solids, aggressive chemicals or biological materials are present in the sample and clogging or junction contamination may occur. The probe is easily cleaned using a simple toothbrush. 2. Due to the elimination of the glass-bulb, the possibility of broken glass is eliminated. So, it can be safely used in food and beverage industries. | <ol style="list-style-type: none"> 1. It is highly susceptible to ground looping and static charge build-up. Proper isolation techniques have to be used to avoid them. |

Maintenance of pH Electrodes

Depending on the process conditions and accuracy and stability expectations, a system's pH electrodes require periodic cleaning and calibration. Electrical properties of the measuring and reference electrodes change with time. Calibration against known pH buffers will correct these changes. However, like batteries, pH electrodes also have finite lifetime.

13.6 Polarography

Invented by a Czech chemist Jaroslav Heyrovský in 1922, polarography, a branch of voltammetry, is an electrochemical method of analysing solutions of reducible or oxidisable substances. In polarography, a dropping mercury electrode (DME) is used as the indicator electrode. The electric potential (or voltage) between the reference and indicator electrodes is varied in a regular manner while the current is monitored.

The current vs. potential curve thus obtained is called a *polarogram*. Its shape depends on the method of analysis selected, the type of indicator electrode used, and the potential ramp that is applied. Figure 13.40 shows five selected methods of polarography. The potential ramps are applied to a mercury indicator electrode, and the shapes of the resulting polarograms are compared.

Of all the five methods, we will discuss only three, namely the linear sweep polarography and the two pulse polarographies.

Linear Sweep Polarography

Suppose we apply a linear sweep (or 'ramp') voltage between a DME, dipped in a dilute solution of hydrochloric acid, and a pool of mercury staying underneath. The voltage applied to the DME is negative so that it acts as a cathode. The polarogram will look like curve 1 of Fig. 13.41. Now we add a small amount of cadmium in the hydrochloric acid that will generate a small concentration of cadmium ions in the solution. Next, we generate another polarogram under the same conditions as before. The second polarogram will look like curve 2 of Fig. 13.41.

Both the curves will contain ripples (not shown). Each ripple corresponds to the life cycle of one drop of mercury.

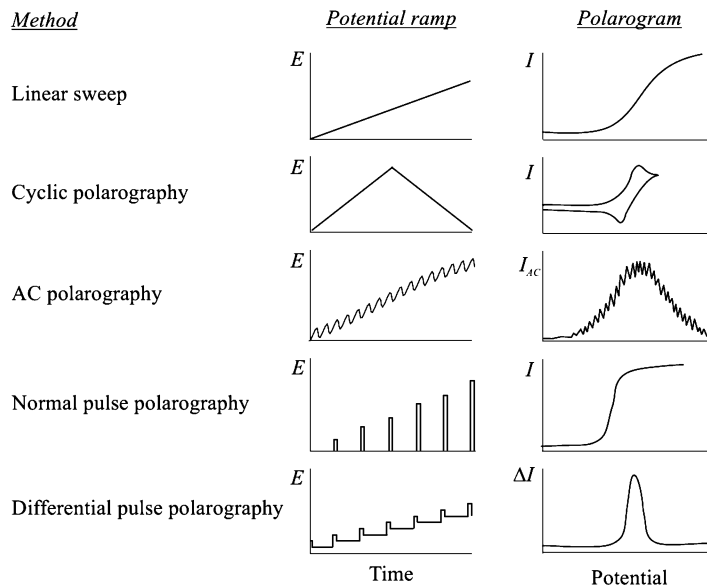


Fig. 13.40 Applied voltage ramps and obtained polarograms in five types of polarography.

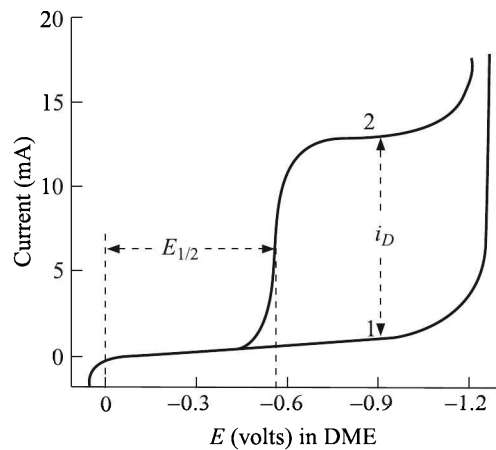


Fig. 13.41 Polarograms of pure hydrochloric acid (curve 1) and cadmium in hydrochloric acid (curve 2).

It is the difference between the currents on the two curves that is of interest to us. The curves are identical up to a voltage of -0.5 V. The reason is that the voltage is not negative enough to bring about the reduction of an appreciable fraction of the cadmium ion at the surface of the drop of mercury. The situation changes dramatically at more negative voltage. The resulting polarogram assumes the form of a *wave* which indicates that the reduction of cadmium ion takes place more and more rapidly as the DME potential becomes more and more negative until it reaches a potential of about -0.7 V. Thereafter it reaches a plateau. The corresponding current, called the *diffusion current* i_D , depends on the concentration of the cadmium ion in the solution.

Another parameter, called the *half-wave potential* $E_{1/2}$, is defined. It corresponds to the potential at the half of the diffusion current (see Fig. 13.41). Under a set of defined experimental conditions, each ion has its own characteristic half-wave potential. So, this value is suitable for qualitative analysis of the sample.

Theoretical relations

In general, the Fick's law²⁷ governs the rate of diffusion of ions to the cathode. Written in mathematical form

$$\frac{\partial d_s}{\partial t} = \alpha D \frac{\partial [C]}{\partial x} \quad (13.32)$$

where $\frac{\partial d_s}{\partial t}$ is the diffusion rate of the sample ion
 α is the surface area of the indicator electrode

D is the diffusion coefficient

$\frac{\partial [C]}{\partial x}$ is the ion concentration gradient between the bulk and the electrode

At the equilibrium condition, the rate of diffusion equals the ion discharge rate at the electrode. So, if n electrons take part in the process, we get from Eq. (13.32)

$$\frac{i_D}{ne} = \alpha D \frac{[C] - [C_e]}{l} \quad (13.33)$$

where e is the electronic charge

$[C_e]$ is the ion concentration at the electrode

$[C]$ is the bulk ion concentration

l is the hypothetical *Nernst layer* around the electrode

While writing Eq. (13.33), it has been assumed that the generated current is solely owing its origin to ion diffusion. But, other factors like adsorption, transference, etc. also play a role in the current generation. Considering all these factors, the diffusion current is given by the Ilkovic equation

$$i_D = 708ND^{1/2}m^{2/3}t^{1/6}[C]$$

where N is the number of electrons transferred per mole of analyte

D is the diffusion coefficient of the analyte in the medium (in cm^2/s)

m is the mass flow rate of Hg through the capillary (in mg/s)

t is the drop lifetime (in seconds)

Instrumentation

A schematic diagram of the set-up is shown in Fig. 13.42. The heart of the instrumentation in polarography measurement is a polarographic cell which consists of a reference and an indicator electrode.

²⁷Named after its discoverer Adolf Eugen Fick (1829 – 1901), a German physiologist.

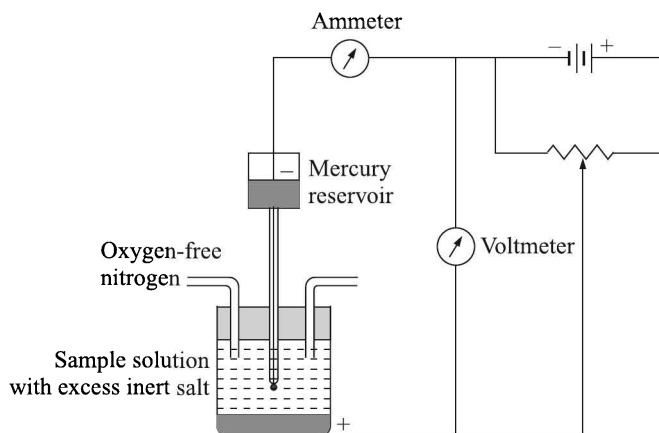


Fig. 13.42 Schematic diagram of a polarography set-up.

Reference electrode. The reference electrode of the cell may be of two types:

1. Internal
2. External

Internal reference electrode. An internal reference electrode is in direct contact with the sample solution. Such electrodes are used

1. In routine analysis where a limited constancy of potential is not important
2. In work where contamination by alkali metal ions or other constituents of a salt bridge would be harmful, and
3. In cells containing a small volume of available sample.

The internal reference cell is generally a pool of mercury so as to make the surface current density small and the potential stable.

External reference electrode. The external reference electrode can be a saturated calomel electrode connected to the sample by a salt bridge.

Indicator electrode. The indicator electrode is a DME, consisting of a mercury drop hanging at the orifice of a fine-bore (diameter $\sim 50 \mu\text{m}$) glass capillary. The capillary is connected to a mercury reservoir so that mercury flows through it at the rate of a few milligrams per second. At the orifice the outflowing mercury forms a drop, which grows until it falls off. The lifetime of each drop is several seconds (usually 2 to 5). Each drop forms a new electrode; its surface is practically unaffected by processes that took place on the previous drop. Hence each drop represents a well-reproducible electrode with a fresh, clean surface. The small droplet ensures a high surface current density and hence quick polarisation of the analyte.

Current-voltage measurement. Polarographic current-voltage curves can be recorded with a simple instrument consisting of a potentiometer or a high impedance voltmeter, and a low impedance milliammeter. The voltage can be varied by manually changing the applied voltage in finite increments, measuring current at each, and plotting current as a function of the voltage. Alternatively, commercial instruments are available in which voltage is increased linearly with time (a voltage ramp), and current variations are recorded automatically.

Interference. The major interference in the measurement arises out of the electrolytic current which results from ion migration rather than diffusion. To inhibit this effect and to make the diffusion current dominant, a large quantity (about 100 times by volume) of an inert supporting electrolyte is added to the sample solution.

Sometimes dissolved oxygen in the solution interferes with measurement through electrochemical reactions. To minimise this interference, oxygen-free nitrogen is bubbled through the solution at a constant rate. This also performs the function of a stirrer.

Accuracy. If measurements are made under proper conditions, the magnitude of i_D constitutes a measure of the concentration of the reducible substance (quantitative analysis). Diffusion currents also result from the oxidation of certain oxidisable substances when the DME is the anode. When the solution contains several substances that are reduced or oxidised at different voltages, the current-voltage curve shows half-wave potential and diffusion current for each. The method is thus useful in detecting and determining several substances simultaneously and is applicable to relatively small concentrations e.g., from 10^{-6} M to about 0.01 M, or approximately 1 to 1,000 ppm.

Pulse Polarography

In the linear sweep method, because the current is continuously measured during the growth of the Hg drop, there is a substantial contribution from capacitive current which comes into play due to the following reasons:

1. As the mercury flows from the capillary end, there is initially a large increase in the surface area. As a consequence, the initial current is dominated by capacitive effects as charging of the rapidly increasing interface occurs.
2. Towards the end of the drop life, there is little change in the surface area which diminishes the contribution of capacitance changes to the total current. At the same time, any redox process which occurs will result in faradaic current that decays approximately as the square root of time (due to the increasing dimensions of the Nernst diffusion layer). The exponential decay of the capacitive current is much more rapid than the decay of the faradaic current; hence, the faradaic current is proportionally larger at the end of the drop life.
3. Because the potential is changing during the drop lifetime (assuming typical experimental parameters of a 2 mV/s scan rate and a 4 s drop time, the potential can change by 8 mV from the beginning to the end of the drop), the charging of the interface (capacitive current) has a continuous contribution to the total current, even at the end of the drop when the surface area is not rapidly changing.

Owing to these problems, the typical signal to noise of a polarographic experiment allows detection limits of only approximately 10^{-6} M. Better discrimination against the capacitive current can be obtained using the pulse polarographic techniques. Pulse polarography can be divided into two categories—normal pulse polarography and differential pulse polarography.

Normal pulse polarography

The pulse polarography is a technique which tries to minimise the background capacitive contribution to the current by eliminating the continuously varying potential ramp, and

replacing it with a series of potential steps of short duration. In normal pulse polarography (NPP), each potential step begins at the same potential at which no faradaic electrochemistry occurs, and the amplitude of each subsequent step increases in small increments. When the mercury drop is dislodged from the capillary by a drop knocker at accurately timed intervals, the potential resets to the initial value in preparation for a new step (Fig. 13.43).

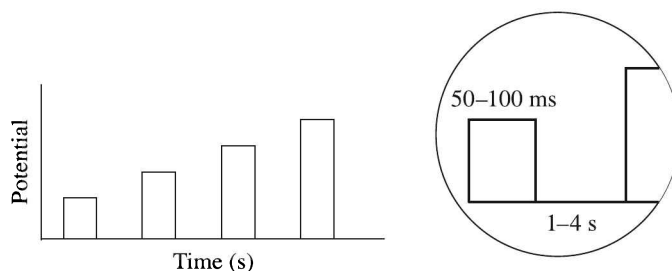


Fig. 13.43 Pulse form applied in normal pulse polarography.

In NPP, the polarogram is obtained by plotting the measured current vs. the step potential. Since the current is not followed during the growth of the mercury drop, the normal pulse polarogram has the typical shape of a sigmoid. The diffusion current is measured just before the drop is dislodged, allowing discrimination against the background capacitive current. This enhances the limits of detection to 10^{-7} or 10^{-8} M.

Differential pulse polarography

For differential pulse polarography, many of the experimental parameters, such as accurately timed drop lifetimes, potential step duration of 50 to 100 ms at the end of the drop lifetime, are the same as those of the normal pulse polarography. Unlike normal pulse polarography, however, each potential step has the same amplitude, and the return potential after each pulse is slightly negative of the potential prior to the step (Fig. 13.44).

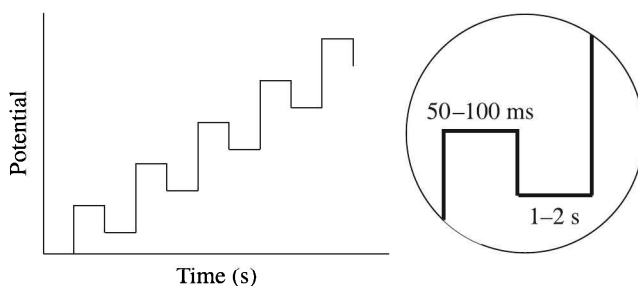


Fig. 13.44 Pulse form applied in differential pulse polarography.

The differential pulse polarogram is obtained by measuring the current immediately before the potential step, and then again just before the end of the drop lifetime. The analytical current in this case is the difference between the current at the end of the step and the current before the step (the differential current). This differential current is then plotted vs. the average potential (average of the potential before the step and the step potential) to obtain

the differential pulse polarogram. Because this is a differential current, the polarogram is like the differential of the sigmoidal normal pulse polarogram. As a result, the differential pulse polarogram is peak shaped.

The differential pulse polarography has even better ability to discriminate against capacitive current because it measures a difference current. That means, it helps to subtract any residual capacitive current that remains prior to each step. Limits of detection with differential pulse polarography are 10^{-8} to 10^{-9} M.

Applications

The majority of the chemical elements can be identified by polarographic analysis, and the method is applicable to the analysis of alloys and to various inorganic compounds. Polarography is also used to identify numerous types of organic compounds and to study chemical equilibria and rates of reactions in solutions. The various applications are listed in Table 13.10.

Table 13.10 Applications of polarographic analysis technique

| <i>Area</i> | <i>Application</i> |
|---|--|
| Inorganic analysis | Predominantly for trace-metal analysis in metallurgy, environmental analysis (air and water contaminants), food analysis, toxicology, and clinical analysis. |
| Organic analysis | In elemental analysis and functional group analysis. |
| Analysis of drugs and pharmaceutical preparations | Determination of vitamins, alkaloids, hormones, terpenoid substances, natural colouring substances and pesticide or herbicide residues in foods. |

13.7 Viscosity Measurement

Before we discuss the measurement procedures, let us recapitulate the definition of viscosity, and the distinction between Newtonian and non-Newtonian fluids.

Coefficient of Viscosity

Consider an area A of the streamline flow of an incompressible liquid. Let the velocities of fluid flow at P and Q (Fig. 13.45) be $(v + dv)$ and v respectively and the distance between P and Q be dz . There will be some friction between the two layers of fluid flow, causing the velocity difference. This frictional backward force is called the 'viscous drag'.

Newton found that the viscous drag is proportional to area of the layers A and the velocity gradient. That is

$$F \propto A$$

$$\propto -\frac{dv}{dz}$$

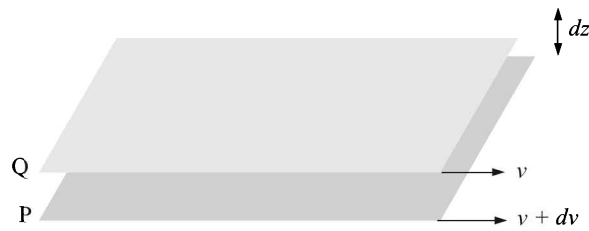


Fig. 13.45 Two layers of streamline fluid flow.

Combining the two,

$$F = -\mu A \frac{dv}{dz} \quad (13.34)$$

where, μ is the *coefficient of viscosity*. Thus, if

$$A = 1 \quad \text{and} \quad -\frac{dv}{dz} = 1, \quad F = \mu.$$

Therefore, the coefficient of viscosity or simply viscosity, is defined as the tangential backward force that acts between two fluid layers of unit area, situated unit distance apart and having unit relative velocity, when the fluid is in streamline motion.

Units of viscosity

The unit of viscosity in the CGS system is called poise²⁸. Even in relation to high-viscosity fluids, the unit that is most usually encountered is the centipoise (cP), which is 0.01 poise. The SI unit for μ is kg/(m·s) which equals 10 poise. The SI unit of viscosity, equivalent to newton-second per square metre (Ns m⁻²), is sometimes referred to as the *poiseuille* (symbol Pl). Pascal-second (symbol Pa·s) is another unit that is often used. One poise is exactly 0.1 Pa·s. One poiseuille is 10 poise or 1000 cP, while 1 cP = 1 mPa·s (one millipascal-second).

Measurement of viscosity serves to determine the resistance of fluids to flow or to measure the molecular weight of polymers. What we have defined above, i.e. μ , is called the *absolute viscosity* or *dynamic viscosity*. Another quantity, called the *kinematic viscosity* ν is defined as

$$\nu = \frac{\mu}{\rho} \quad (13.35)$$

where ρ is the density of the fluid. Its CGS unit is stokes. It is most usually encountered as the centistokes (cSt) (= 0.01 stokes).

Newtonian and non-Newtonian fluids

Our definition of viscosity rests on the Newtonian relation given by Eq. (13.34). Fluids which obey this relation are called *Newtonian fluids*, whereas non-Newtonian fluids do not obey this relation. Petrol, kerosene, mineral oils, water and salt solutions in water are a few examples of Newtonian fluids while printers' ink, starch, peanut butter, tar, chewing gum are some examples of the other kind.

²⁸Named after the French physician and physiologist Jean Louis Marie Poiseuille (1797–1869).

Methods of Viscosity Measurement

Principles

Measurements of viscosity are carried out on the basis of one of the following three phenomena where this property plays a major role:

1. Flow through a capillary tube
2. Drag experienced by a falling ball through a fluid
3. Drag experienced by one of the concentric cylinders carrying fluid between them when the other cylinder is rotating

We will discuss them in that order.

Flow through a capillary tube. According to this formulation, the viscosity μ of a streamline flow of a Newtonian fluid through a capillary of radius a and length l is given by

$$\mu = \frac{\pi \Delta p a^4}{8Vl} \quad (13.36)$$

where Δp is the pressure difference between the ends of the tube

V is the volume rate of flow through the tube.

Equation (13.36), known as the Hagen-Poiseuille formula, needs to be corrected for a few simplifying assumptions that were made to arrive at this formula. However, we need not delve into that.

If CGS units are used for different factors, Eq. (13.36) yields the value of viscosity in poise.

Viscous drag experienced by a falling ball through the fluid. The viscous drag D experienced by a small solid sphere falling through a viscous liquid is given by the Stokes'²⁹ formula

$$D = 6\pi\mu rv \quad (13.37)$$

where r is the radius of the spherical ball

v is the terminal velocity of the ball.

If

- (a) v is measured by noting the transit time of the sphere between two marks which are not within 10 cm from the top or bottom of the vessel containing the fluid
- (b) the vessel diameter is at least 10 cm and
- (c) the sphere diameter is less than 0.2 cm

then the viscous drag equals the resultant force arising from the weight of the sphere minus the buoyancy experienced by it. That is,

$$6\pi\mu rv = \frac{4}{3}\pi r^3(\rho - \sigma)g \quad (13.38)$$

²⁹George Gabriel Stokes (1819 – 1903), was a British mathematician and physicist.

where ρ is the density of the material of the sphere
 σ is the density of the liquid, and
 g is the acceleration due to gravity.

On rearranging Eq. (13.38), we get

$$\mu = \frac{2}{9} \cdot \frac{(\rho - \sigma)g}{v} r^2$$

If t is the transit time of the sphere to cross the distance l ,

$$\mu = \frac{2}{9} \cdot \frac{\rho - \sigma}{l} r^2 g t \quad (13.39)$$

Equation (13.39) shows that $r^2 t$ is a constant for a given liquid in a given experimental set-up. Therefore, a plot of r^2 vs. t^{-1} should be a straight line, the slope of which will yield the value of μ .

Rotating concentric cylinders. If two concentric cylinders, having the fluid in between, rotate with a relative angular velocity ω , the stationary cylinder experiences a torque τ . Then the viscosity μ can be obtained from the relation

$$\mu = \frac{\tau}{4\pi l \omega} \left[\frac{1}{a^2} - \frac{1}{b^2} \right] \quad (13.40)$$

where a is the radius of the stationary cylinder (inner)
 b is the radius of the rotating cylinder (outer)
 l is the height of the liquid level in the cylinder.

Measuring instruments

Capillary flow-based. The most common method is shown in Fig. 13.46(a). The process fluid is allowed to flow through a capillary of length l and radius a . The head at the entry to the capillary is kept constant by an arrangement as shown in the diagram. The volume rate of flow is found by measuring the volume collected in a known time interval.

The viscosity μ can be found using Eq. (13.36). An arrangement, shown schematically in Fig. 13.46(b), allows the capillary flow-based viscometer to measure viscosity of a process fluid in industry.

The pump should generate a constant flow. Rather than using Eq. (13.36), these viscometers are calibrated against Δp generated by fluids of known viscosities.

Ostwald viscometer. The Ostwald³⁰ viscometer utilises the same principle, though it offers a comparison between viscosities of two fluids. The viscometer, as shown in Fig. 13.47, consists of a U-tube of glass with two bulbs and a capillary tube in between. While one arm of the tube is open with a funnel at the end, the other arm has a stopcock. The viscometer may be kept immersed in a liquid bath of fixed temperature.

³⁰Friedrich Wilhelm Ostwald (1853 – 1932) was a German chemist. He received the Nobel Prize in Chemistry in 1909.

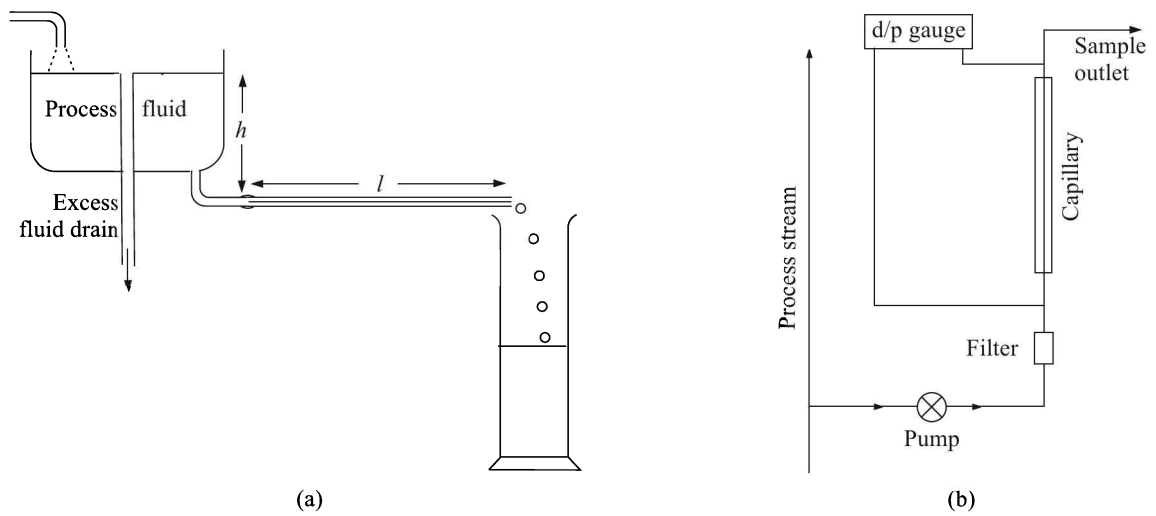


Fig. 13.46 (a) Capillary flow method, and (b) capillary flow-based measurement arrangement as used in industry. The arrangement is horizontal though it looks vertical.

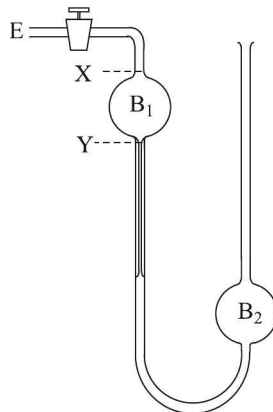


Fig. 13.47 Ostwald viscometer.

Operation.

- Step 1. Let the volume between marks X and Y of bulb B_1 be V . The first liquid of volume V is poured through funnel at the open end.
- Step 2. The stopcock is opened and air is slowly sucked through the end E so that the poured liquid is drawn into bulb B_1 to occupy the volume between X and Y. The stopcock is then closed.
- Step 3. The stopcock is opened and simultaneously a stop-watch is started. When the bulb B_1 is just empty, the stop-watch is stopped and the elapsed time is noted. Let it be t_1 .
- Step 4. Steps 1 to 3 are repeated for the second liquid. Let the elapsed time this time be t_2 .

Theory.

Let l be the length of the capillary tube
 a be the radius of the capillary tube
 h be the average height of the liquid in the bulb when it falls through the capillary
 V_1, V_2 be the rates of flow for the two liquids
 ρ_1, ρ_2 be the densities of two liquids
 μ_1, μ_2 be the coefficients of viscosities of two liquids.

Then from Poiseuille's relation, we have for the two liquids

$$\mu_1 = \frac{\pi p_1 a^4}{8V_1 l} = \frac{\pi(h\rho_1 g)a^4}{8(V/t_1)l} \quad (13.41)$$

and

$$\mu_2 = \frac{\pi p_2 a^4}{8V_2 l} = \frac{\pi(h\rho_2 g)a^4}{8(V/t_2)l} \quad (13.42)$$

Dividing Eq. (13.41) by Eq. (13.42), we get

$$\frac{\mu_1}{\mu_2} = \frac{\rho_1 t_1}{\rho_2 t_2} \quad (13.43)$$

Thus, if ρ_1 and ρ_2 are known, μ_1/μ_2 can be found out.

- Note:*
1. The same apparatus can be used to compare viscosities of the same liquid at two different temperatures by varying the temperature of the bath.
 2. The time necessary to make a viscosity measurement by this viscometer is large. Hence, it cannot be used for samples that evaporate or deteriorate when exposed to atmosphere.

Saybolt or Redwood viscometer. Used as a standard for testing petroleum products, this viscometer is known as Saybolt viscometer in USA and Redwood viscometer in UK (also in India). It consists of a cup and orifice assembly as shown in Fig. 13.48. A charge of 60 cm^3 of

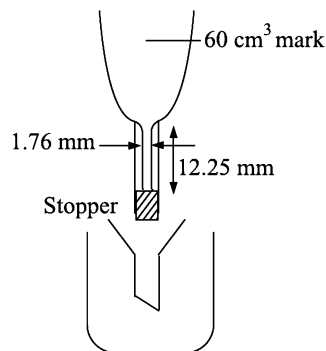


Fig. 13.48 Saybolt (Redwood) viscometer.

the test fluid is taken in the cup. The stopper is opened and time t taken by the fluid to empty the cup by flowing through the capillary orifice of diameter 1.76 mm and length 12.25 mm is noted.

Kinematic viscosity ν is calculated from either of the equations (13.44) whichever is applicable.

$$\nu = \begin{cases} 0.226t - \frac{195}{t} & \text{centistokes for } t < 100 \text{ s} \\ 0.220t - \frac{135}{t} & \text{centistokes for } t > 100 \text{ s} \end{cases} \quad (13.44)$$

Accuracies $\sim \pm 0.1\%$ of the reading can be achieved if the standard procedure laid down by ASTM is followed.

Falling-ball viscometer. The Stokes formula forms the basis of this viscometer. Equation (13.39) which is the working formula, can be written as

$$\mu = (\rho - \sigma)Bt$$

where B is the ball constant. We describe the procedure for measuring viscosity of a representative fluid castor oil by this method.

Castor oil is placed in a glass cylinder. Spheres of known diameter are dropped into the liquid. Time required for one sphere to travel between two fiducial marks is measured and the viscosity may be calculated therefrom. Possible errors arising out of wall and limited depth of the vessel are eliminated if the marks are located at a distance of, at least, 10 cm from the top and bottom, the vessel diameter is at least 10 cm, and the diameter of spheres does not exceed 0.2 cm.

It is better to determine μ from the slope of the r^2 vs. t^{-1} plot as discussed before. Viscosity range of 0.01 to 10^6 cP can be covered with the help of several calibrated balls of different sizes. Depending on the accuracy of time interval measurement, an accuracy of $\pm 0.1\%$ can be achieved.

In industrial situation, an arrangement like that shown in Fig. 13.49 can measure the viscosity of process fluid automatically.

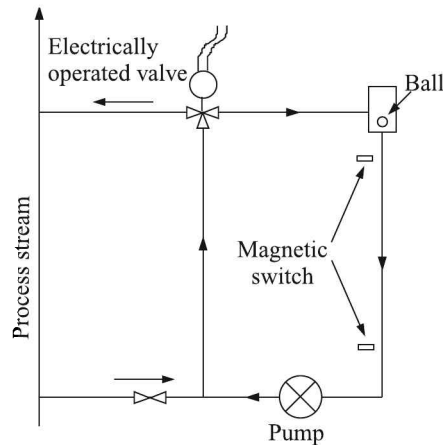


Fig. 13.49 Falling-ball measurement of viscosity in industry.

The sample pump raises the ball and sample to the top of the fall tube. Then the pump flow stops and the ball starts falling, actuating two magnetic switches on the way. The elapsed time between two switch activations is a measure of viscosity of the fluid.

Rotational viscometer. Rotational viscometers, perhaps, are the most widely used among viscometers. They are generally of two types—concentric cylinder type, and cone-and-plate type.

In concentric cylinder type viscometers, one of the cylinders is kept static while the other is rotated. In Fig. 13.50(a), the outer cylinder is rotating at a fixed angular velocity. The test liquid is placed in the gap between the cylinders. The rotational speed, which should not be too high to generate turbulence, generates a torque τ on the inner suspended cylinder. μ may be calculated using Eq. (13.40).

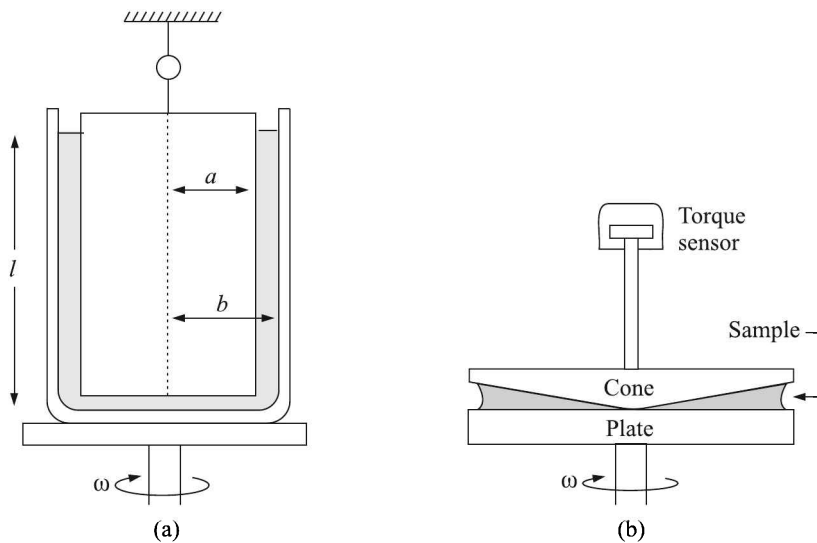


Fig. 13.50 Rotational viscometer: (a) concentric cylinder type, and (b) cone-and-plate type .

However, while deriving Eq. (13.40), the torque acting at the bottom of the cylinder was not taken into account. This error can be eliminated by measuring torques for two heights of the liquid and subtracting the results. Thus, if l and l' are two heights and the corresponding torques are τ and τ' ,

$$\mu = \frac{\tau - \tau'}{4\pi\omega(l - l')} \left(\frac{1}{a^2} - \frac{1}{b^2} \right)$$

The torques τ and τ' can be determined by knowing the rigidity modulus n of the suspension wire and measuring angular deflections θ and θ' corresponding to τ and τ' . Then,

$$\tau - \tau' = n(\theta - \theta')$$

The instrument is inexpensive and can be used for viscosity range from 0.1 to 5000 poise. Even it is good for determining force vs. flow relationships of non-Newtonian fluids.

The other rotational type viscometer is the cone-and-plate viscometer [Fig. 13.50(b)]. It consists of a flat plate and a cone with a small angle ($< 1^\circ$). The tip of the cone nearly touches

the plate. The fluid sample fills the gap between the plate and the cone and stays there due to its surface tension (capillary action).

The geometry of a cone-and-plate viscometer provides uniform shear rate and stress throughout the fluid sample at a given angular velocity, because here the sample thickness increases with the increase in the tangential velocity $v(= \omega r)$ at higher radius. Secondly, since the thickness of fluid layer is small, there will be less generation of heat owing to the churning motion of the fluid.

A suitable design of the cone and plate, and rotational speed variation may allow measuring viscosities from 10^{-4} to 10^8 poise by this method.

Air bubbles present in samples will produce a large error in viscosity measurement by rotational methods.

Float viscometer. The two-float viscometer uses the upper float to monitor the flow rate while the lower float is sensitive to the viscosity of the flowing fluid (Fig. 13.51).

During a measurement, the fluid flow rate is maintained constant by adjusting the needle valve so that the upper float maintains a constant position. Under such condition, the position of the viscosity float indicates the viscosity of the fluid on a scale graduated on the tube. The throttle valve is necessary to produce a pressure gradient in the line so that the outlet from the float viscometer can be discharged to the process stream. If the outlet is not returned to the process stream, the throttle valve is not necessary.

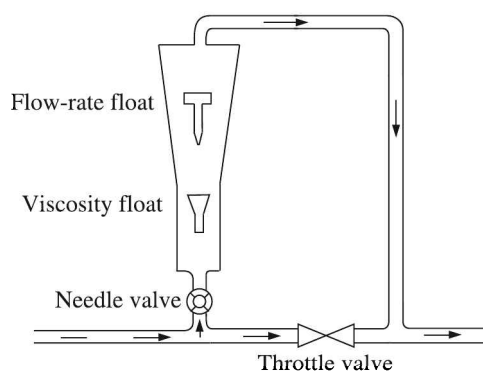


Fig. 13.51 Float viscometer.

The range of this type of viscometer is not high—generally from 0.3 to 250 cP and so is its accuracy which is $\sim \pm 3\%$ on an average.

Coriolis mass flow viscometer. The Coriolis mass flowmeter generates an effective rotational motion of the fluid flowing through it. Therefore, the fluid experiences a shear stress that alters the velocity profile of the flowing fluid inside the tube. The shear forces dampen the tube oscillation, thus necessitating a larger drive force to maintain the oscillation. This excess drive force which is necessary to maintain a sustained oscillation overcoming the damping caused by the viscosity of the fluid, is a measure of the viscosity. This principle has been utilised to construct viscosity meters from CMF meters of late.

We give in Table 13.11 the viscosity values of a few common fluids. Fluids with variable compositions, such as honey, can have a wide range of viscosities.

Table 13.11 Viscosity of a few common fluids

| <i>Type</i> | <i>Fluid</i> | <i>Viscosity (cP)</i> |
|------------------|----------------|-----------------------|
| Gas (at 0°C) | Hydrogen | 0.0084 |
| | Air | 0.0174 |
| | Xenon | 0.0212 |
| Liquid (at 25°C) | Ethyl alcohol | 0.248 |
| | Acetone | 0.326 |
| | Methanol | 0.597 |
| | Propyl alcohol | 2.256 |
| | Benzene | 0.64 |
| | Water | 1.0030 |
| | Nitrobenzene | 2.0 |
| | Mercury | 17.0 |
| | Sulphuric acid | 30 |
| | Olive oil | 81 |
| | Castor oil | 985 |
| Glycerol | 1,485 | |
| Molten polymers | 100,000 | |

13.8 Consistency Measurement

The coefficient of viscosity, or simply viscosity, is defined for Newtonian fluids that generate a linearly proportional flow when subjected to a shearing stress. Non-Newtonian fluids, however, do not maintain a linear relation between the applied shearing stress and the generated flow. For them, another parameter, called *consistency* is defined. Mechanically, consistency is the resistance to deformation or shear by fibrous materials. These materials include wood pulp, tomato paste, flour dough, drilling mud among others.

In fact, consistency control is of fundamental importance in paper and pulp industries. We will study consistency from the viewpoint of these industries though the definition and methods of measurement are applicable to other relevant industries as well.

Definitions

Quantitatively, consistency, C_s , is defined as

$$C_s = \frac{f}{w} \times 100 \quad (13.45)$$

where f is the weight of the bone dry fibrous material present in a certain volume of the pulp/stock slurry, and

w is the weight of the same volume of the pulp/stock slurry

In this context, it needs to be mentioned that a pulp slurry consists of fibrous material and water while a stock slurry contains additives, such as fillers and chemicals, over and above the pulp slurry.

Table 13.12 gives us an idea about the consistency ranges and their requirements of measurement in the paper and pulp industries.

Table 13.12 Consistency ranges and their requirement of measurement

| Range | Requirement (approx.) | Application area |
|-------|-----------------------|--------------------------------------|
| 0–1% | 10% | Wet end of a paper producing machine |
| 1–8% | 75% | Paper and pulp mill |
| 8–15% | 15% | Pulp mill |

Methods of Measurement

Gravimetric method

Recommended as the industry standard³¹ by the TAPPI³², the method consists of manually collecting a selected sample and weighing it in the wet and dry states.

Though simple, the method is tedious for a continuous control and its repeatability is around 10% because of adopting manual means.

Online methods

Mechanical. Quite a few mechanical methods are used in industries. Some of them are

1. A stationary blade/rotor is submerged in the pulp/stock slurry flowing out at a constant rate and the mechanical shear force/torque experienced by the blade/rotor is measured (Fig. 13.52).
2. The pressure drop caused by a constant flow through a pipe of fixed length is measured.
3. The head needed to maintain a constant flow through a fixed pipe is calibrated in terms of consistency.
4. The amplitude of vibrations sensed by an outer cylinder, when the inner cylinder is vibrated sinusoidally, is calibrated to indicate consistency.

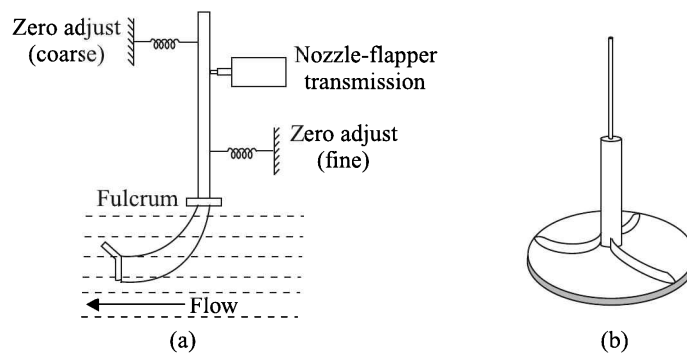


Fig. 13.52 Consistency sensing: (a) stationary blade, and (b) rotor.

³¹Test method T240 om-88

³²Technical Association of the Pulp and Paper Industry, USA

Electrical. Generally two methods are common:

1. The power required to rotate a propeller, submerged in the pulp /stock slurry, at a constant rpm is monitored.
2. The electrical conductivity of the pulp/stock slurry is measured.

Radiative. The intensity of transmitted/reflected visible light, microwave, ultrasonic wave or γ -ray is measured.

However, all online measurements provide short-term information about pulp/stock slurry consistency which is influenced by factors like (i) furnish³³, (ii) freeness³⁴, (iii) velocity of slurry flow, (iv) temperature and (v) pressure. Therefore, online consistency measurements provide only comparative indications and should be used to sense trend or track temporal consistency variations.

13.9 Turbidity Measurement

Before we go into the methods of measurement of turbidity, let us discuss what turbidity is and why it does occur.

Turbidity and Scattering of Light

Water or other liquids appear cloudy or hazy if there are suspended particles in them. Turbidity is a measure of the degree to which the liquid loses its transparency due to the presence of suspended particulates. Basically, it is caused by the scattering of the incident radiation by suspended particles. What happens when a light beam strikes a liquid sample containing finely distributed particles is shown in Fig. 13.53.

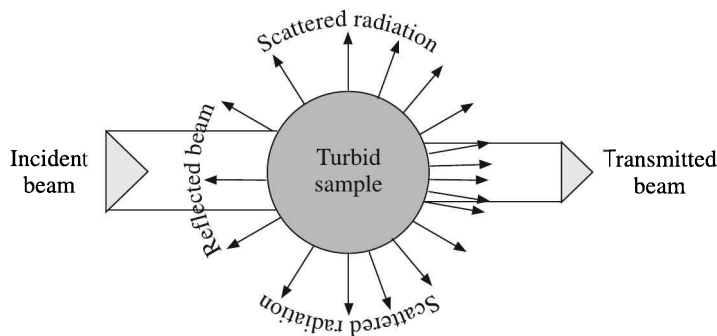


Fig. 13.53 Scattering of light by a turbid liquid sample.

We see that the intensity of the incident beam gets reduced owing to three factors:

1. Reflection from the wall of the container
2. Scattering caused by the suspended particles
3. Absorption of light by the liquid as well as suspended particles

³³Variation in fibre lengths and blends.

³⁴Ability of the suspension to release water.

The pattern of the intensity of the scattered light depends on the size of the particles vis-a-vis the wavelength of the incident radiation. If ϕ is the diameter of the particle and λ is the wavelength of the incident radiation, then the scattered radiation is

1. Nearly isotropic for $\phi < 0.1\lambda$
2. Concentrated in the forward direction for $\phi \sim 0.25\lambda$
3. Highly concentrated in the forward direction with maxima and minima at wider angles for $\phi > \lambda$ as shown in Fig. 13.54

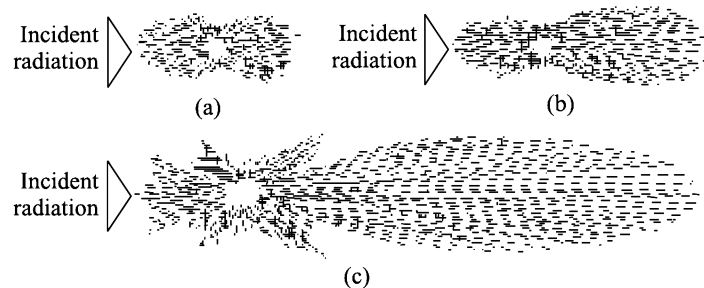


Fig. 13.54 Pattern of scattered light vis-a-vis particle size— (a) $\phi < 0.1\lambda$, (b) $\phi \sim 0.25\lambda$, and (c) $\phi > \lambda$.

Apart from the size of the suspended particles, angular variation of the intensity of the scattered radiation depends on the colour of the incident light as well as the shape of particles.

Isotropic Scattering and Turbidity

Diffused scattering of light is often called the *Tyndall*³⁵ effect. For Tyndall effect, light waves scattered by different particles possess perfectly random phases. In such cases, for N particles, the resultant intensity is just N times that reflected from an individual one. This can be shown mathematically as follows.

Let a be the individual amplitude of reflected waves, and

N be the number of wave trains superposed.

Then, the amplitude of the resultant wave will be the amplitude of motion of a particle undergoing N simple harmonic motions at once, each of amplitude a . If these motions were all in the same phase, the resultant amplitude would be Na and the intensity N^2a^2 , or N^2 times of one wave. However, in a diffused reflection, the phases are distributed at random. A graphical method of compounding amplitudes would look like that given in Fig. 13.55.

The phases $\alpha_1, \alpha_2, \dots$ take perfectly arbitrary values between 0 and 2π . If A = the resultant amplitude, the intensity of the diffused reflection is $I = A^2$.

Now,

$$\begin{aligned} A^2 &= [a^2(\cos \alpha_1 + \cos \alpha_2 + \dots + \cos \alpha_N)^2 + a^2(\sin \alpha_1 + \sin \alpha_2 + \dots + \sin \alpha_N)^2] \\ &= a^2 \sum_{i=1}^N \cos^2 \alpha_i + a^2 \sum_{i=1}^N \sin^2 \alpha_i \\ &= a^2 N \end{aligned}$$

³⁵Named after John Tyndall (1820 – 1893), a British “natural philosopher”.

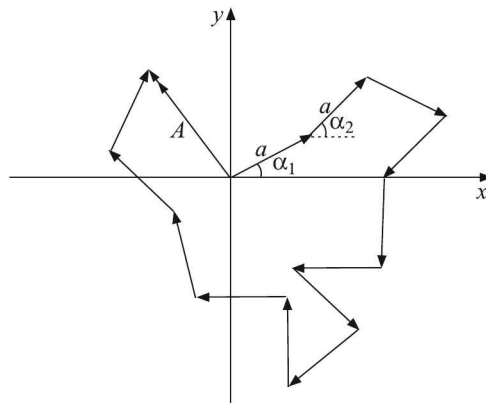


Fig. 13.55 Resultant amplitude of diffused reflection

because, cross terms like $2 \cos \alpha_1 \cos \alpha_2$ take both positive and negative values and when N is large, they will cancel out. So,

$$I \sim Na^2 \quad (13.46)$$

Thus, we see that the intensity of the diffused scattering is linearly related to the number of suspended particles in the liquid, i.e. its turbidity. However, this linear relation exists up to about 40 NTU³⁶ beyond which the intensity no longer increases with the number of suspended particles, as shown in Fig. 13.56.

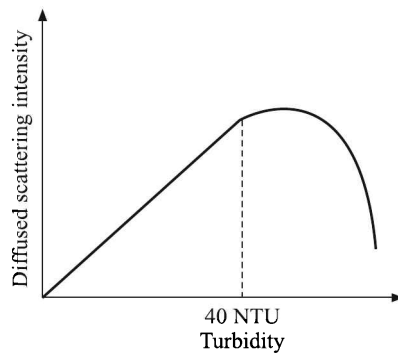


Fig. 13.56 Turbidity vs. diffused scattering intensity curve.

Suspended Particles

Suspended particles, especially in water of water-bodies like lakes, may be inorganic and/or organic. Sand ($\phi \sim 0.09$ to 1.5 mm), silt ($\phi \sim 0.005$ to 0.05 mm) or clay ($\phi < 0.002$ mm) particles may constitute inorganic suspended particles while typically phytoplanktons and algae constitute the organic particulate matter. The study of their concentration is important for the following reasons:

³⁶see Section 13.9 at page 591 for definition.

1. Organic particulates may harbour pathogens like *Giardia*, thus increasing the possibility of causing waterborne diseases. Inorganic particles, though do not have notable health effects, are harmful for industrial processes because they may clog or scour pipes and machinery.
2. For water bodies, the inorganic particulate matter reduces sunlight penetration thereby suppressing photosynthetic activity of organic particulate matter. This, in turn, leads to fewer photosynthetic organisms which serve as food for aquatic life.
3. An excess of organic particulates in a water body, on the other hand, depletes dissolved oxygen in water because although photosynthetic by day, algae respire at night, using valuable dissolved oxygen. Extensive depletion of dissolved oxygen often results in fish kills.

Units of Turbidity

Jackson turbidity unit (JTU). Jackson turbidity unit (JTU) was originally defined in terms of the inverse length of a column of fluid needed to completely obscure a candle flame viewed through it. Now one JTU is defined as the turbidity caused by the distributed presence of 1 mg of diatomaceous fullers earth³⁷ (an inert material) or 1 mg of finely powdered silica (SiO₂) in 1 litre of distilled water.

Formazin turbidity unit (FTU). The internationally accepted unit for turbidity is the formazin³⁸ turbidity unit (FTU). 1 FTU equals 10 JTU.

Formazin suspension is produced by the polymerisation of hexamethylene-tetramine and hydrazin sulphate under strictly controlled conditions.

Nephelometric turbidity unit (NTU). The USEPA³⁹ defines nephelometric turbidity unit (NTU) which is based on formazin. The ISO⁴⁰ refers to its units as formazin nephelometric unit (FNU).

However, all these primary standards lack long term stability and reproducibility. Usually a secondary standard, supplied by the instrument maker, is used to calibrate most of the modern turbidimeters. Among them, the turbid glass standard is the best. It consists of a specially formed glass cube in which small particles are uniformly embedded.

To have an idea of the uses of units, the ranges of accepted turbidity of water for different uses are given in Table 13.13.

Table 13.13 Acceptable ranges of turbidity of water

| <i>Designated uses</i> | <i>Range in NTU</i> |
|---------------------------------|---------------------|
| Human consumption | 1 to 5 |
| Recreation (e.g. swimming pool) | 5 |
| Aquatic life | 10 to 50 |

³⁷ Also called *kisselguhr*.

³⁸ A liquid having the appearance of milk.

³⁹ United States Environmental Protection Agency.

⁴⁰ International Standards Organisation.

Methods of Measurement

Usually turbidity measurement is considered a cheap estimate of Total Dissolved Solids (TDS) concentration. TDS measurement is tedious and time-consuming. It involves:

- Step 1. Filtering a known volume of the sample through a pre-weighed filter so that all suspended particles of diameter 1 μm and above are collected.
- Step 2. Drying the filter overnight by keeping it at $\sim 103^\circ\text{C}$ to remove all liquid in the residue and filter paper.
- Step 3. Re-weighing the filter.

The process demands good attention when the TDS is low. The turbidity measurement, on the other hand, is quick and reliable for over a good range of TDS.

We have already discussed that a liquid appears turbid because of scattering of light by suspended particles. So, turbidity is really a measure of how the liquid scatters light. It can be measured by shining a controlled light source onto a sample and estimating the absorbed or scattered radiation. But in water-bodies a visual measurement of clarity of water is normally made. Though turbidity is different from clarity, which is measured in terms of length, the two are interrelated. We first discuss two widely used methods of clarity measurement, namely Secchi disc and turbidity tube.

Secchi disc

A *Secchi*⁴¹ disc is a circular plate of 10 cm diameter. It is divided into quadrants, and alternate quadrants are painted black [Fig. 13.57(a)].

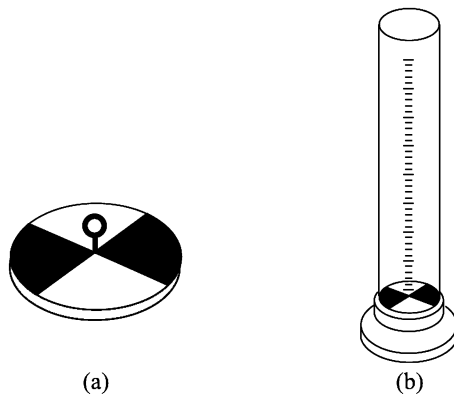


Fig. 13.57 (a) Secchi disc, and (b) turbidity tube.

The disc is attached to a cord and is slowly lowered into water until it is no longer visible. The corresponding depth of water is called the *Secchi depth*. A higher Secchi depth indicates higher clarity of water. The rule of thumb is, sunlight can penetrate 2 to 3 times the Secchi depth and support aquatic flora.

Though it has the advantage of being an in situ measurement, it is a visual measurement, the accuracy of measurement having dependence on the eyesight of the measurer, sun's glare on water, time of the day, etc. Also, the method cannot be used at shallow water.

⁴¹Pronounced as sek'kē. It was invented in 1865 by the Italian astronomer Pietro Angelo Secchi.

Turbidity tube

This is an adaptation of the Secchi depth measurement. The Secchi pattern, painted at the bottom of a measuring cylinder having a graduation of length, constitutes the turbidity tube [Fig. 13.57(b)].

The sample water is poured into the cylinder which is viewed from the top. The length of the water column that makes the pattern disappear is the required Secchi depth.

Forward scattering turbidimeter

Optical measurements of turbidity are carried out by measuring the intensities of the incident and absorbed/scattered light. If I_0 and I are the intensities of the incident and transmitted radiations, the well known Beer-Lambert relation connecting them is given by

$$I = I_0 \exp(-kNl) \quad (13.47)$$

where N is the number density of suspended particles

l is the length of the path traversed by light in the turbid medium

k is the constant.

The product kN is often called *turbidity attenuation coefficient*.

Equation (13.47) is useful for measuring turbidity by measuring transmittance or forward scattering. The measurement can be made by single beam and dual-beam methods.

Single beam method. The arrangement is shown in Fig. 13.58. The diagram is self-explanatory. I_0 can be measured by using clear (i.e. not turbid) liquid in the sample cell. However, the measurement is susceptible to errors arising out of decay of the source and presence of colour in the liquid.

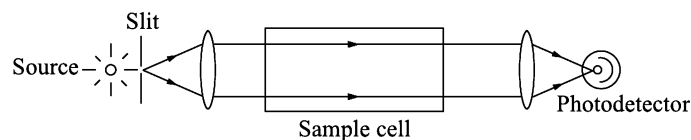


Fig. 13.58 Single beam forward scattering type turbidimeter.

Dual-beam method. To minimise the effect of light source decay and to eliminate the effect of colour of the sample liquid, the dual-beam forward scattering type turbidimeter was developed. The arrangement is shown in Fig. 13.59. Oscillating about 600 times per second, the mirror directs the light alternately to the measuring and reference cells. The measuring cell contains the sample fluid while the reference cell contains the same fluid from which suspended particles have been filtered out. The photocurrent generated by the photodetector is thus a measure of the intensity differential of the light emerging from the measuring and reference cells. This photocurrent is employed to modulate the opening of a mechanical shutter so that the resulting photocurrent is zero. The more the turbidity of the fluid, the further the shutter needs to be closed. Therefore, the position of the shutter can be graduated in units of turbidity.

Forward scattering or transmission type turbidimeters are suitable for measurements of rather high turbidity. At low turbidities, the absorption is small and hence the error in measurement becomes high.

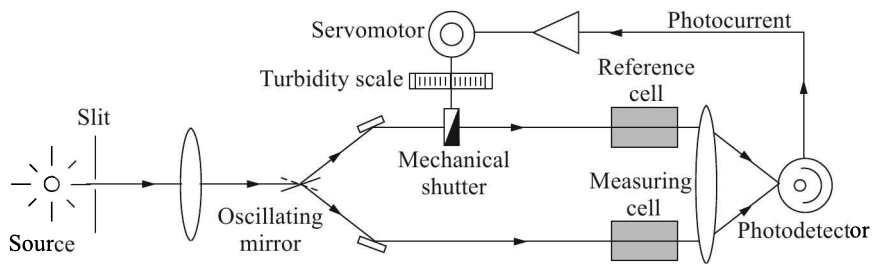


Fig. 13.59 Dual-beam forward scattering type turbidimeter.

Nephelometer

Nephelometers⁴², which utilise measurement of scattered radiation perpendicular to the incident radiation, are suitable for turbidity measurements over a wide range. We have seen in Section 13.9 at page 589 that the scattered radiation intensity varies linearly with the number density of suspended particles over a wide range.

Dual-beam nephelometer. This nephelometer uses a single light source which is split by an oscillating mirror into two beams—a measuring- and a reference-beam. The arrangement, shown in Fig. 13.60, is very similar to the corresponding transmission measurement except that the photodetector is placed perpendicular to both the reference and measuring cells.

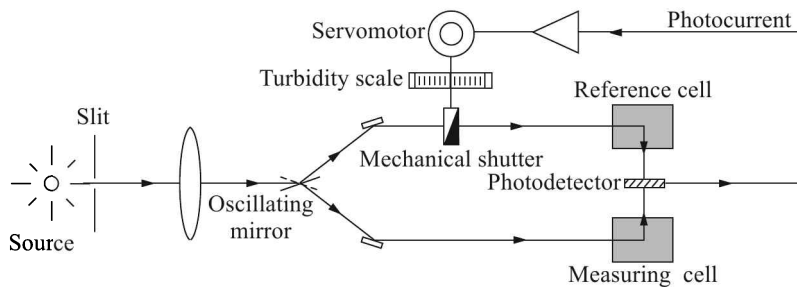


Fig. 13.60 Dual-beam nephelometer.

Since the measurement is made differentially with a single beam and a single photodetector, this method reduces the need for frequent calibration, and when used with a monochromatic light source, almost eliminates the need for calibration. But the method does not address the problem of measurement when turbidity or particle size is high. The four-beam method, described below, takes care of this problem.

Four-beam nephelometer. As shown in Fig. 13.61, the four-beam nephelometer uses two light sources and two photodetectors.

The measurement is made in two phases. During the first phase, LS1 pulses a beam directly onto PD1. Simultaneously, the PD2 measures scattered intensity at 90° . During the second phase, LS2 pulses a beam directly onto PD2 while PD1 measures the scattered radiation at 90° . These two phases of measurement provide two measurements of transmission

⁴²Derived from the Greek word for cloud, *nephelē*.

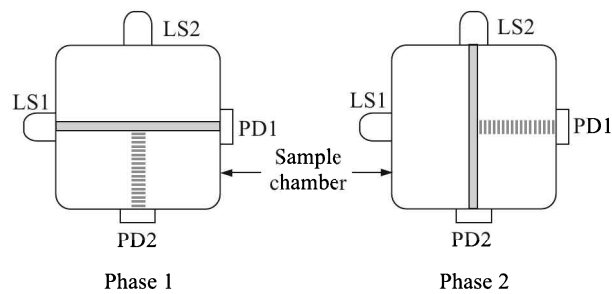


Fig. 13.61 Four-beam nephelometer. LS1, LS2: light sources, PD1, PD2: photodetectors.

and two of scattered radiation. A microprocessor uses a ratiometric algorithm to calculate the turbidity value from the four readings. This eventually eliminates the errors arising out of all interferences—colour, source decay, particle size.

Back-scatter nephelometer. Turbidimeters that have sample chambers are not free from the problem of deposit build-up on the walls of the chamber. This hinders the passage of the light beam through the sample and acts as a source of error. Surface-scatter nephelometers do not suffer from this problem. The schematic diagram of a back-scatter nephelometer is shown in Fig. 13.62. The figure explains itself.

The disadvantage of these meters is that they operate (i) at low sample flow rates and (ii) at atmospheric pressures.

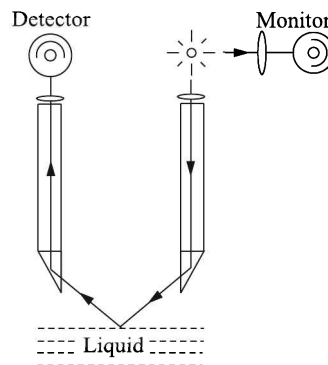


Fig. 13.62 Back-scatter nephelometer.

13.10 Opacity Measurement

Opacity of gases is akin to turbidity of liquids. The measurement of opacity is necessary for monitoring

1. The quality of ambient air
2. The stack emission
3. The visibility especially at airports
4. The particulate concentration in clean rooms in semiconductor production facilities
5. Dry powder or slurry processes such as pigments, catalysts, clays, cement, plastic, etc.

Suspended particles of dust and smoke in gas absorb, reflect and scatter incident radiation. Quantitative study of scattering of light by fine particles ($\phi < \lambda$) was first made by Rayleigh⁴³ who showed that the intensity of the scattered radiation is proportional to λ^{-4} . However, if the particle size is bigger than the wavelength of radiation, the intensity attenuation of the incident light is not due to Rayleigh scattering but mainly due to diffraction. Which is why tobacco smoke ($\phi < \lambda$) has a bluish tinge while chalk dust ($\phi > \lambda$) falling across a beam of light does appear greyish. This colour variation with particle size was first experimentally studied by Tyndall.

Opacity measurement for visibility studies should be carried out in such a way that it corroborates that observed by a human eye. The human eye is not equally sensitive to the entire range of the visible spectrum. It has maximum sensitivity around the region of $\lambda = 0.55 \mu\text{m}$, which is called the *photopic region*. So, the opacity measurement light should emit radiation in the range of 0.38 to 0.78 μm to reproduce human eye perceptions. Most particle sizes in opacity studies lie between 0.1 and 50 μm .

Units and Definitions

From Eq. (13.47), the transmittance T of a ray of light through a sample is given by

$$T = \frac{I}{I_0} = \exp(-kNl) \equiv 10^{-k'Nl}$$

where k' is a constant. The opacity, O is defined as

$$O = 1 - T \quad (13.48)$$

Equation (13.48) can be utilised to determine the dust loading in the stack in mg/m^3 . Let us define optical density D as

$$D = -k'Nl$$

Then, it is easy to see that D and O are interconnected as

$$D = \log_{10} \left(\frac{1}{1 - O} \right) \quad (13.49)$$

From Eqs. (13.48) and (13.49) one may find the interrelation between D , T and O as given in Table 13.14.

Table 13.14 Interrelation between D , T and O

| D | T | O |
|-----|--------------|----------------|
| 1.0 | < 10% | > 90% |
| 2.0 | $\simeq 0\%$ | $\simeq 100\%$ |

So, when optical density reaches 2.0, the sample appears completely opaque. The corresponding dust loading is $250 \text{ mg}/\text{m}^3$. These data give the dust loading vs. optical density curve as given in Fig. 13.63.

Smoke density monitors are available commercially in different ranges. Some of the standard ranges of such monitors are given in Table 13.15.

⁴³Pronounced as *ráyli*. John William Strutt Rayleigh (1842–1919) was a British physicist.

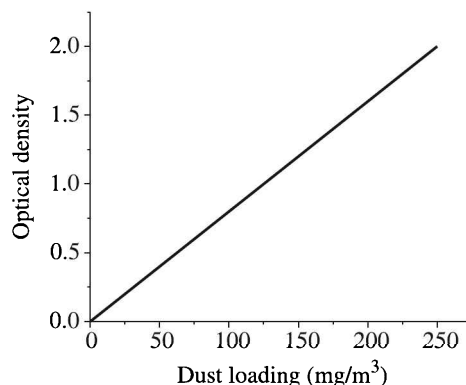


Fig. 13.63 Optical density vs. dust loading curve.

Table 13.15 Standard ranges of stack monitors

| D (mg/m ³) | % O |
|--------------------------|----------|
| 0 – 0.18 | 0 – 33.9 |
| 0 – 0.45 | 0 – 64.5 |
| 0 – 0.90 | 0 – 87.4 |
| 0 – 1.80 | 0 – 98.4 |

Apart from units of dust loading and % opacity, opacity is sometimes expressed in terms of Ringlemann card numbers. These cards, five in number, present graduated shades of grey from white to black. The opacity is determined by comparing the stack gas colour with that of Ringlemann cards.

Methods of Measurement

Dust loading measurement

Air is drawn by pump at a constant flow rate through a specially shaped inlet (Fig. 13.64) where particulate matter is collected according to size fractions on different filters. Each filter is weighed before and after use to determine the net mass collected. The total volume of gas filtered is known from the flow rate and the time of operation.

The recommended range of dust loading measurement by this method is 30 to 300 $\mu\text{g}/\text{m}^3$. Humidity and adsorbed water may interfere with the measurement.

Light scattering method

Nephelometry. It is a common experience to have seen small dust particles flying around in a ray of light coming through a window while the the same particles are not visible if we go out in the sun. This happens because the small concentration of dust particles appear visible owing to Tyndall scattering when viewed from a side in the window light. Nephelometers utilise this effect to measure particulate density in gas much in the same way as in the turbidity measurement. The more the scattered light intensity in a nephelometer, the more the opacity

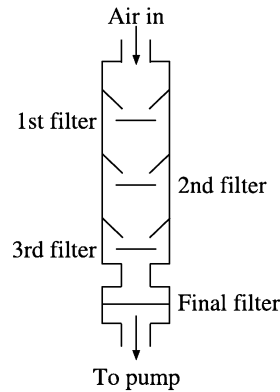


Fig. 13.64 Dust collection.

and the less the visibility. So, nephelometers can be calibrated in terms of visibility range as well, the usual span of visibility being 1.6 km to 160 km.

Open path spectrometry. Highly sophisticated spectrometric methods are available not only to determine concentration of particulate matter but also to analyse the percentage of different constituents of a gas mixture or air. Four such spectrometric techniques

1. Open path Fourier transformed infrared (OP-FTIR)
2. Open path ultraviolet (OP-UV)
3. Open path tunable diode laser (OP-TDLAS)
4. Open path hydrocarbon

are available. A discussion of these techniques is beyond the scope of this text. However, to have a feel of the methods, we describe in brief the OP-TDLAS technique. The arrangement is shown in Fig. 13.65.

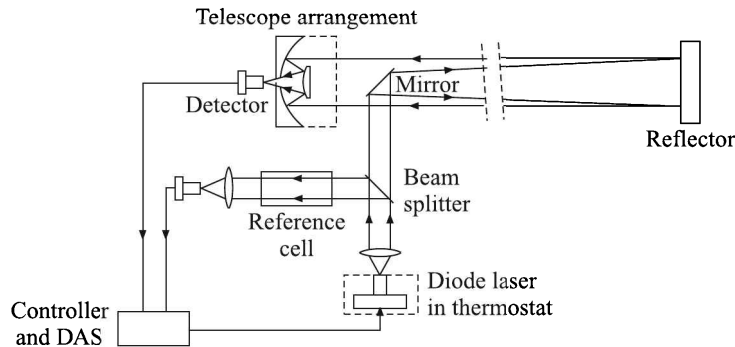


Fig. 13.65 TDLAS arrangement

A diode laser beam is split into a transmitted and a reflected beam by a beam splitter. The transmitted beam, reflected by a mirror, travels through a distance of about 1 km when it interacts with target gas molecules and/or aerosols and eventually gets reflected by reflector array. This array sends the light back to a telescopic arrangement which focuses it to a photodetector.

The beam reflected at the beam splitter goes through a reference cell that contains a sample of the analyte gas. The reference gas cell helps to tune the tunable diode laser to emit the light of desired frequency which corresponds to a signature line of the target molecule/aerosol. An analysis of the resulting spectra yields the required data of particulate composition and percentage of constituents.

Tapered element oscillating microbalance (TEOM)

The TEOM consists of a tapered glass element with a filter attached to it. The element is made to oscillate in its natural frequency. When air is drawn through the element, the frequency of oscillation decreases owing to particulate deposition on the filter. So, the frequency of oscillation of the element can be calibrated in terms of dust loading.

A variant of the microbalance uses two identical piezoelectric crystals, oscillating in their resonant frequency. When particulate matter is trapped on the sensing crystal, it alters its frequency as a result. A measurement of the frequency change in respect of the reference crystal provides the information about dust loading. Commercially available instruments show a frequency change of ~ 2 kHz/ μg of mass deposited. However, the method of trapping of the particulate matter on the crystal is not thoroughly satisfactory. Impacting and electrostatic methods have been tried though both suffer from weaknesses.

Beta attenuation

The transmission of beta particles⁴⁴ gets attenuated when it encounters particulate matter deposited on a filter tape. The attenuation is a measure of the mass deposited on the filter. The required arrangement for this measurement is shown in Fig. 13.66.

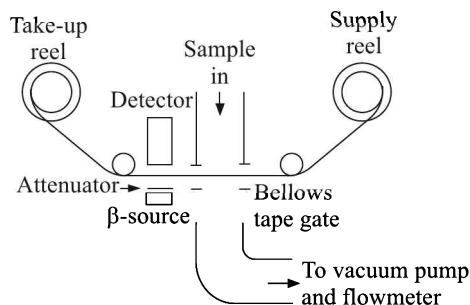


Fig. 13.66 Dust loading measurement by β -attenuation method.

Automated samplers draw air through a continuous filter tape where particulate is collected. Attenuation of the β -ray is first measured in the clean state and then in the dirty state. Flow rate through the tape has to be accurately controlled. Low energy β -ray emitted by ^{14}C is normally used from human exposure considerations. The attenuation is somewhat more responsive to hydrogen compounds though, in general, it is not sensitive to chemical composition of the particulate matter.

⁴⁴Electrons of energy between 0.01 and 1 MeV emitted by radioactive substances.

Triboelectricity generation

Theoretically, particulate impact on the surface of an electrode should generate triboelectricity⁴⁵. The amount of charge thus generated should be a measure of the particulate mass. But devices based on this principle have failed to gain wide acceptance owing to a doubtful correlation between the transferred charge and particulate concentration.

Surface ionisation

Instruments based on the principle of surface ionisation consist of a hot wire and a concentric cylindrical electrode. Airborne particles decompose when they come in contact with the hot wire. The resulting release of ions, collected at the electrode, generates an ion current. Through an analysis of an average ion current and heights of peaks and the rate of their occurrence, the instrument provides information about both mass concentration and relative distribution of particle sizes.

The disadvantage of this method is that it offers measurement at a point and therefore, fails to provide information about the overall stack.

Review Questions

- 13.1 (a) Describe the IR absorption method of measuring moisture in a suitably formed sample.
(b) Define the following terms:
(i) Absolute humidity
(ii) Specific humidity
(iii) Relative humidity
(iv) Dew point.
- 13.2 (a) What is the working principle of electrolysis-type hygrometer?
(b) Describe the set-up of a commercial-type dew-point meter.
(c) How can the moisture content of granular materials be obtained? List the advantages of the capacitance-type moisture measurement.
- 13.3 (a) Define the term pH of a solution. What is the working principle of pH measurement?
(b) Explain the construction and working principle of calomel electrode and glass electrode for measuring the pH of an unknown solution.
- 13.4 (a) Define (i) Absolute humidity, and (ii) Relative humidity.
(b) What is a psychrometer? Describe an industrial psychrometer and explain its operation.
(c) Describe the IR absorption method of measuring moisture in a suitably formed sample.
- 13.5 (a) How is the conductivity of an aqueous solution defined? How does dilution affect the conductivity? What is molar conductivity?
(b) Explain briefly the working of an electrodeless (i.e. toroidal) conductivity analyser. What is the chief advantage of such type of measurement?

⁴⁵Electricity generated by friction.

13.6 Explain with diagram the operation of a liquid level measurement system using an electrical type transducer.

13.7 Indicate the correct answer:

- (a) Measurement of viscosity involves measuring
- (i) Frictional force
 - (ii) Coriolis force
 - (iii) Centrifugal force
 - (iv) Buoyant force
- (b) When the reading of a pH meter changes from 5 to 7, the hydrogen ion concentration of the solution is
- (i) Halved
 - (ii) Doubled
 - (iii) Increased 100 times
 - (iv) Decreased 100 times
- (c) C_1 and C_2 are the activities of the ions on the two sides of a membrane. The Nernst potential developed across the membrane is proportional to
- (i) $\frac{C_1}{C_2}$
 - (ii) $\frac{C_1^2}{C_2^2}$
 - (iii) $\log_e \frac{C_1}{C_2}$
 - (iv) $\exp \frac{C_1}{C_2}$
- (d) The value of pH of a solution is 4. It indicates that the concentration of hydrogen ions is
- (i) 10^{-4} g/litre and the solution is acidic
 - (ii) 10^{-4} g/litre and the solution is alkaline
 - (iii) 10^{-4} mg/litre and the solution is acidic
 - (iv) 10^{-4} mg/litre and the solution is alkaline
- (e) A capillary viscometer, with known dimensions, is used for measuring the dynamic viscosity of oil. In order to obtain viscosity, it is necessary and sufficient if one measures
- (i) pressure drop across the capillary
 - (ii) volume of fluid collected in a given period of time
 - (iii) both (i) and (ii)
 - (iv) not only (i) and (ii), but also one must ensure that the flow is laminar
- (f) In a falling-ball viscometer, the ball attains terminal velocities of 0.01 m/s for oil A and 0.002 for oil B. Assuming the oils have the same density and oil A has a kinematic viscosity of 5×10^{-3} m²/s, the kinematic viscosity of oil B in m²/s is

- (i) 15×10^{-3}
- (ii) 20×10^{-3}
- (iii) 25×10^{-3}
- (iv) 30×10^{-3}

13.8 In a laminar flow experiment, fluid A is pumped through a straight tube and the volumetric flow rate and pressure drop per unit length are recorded. In a second straight tube having twice the internal diameter of the first one, fluid B records the same pressure drop per unit length at the same volumetric flow rate. Assuming fully developed flow conditions in the tubes, the ratio of the dynamic viscosity of fluid B to that of fluid A is

- (a) 16
- (b) 32
- (c) 64
- (d) 128

13.9 A glass electrode with a sensitivity of 59 mV/pH and resistance of $10^9 \Omega$ is used to measure pH with a range of 0–14. The electrode is connected to a recorder of input range 0–100 mV and resistance 100Ω using a buffer amplifier with output resistance 100Ω .

- (a) Calculate the impedance of the amplifier and the sensitivity of the recorder scale to obtain an accurate reading of pH.
- (b) The resistance of the electrode has increased to $2 \times 10^9 \Omega$ due to chemical action over time. Calculate the resulting error for a true pH of 7.0.

13.10 Viscosity μ is given by

$$\mu = \frac{\pi r^4 (p_1 - p_2)}{8QL}$$

where r (radius of the capillary tube) = 0.5 ± 0.01 mm
 p_1 (pressure at the inlet) = 200 ± 3 kPa
 p_2 (pressure at the exit) = 150 ± 2 kPa
 L (length of the capillary tube) = 3 m
 Q (volume flow rate) = 4×10^{-7} m³/s

- (a) Calculate the viscosity and specify the unit
- (b) Calculate absolute error
- (c) Calculate root sum square error.

Analytical Instrumentation

Any analysis of a species in a mixture is made by its specific properties which are different from others. The mixture may be in any of the three states of matter—gas, liquid or solid. We will begin with the industrial gas analysis, though we will deal with the general methods of analysis of other states of matter later.

14.1 Industrial Gas Analysis

Gases are distinguishable by their

1. Thermal properties
2. Magnetic susceptibilities
3. Absorption/emission of electromagnetic radiation
4. Chemical or electrochemical reactions, with the exception of inert gases

Of course, considering at micro levels, all these properties arise out of molecular and/or atomic behaviour of gases, and there are still other methods such as mass spectrometry, NMR, chromatography, etc. which exploit the micro level behaviour of atoms or molecules to quantify the presence of gases in a mixture.

In industrial gas analysis, primarily the principle enumerated above is followed. Instead of presenting methods for analysis of all industrial gases, we will first discuss two simple methods based on thermal properties of gases.

Methods based on Thermal Properties

Two such methods are commonly used. They are based on

1. Thermal conductivity
2. Heat of reaction (or catalytic combustion) study

Thermal conductivity analyser

Thermal conductivity analyser is one of the most widespread and universal methods of gas analysis in industries. The success of application of the method to determine the composition of a gas mixture depends on how widely the coefficients of thermal conductivities of constituent gases differ. If these values are pretty close, the method is not likely to yield reliable results.

Thermal conductivity coefficients of common gases. Before we discuss the method of gas analysis, let us have a look at the values of the coefficients of thermal conductivity λ for the common industrial gases (Table 14.1).

Table 14.1 Coefficients of thermal conductivity λ_0 of gases at 0°C

| <i>Gas</i> | $\lambda_0 \times 10^4$ (W/m-deg) | $\beta \times 10^4$ (/deg) | <i>Gas</i> | $\lambda_0 \times 10^4$ (W/m-deg) | $\beta \times 10^4$ (/deg) |
|-----------------|--------------------------------------|-------------------------------|-------------------|--------------------------------------|-------------------------------|
| Acetylene | 190.0 | 48 | Helium | 1457.0 | 18 |
| Air | 244.0 | 28 | Hydrogen | 1740.0 | 27 |
| Ammonia | 218.0 | 48 | Hydrogen sulphide | 131.0 | – |
| Argon | 167.0 | 30 | Methane | 302.0 | 48 |
| Butane | 135.0 | 72 | Methyl iodide | 47.3 | – |
| Carbon dioxide | 146.0 | 48 | Nitrogen | 243.0 | 28 |
| Carbon monoxide | 236.0 | 28 | Oxygen | 246.0 | 28 |
| Chlorine | 78.7 | – | Pentane | 130.0 | – |
| Chloroform | 66.0 | – | Propane | 150.0 | 73 |
| Ethane | 182.0 | 65 | Sulphur dioxide | 85.4 | – |
| Ethylene | 175.0 | 74 | Xenon | 51.9 | – |

Relations for gas analysis. The recorded values in the second column of Table 14.1 correspond to 0°C. At any other temperature $T^\circ\text{C}$, if the coefficient of thermal conductivity is denoted by λ_T , it is related to λ_0 as

$$\lambda_T = \lambda_0(1 + \beta T)$$

where β , the temperature coefficient of thermal conductivity, is shown in the third column of Table 14.1. It needs to be mentioned in this context that β is also temperature-dependent. But the fixed values given in Table 14.1 are valid over the temperature range from 0°C to 100°C.

For a mixture of polar and nonpolar gases, the resulting coefficient of thermal conductivity λ_m is given by

$$\lambda_m = \frac{c_1 \lambda_1 \sqrt[3]{M_1} + c_2 \lambda_2 \sqrt[3]{M_2} + \dots}{c_1 \sqrt[3]{M_1} + c_2 \sqrt[3]{M_2} + \dots} \quad (14.1)$$

where λ_i is the coefficient of thermal conductivity of the i th component
 c_i is the concentration in volume per cent of the i th component
 M_i is the molecular weight of the i th component.

The variation of thermal conductivities of three binary mixtures is shown in Fig. 14.1.

The scale on the left of the graph pertains to the mixture of CO_2 and H_2 while that on the right pertains to the other two mixtures. It may be seen from the graph that in the case of a mixture of a polar and a nonpolar gas (curve 3), the variation is maximum. For such cases, the λ_m is given by the relation

$$\lambda_m = (c_n \lambda_n + c_p \lambda_p) \left(1 + \frac{c_n - c_p^2}{3.5} \right) \quad (14.2)$$

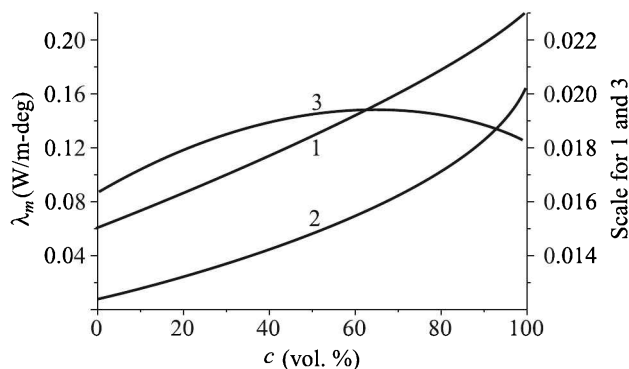


Fig. 14.1 Variation of thermal conductivities of gas mixtures. Curves: 1 for Ar + N₂, 2 for CO₂ + H₂, and 3 for C₂H₂ + CH₃OH (vapour).

where subscripts n and p indicate nonpolar and polar components respectively. However, in the case of a gas mixture of polar or nonpolar components whose molecular weights are close, the relevant equation is

$$\lambda_m = c_1 \lambda_1 + c_2 \lambda_2 \quad (14.3)$$

Equations (14.1), (14.2) and (14.3) are all empirical relations. They can all be extended for multicomponent mixtures. But, we must remember that since we will measure only λ_m , we need to treat the multicomponent mixture as a binary mixture, with x as the concentration of one component and $(1 - x)$ as that of other components taken together.

With this background, let us now look at the method of measurement of the coefficient of thermal conductivity.

Method of measurement. In thermal conductivity analysers, the coefficients of thermal conductivity of the sample gas and a known gas mixture are compared. Owing to its experimental complexity and low accuracy, the absolute value of the coefficient of thermal conductivity is not measured.

The schematic diagram of a thermal conductivity analyser is shown in Fig. 14.2. The sensing element, situated at the middle of the cells, is a resistor wire. It is made of a material having a high temperature coefficient of resistance like platinum, tungsten, nickel or chromium coated¹ copper.

The resistance wires are heated by sending a constant current through them. The ratio of voltages of the two cells may be measured by standard methods. Alternatively, the current terminals may be dispensed with and voltage terminals may be connected to the two arms of a Wheatstone bridge. The former method is preferred because it almost eliminates the so-called *end cooling* of the resistance wires.

Suppose, we want to determine the percentage composition of hydrogen and carbon dioxide in a mixture of two gases. Then, initially, pure hydrogen is passed through both the cells at a certain flow rate and the ratio of voltages is set to 1:1. Next, the sample gas mixture is passed through the sample cell and hydrogen is passed through the reference cell while maintaining the previous flow rate for both.

¹Because copper is susceptible to oxidation.

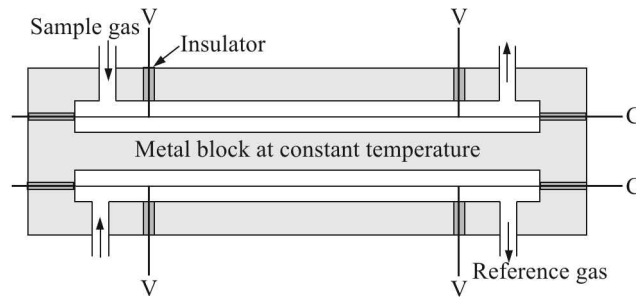


Fig. 14.2 Schematic diagram of a thermal conductivity analyser. C-C indicate current leads and V-V indicate voltage measurement leads.

Heat transfer from the hot wires to the metal block, maintained at a constant temperature, takes place by conduction through the gas and terminals (end cooling), convection, and radiation. Convection, owing its origin to gravity, can be minimised by using a cell of low diameter and keeping it horizontal. The radiation can be minimised by polishing the inner wall of the cell. In this way, in commercial thermal conductivity cells the various types of heat transfer are:

1. Convection – 2%
2. Radiation – 1%
3. End cooling – about 30%

But because of the comparison of thermal resistance changes of the reference and sample cells, these errors become negligibly small.

The bridge unbalance voltages can be calibrated for different gases at different concentrations. From these data the composition of an unknown mixture having known components can be figured out using Eqs. (14.1), (14.2) or (14.3).

However, if a single cell is used to measure thermal conductivity then, assuming heat transfer takes place only by conduction through the gas, we can express the voltage on the measuring diagonal of an equal-arm Wheatstone bridge by

$$V = \frac{IRR_1R_0\alpha}{k(R + R_1)} \cdot \frac{\lambda_1 - \lambda_2}{\lambda_1\lambda_2} \quad (14.4)$$

where I is the current passing through the sensing element
 R is the resistance of the sensing element in the cell
 R_1 is the resistance of the contiguous arm of the bridge
 R_0 is the resistance of the sensing element at 0°C
 α is the temperature coefficient of resistance of the sensing element
 k is the geometrical factor of the cell
 λ_i is the thermal conductivity of the i -th component of the gas mixture.

From Eq. (14.4) it is clear that apart from the thermal conductivities of component gases, the voltage depends on the current passing through the sensing element. The heating caused by this current changes the value of the thermal conductivity. So, it is absolutely necessary to keep the current constant at a known value.

The pressure of the gas does not affect the value of the thermal conductivity if it is maintained in the range of 5 to 10 cm of Hg.

Catalytic combustion analyser

The principle is to mix the sample, which must be a fuel gas, with oxygen and oxidise it through combustion in one cell while keeping the mixture intact in a reference cell. The comparison of two identical resistors placed in both the cells will give the quantity of heat generated in combustion and thereby the percentage of oxygen in the sample gas. A typical set-up is shown schematically in Fig. 14.3.

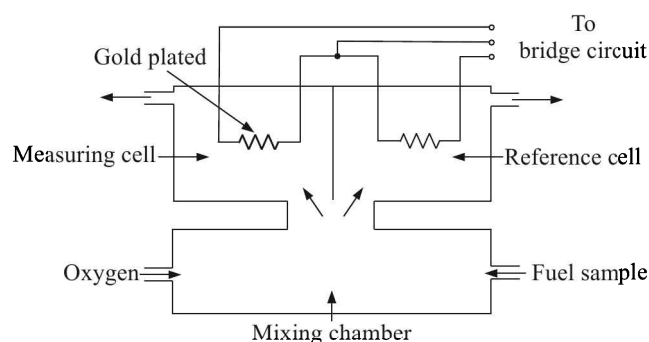


Fig. 14.3 A schematic set-up of catalytic combustion analyser.

The sample fuel gas and oxygen are mixed in the mixing chamber. The flow rate of gases is typically about $150 \text{ cm}^3/\text{hour}$. The gas mixture then enters two cells—the measuring cell and the reference cell. Both cells are provided with a filament resistance of equal value, but the measuring cell filament is provided with a catalytic noble metal surface that can oxidise the fuel. The reference cell filament only compensates the variations of temperature and conductivity of the sample. So, once the gas mixture enters the measuring cell, a combustion takes place. The resulting temperature changes the resistance of the measuring filament. The resistance of the measuring cell filament and that of the reference cell filament, where no combustion takes place, are compared in a bridge to sense the change in temperature. The bridge unbalance current, which is proportional to the temperature difference between the measuring and reference filaments is, therefore, proportional to the sample fuel gas concentration that supports combustion.

This procedure is also known as *heat of reaction method* of analysis.

Now we will discuss some specific methods of analysis of two gases individually, and then discuss general methods that apply to all gases.

Oxygen Analysis

Analysis of the presence of oxygen is important for two types of applications;

1. Applications where oxygen is necessary for oxidation and combustions
2. Applications where the contamination of oxygen needs to be prevented such as in the production of pure inert gases

Apart from general analytical methods, common for all gases, the oxygen-specific analysers can be divided into three categories, namely

1. Paramagnetic analyser
2. Electrochemical analyser
3. Catalytic combustion analyser

Among these three, we will discuss only the first two, the third having been discussed before in the previous section.

Paramagnetic oxygen analyser

A paramagnetic substance, as it is well known, is drawn to the stronger area of magnetic field while a diamagnetic substance behaves just in the opposite way. Among gases, oxygen, nitric oxide (NO) and nitrogen dioxide (NO₂) are paramagnetic. Hydrogen and carbon monoxide (CO) too are feebly paramagnetic while the rest of the gases are diamagnetic at ordinary temperatures. Among the paramagnetic gases, their relative magnetic susceptibilities, taking oxygen's as 100, are shown in Fig. 14.4.

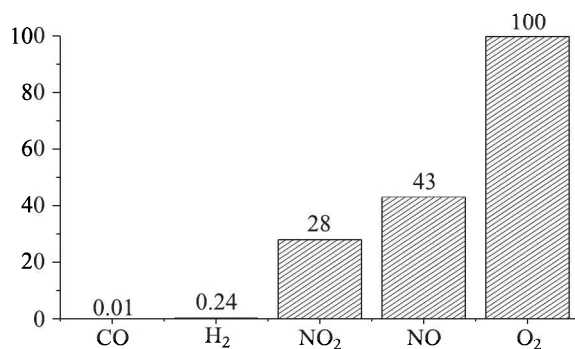


Fig. 14.4 Relative susceptibilities of paramagnetic gases.

However, other paramagnetic gases are not normally present in processes where oxygen needs be monitored. Which is why the paramagnetic property of oxygen can be utilised in its analysis. Three types of such oxygen analysers are in use. They are

1. Deflection type
2. Magnetic wind type
3. Differential pressure type

Deflection-type oxygen analyser. First introduced by the legendary chemist Linus Pauling² and his co-workers in 1946, this paramagnetic oxygen analyser is schematically shown in Fig. 14.5.

It consists of a glass dumbbell suspended by quartz fibre between the poles of a permanent magnet. The dumbbell is filled with nitrogen or some other gas of low magnetic susceptibility.

²Linus Carl Pauling (1901 – 1994) was an American chemist, biochemist, peace activist, author, and educator. He was awarded the Nobel Prize in Chemistry in 1954 and in 1962, the Nobel Peace Prize.

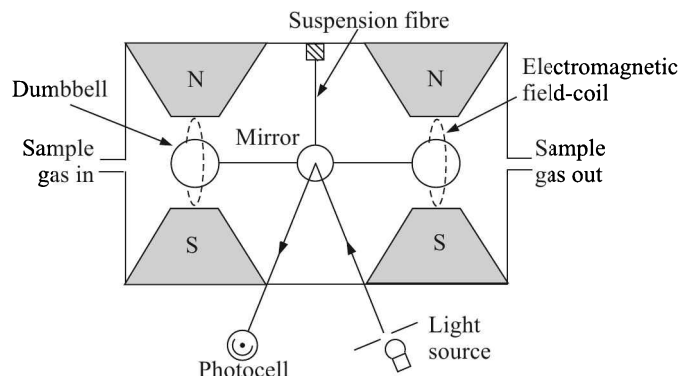


Fig. 14.5 Deflection-type paramagnetic oxygen analyser.

Magnetic pole pieces are in the shape of a wedge so that the field between them is non-uniform. Without the sample gas, the dumbbell stays slightly off from the strongest part of the magnetic field because of its nitrogen content. As the sample gas is let in, its oxygen content occupies the strongest part of the magnetic field and thus displaces the dumbbell further to cause a deflection of the light incident on the mirror in the suspension fibre. The force F that will be acting on each dumbbell is

$$F = K(\chi - \chi_d)$$

where χ is the magnetic susceptibility of the sample gas

χ_d is the magnetic susceptibility of the material and content of the dumbbell

K is a function of the magnetic field and its gradient.

Displacement of the dumbbell upsets the light balance in the photocell. This produces a proportional current in the electromagnetic field coils to bring the dumbbell back to its original position. This proportional current can be calibrated to the percentage of oxygen content in the sample gas.

Though sensitive, the instrument is delicate and it has to be installed on a vibration-free pedestal. Also, the sample gas needs to be filtered because dirt in the sample is likely to cause problems.

Magnetic wind-type oxygen analyser. Paramagnetic susceptibility χ_p of a substance depends on the absolute temperature T according to the Curie³ law

$$\chi_p = \frac{C}{T} \quad (14.5)$$

where C is a constant. According to Eq. (14.5), the susceptibility of a paramagnetic substance decreases with increasing temperature. This property of paramagnetic oxygen is utilised to generate a flow which has been termed *magnetic wind*. Figure 14.6 shows the schematic diagram of such a magnetic wind generating oxygen analyser.

The instrument consists of a ring in the middle of which lies a horizontal glass tube. This horizontal tube houses two wound resistors which form two arms of a bridge. One of the

³Pierre Curie (1859 – 1906) French physicist who shared Nobel prize in 1903 with his wife Marie.

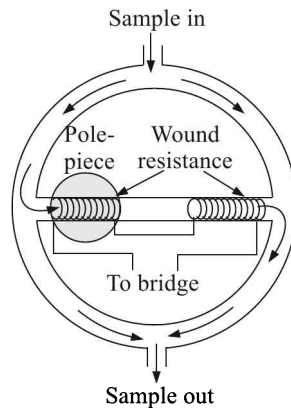


Fig. 14.6 Schematic diagram of magnetic wind-type oxygen analyser.

wound resistors is placed between the pole pieces of a magnet and a current flows through the resistors so that they produce heat. In the absence of oxygen in a gas flowing through the ring, the wound resistors are heated to the same extent and the bridge to which they are connected remains balanced.

If the sample gas contains oxygen, it is drawn to the left resistor where it gets heated to lose some susceptibility. Then the nearest cold gas on the left having more susceptibility, gets drawn to the magnetic field, pushing the hot gas there to the right. Thus, a magnetic wind is generated from the left to the right of the horizontal tube. As a result, the left resistor region becomes colder than the right one, throwing the bridge out of balance. The resulting bridge unbalance current will be proportional to the percentage of oxygen in the sample gas.

The instrument, though robust, suffers from the following sources of error:

1. Heating and cooling of resistors are not only function of the flow rate of the magnetic wind, but also function of the composition and pressure of the sample gas. Different gases possess different thermal conductivities and viscosities which affect heat transfer and flow rate, upsetting the calibration.
2. Even if the background gases in the sample remain the same with variation in oxygen content only and pressure change is compensated for, the calibration curve becomes nonlinear at higher magnetic field strengths. Though it is linear at lower magnetic fields, but there the sensitivity is less (Fig. 14.7).
3. Hydrocarbon and other combustible gases in the sample react on the heating coils, degrade them and the calibration becomes questionable.

This instrument is also known as *thermo-magnetic analyser*.

Differential pressure-type oxygen analyser. The differential pressure-type oxygen analyser uses two gases—reference and sample gas—and generates a pressure difference between them utilising the paramagnetic property of oxygen. The pressure difference gives rise to a flow in a horizontal tube. The flow rate, which is a function of the percentage of oxygen in the sample gas, is measured. The schematic diagram of the arrangement, shown in Fig. 14.8, actually lies in the horizontal plane.

Air may be used as the reference gas. Both the reference and sample gas flow through the sample chamber where one channel is kept under a strong magnetic field while the other

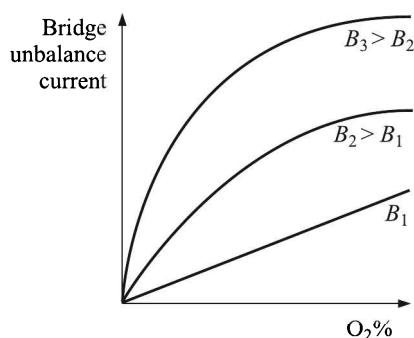


Fig. 14.7 Bridge unbalance current vs. $O_2\%$ curves at different magnetic fields.

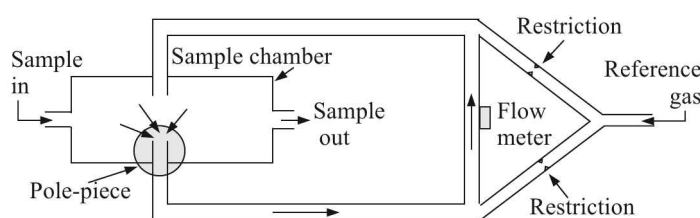


Fig. 14.8 Differential pressure-type oxygen analyser.

channel is free from magnetic influences. On account of its paramagnetic properties, oxygen gas builds up at the channel under the magnetic field causing a pressure differential between the two arms of the flow tube. This leads to a flow in the channel connecting the two arms. The flow is measured using an appropriate transducer.

The differential pressure-type oxygen analyser is rugged. But its functioning is affected by the vibration of the arrangement. Also, it is not suitable for analysis of oxygen if it is present as a trace.

Electrochemical oxygen analyser

Electrochemical oxygen analysers can be of three categories

1. High temperature zirconia fuel cell analyser
2. Ambient temperature galvanic analyser
3. Polarographic analyser

Zirconia fuel cell analyser. Zirconia fuel cell analyser uses a ceramic, zirconium oxide, as the solid electrolyte having porous platinum electrodes on its inner and outer surfaces. The cell is housed in a furnace that is maintained at 800°C . The sample gas is allowed to flow in the inner side of the cell while the flow of a reference gas, generally dry air, is maintained at the outer side of the cell (Fig. 14.9).

Oxygen ionises to O^{2-} ions as it comes in contact with electrode surfaces at 800°C temperature. The electrode facing the gas with higher concentration of oxygen generates oxygen ions, whereas the other electrode facing gas with a lower concentration of oxygen converts oxygen ions to neutral gas.

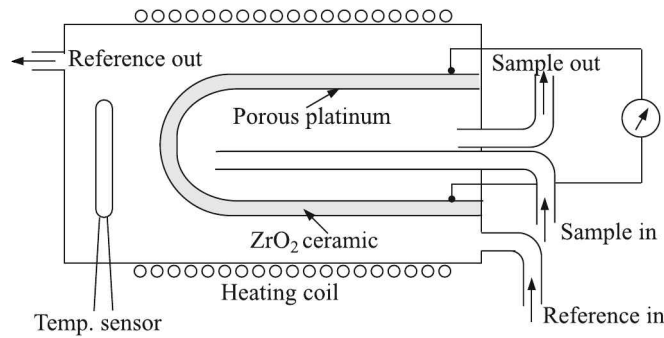


Fig. 14.9 Zirconia fuel cell oxygen analyser.

Electrolytic reactions are as follows:



The resulting emf can be calculated from the Nernst equation

$$E = \frac{2.303RT}{nF} \log \frac{p_r}{p_s} \quad (14.6)$$

where R = gas constant = 8.313 J/(mol-degree)

T = absolute temperature = (273.15 + 800) = 1073.15 K

F = faraday = 96,500 C/mol

p_r = partial pressure of O₂ in the reference gas

p_s = partial pressure of O₂ in the sample gas

n = number of electrons involved in the reaction = 4

Substituting the values of R , T , F and n in Eq. (14.6), we get

$$E = 53.23 \log \frac{p_r}{p_s} \text{ mV} \quad (14.7)$$

The typical cell output vs. oxygen concentration plot is shown in Fig. 14.10.

The following points emerge from Eqs. (14.6) and (14.7):

1. The cell output directly depends on the temperature of the cell. Hence, the temperature of the furnace needs to be held constant.
2. Since the partial pressure of O₂ in the ambient gas stays almost constant, the cell voltage depends logarithmically on the partial pressure of O₂ in the sample.
3. When $p_r = p_s$, $E = 0$. For $p_s > p_r$, the emf becomes negative, i.e. the electrodes change polarity.

The accuracy of measurement by this method is typically $\pm 0.1\%$ of O₂ when its concentration is 2%. The response time is nearly 5 s.

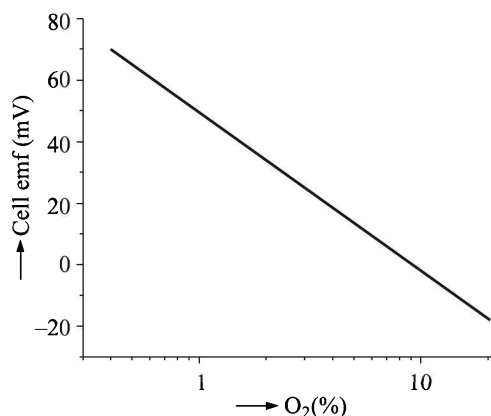
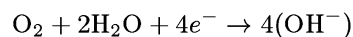


Fig. 14.10 Cell output vs. O₂ (%) plot.

Galvanic analyser. The electrical current of a galvanic analyser, which has appropriate electrodes and an appropriate electrolyte, depends upon the oxygen concentration of the electrolyte. The oxygen content of the electrolyte becomes equal to that of the sample through absorption. The cell acts spontaneously, without application of any external voltage. During the cell action the cathode reduces oxygen into hydroxide, accepting four electrons for each molecule of oxygen in the process



The anode, say lead, reacts with the OH⁻ ions releasing four electrons



These electrons cause a current to flow through the electrolyte. The magnitude of the current is proportional to the oxygen concentration of the electrolyte. Figure 14.11 shows the sketch of a probe-type galvanic oxygen analyser.

The cathode has to be noble metal—silver or gold—so that the cathode potential can reduce O₂ when the circuit is closed. The anode can be a base metal like lead, cadmium, copper or zinc. Electrolytes are generally KOH or KHCO₃ so that there is minimal dissolution of the anode when the circuit remains open. Diffusion of oxygen through the membrane, normally Teflon, initiates the cell action when the circuit is closed.

The ion current I at a given temperature is given by the equation

$$I = \frac{nF\alpha P_m C_s}{t}$$

where n is the number of electrons involved in the reaction

F is the faraday (96,500 C/mol)

α is the surface area of the cathode

P_m is the permeability coefficient of the membrane

C_s is the oxygen concentration of the sample

t is the thickness of the membrane

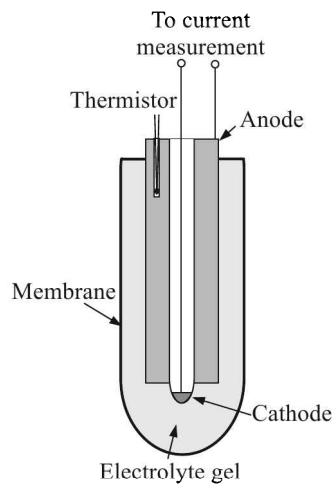


Fig. 14.11 Probe-type galvanic cell oxygen analyser.

Since electrons are supplied by dissolving the anode, the life of the cell is limited. Also, the cell functioning is highly temperature dependent and therefore it is necessary to make temperature compensation arrangement.

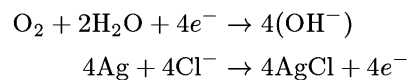
Instead of smooth electrode, the cathode can be a gauze of silver and the volume of the electrolyte can be reduced to increase the speed of response and sensitivity. The anode can be made of porous lead saturated with the electrolyte and dipped in a reservoir over which the sample gas flows. A third electrode may be added with a potential applied to it. This will extend the life of the anode, though limiting the current to some extent. This arrangement is known as *Hersch cell*.

Some background gases, like chlorine and other halogens, high concentrations of CO_2 , H_2S and SO_2 , are likely to contaminate the cell.

Galvanic detectors can be miniaturised with very low voltage and current levels and can be used from ppm level measurement to breathing air applications (20 to 25%).

Polarographic analyser. The polarographic cell is very similar to the galvanic analyser. But here both the electrodes are made of noble metals and a small voltage, about 0.8 V, is applied across the electrodes. The electrolyte is generally KCl (Fig. 14.12).

The oxidation-reduction equations for a gold-silver cell with KCl electrolyte are



As in galvanic cells, the generated current is proportional to oxygen concentration of the sample. The current is also temperature-dependent and therefore, the analyser calls for temperature compensation.

Otherwise rugged, the analyser, however has a low speed of response. This has limited its application to measurement of dissolved oxygen in liquids.

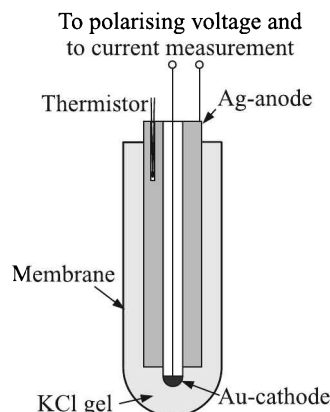


Fig. 14.12 Probe-type polarographic cell.

Other oxygen analysers

Mass spectrometer. Oxygen analysis through mass spectrometry is done for semiconductor, fermentation and pharmaceutical industries. Mass spectrometer is a general analytical instrument which can be used for analysis of other samples. We will consider the instrument later in Section 14.3.

Gas chromatograph. What has been stated for mass spectrometer vis-à-vis oxygen analysis holds true for gas chromatograph as well. We will consider the instrument in Section 14.2.

IR spectrometer. IR spectrometers are useful for analysis of substances that have absorption lines in the IR region of the spectrum. Normally, substances having considerable dipole moments do have absorption lines in the IR region. Homonuclear molecules like O_2 possess little dipole moments and so are not likely to have IR absorption lines.

But as an exception, O_2 does have an absorption line at 760 nm which corresponds to a transition from a metastable state of oxygen and is, therefore, classified as a forbidden transition. Normal spectrometry does not allow study of such transitions. However, with the advent of tunable diode laser absorption spectroscopy (TDLAS), this forbidden transition can be studied for oxygen analysis.

IR spectroscopy will be considered in Section 14.4.

Carbon Monoxide Analysis

Other than the standard methods—such as non-dispersive infrared (NDIR) analyser, gas chromatography—three methods based on the oxidation of CO to CO_2 are discussed here.

Catalytic analysis

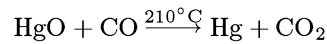
A catalyst, called *hopcalite*⁴—a granular mixture of oxides of copper, cobalt, manganese and silver—oxidises CO to CO_2 . The resultant temperature may be monitored and calibrated for CO concentration measurement. It is necessary to guard against interference by hydrocarbons.

Hopcalite is used in gas masks.

⁴(Johns) Hop(kins University and University of) Cal(ifornia) + *ite*.

Mercury vapour analyser

Hot mercuric oxide oxidises CO to CO₂ and gets itself reduced to metallic mercury vapour



The concentration of released mercury vapour may be assessed by photometry. As low as 0.025 ppm concentration of CO may be traced by this method (Fig. 14.13 for set-up) and a 10% change in this concentration may be detected. But hydrogen and hydrocarbons interfere in the measurement.

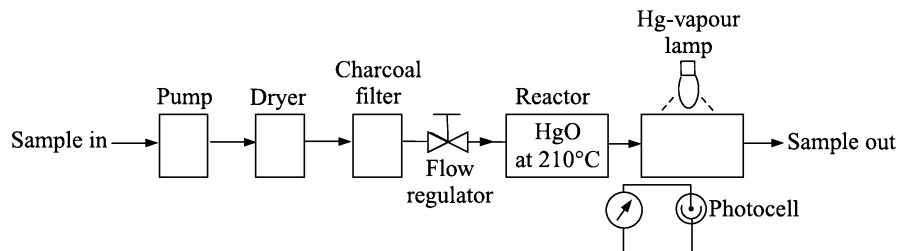


Fig. 14.13 Set-up of mercury vapour analyser for CO.

Electrochemical analyser

Iodine pentoxide reacts with CO at 150°C to liberate iodine. A galvanic cell may be constructed where this liberated iodine is absorbed by an electrolyte and will reach the cathode where it will be reduced. The resulting emf may be measured by a suitable arrangement.

Interference caused by hydrogen, hydrogen sulphide, acetylene, etc. needs to be tackled through their absorption by suitable agents.

Carbon dioxide and methane may be analysed by standard methods of gas analysis as discussed before.

14.2 Chromatography

A chromatograph is an analytical instrument that helps separate components of a mixture — gas or liquid—and estimate their relative abundance in the mixture.

The principle of operation of the instrument is simple and elegant. The analyte is dissolved in a *mobile phase*, which may be gas, liquid or supercritical fluid. The mobile phase, with the sample injected into it, is then forced through a *stationary phase*, which may be granular solid, granular solid soaked with a liquid or ion-exchange resin (Fig. 14.14). The tube, packed with the stationary phase, is called the *column*.

The stationary phase is so chosen that the components of the sample have different solubilities in it. As a result, the component that has a higher solubility will take longer time to pass through the column containing the stationary phase than the one which is not so soluble in the stationary phase. In fact, the cause of retardation may be

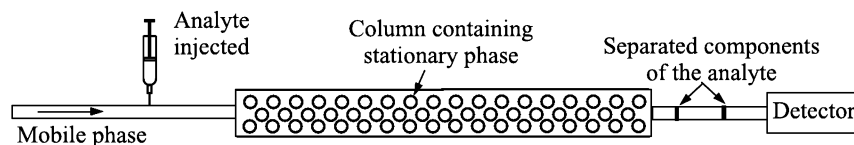


Fig. 14.14 Operation schematic of a chromatograph.

1. Adsorption
2. Solubility
3. Chemical bonding
4. Polarity or
5. Molecular filtration

of the sample. As a result of differential retardation, different components move through the column at different flow-rates and exit at different times from the column. A detector, placed at the exit of the column, will thus detect different components of the gas at different times. The resulting time vs. signal curve, called *chromatogram*, looks like a series of peaks having different peak heights (Fig. 14.15) which correspond to different components of the sample. In this way, the components of the analyte get separated after passing through the column.

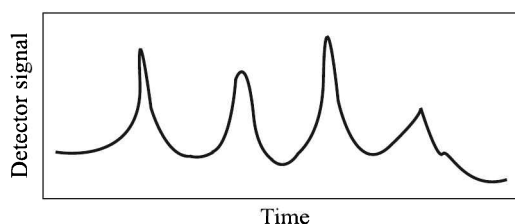


Fig. 14.15 Chromatogram.

Classification

Chromatography is broadly classified into two categories:

1. Gas chromatography
2. Liquid chromatography

The 'gas' and 'liquid' qualifiers actually indicate the state of the mobile phase as well as of the sample. Gas and liquid chromatography, in turn, can be divided into a few classes depending on the composition of the stationary phase. The entire classification is shown in Table 14.2

It may be noted here that

1. The first letters of the acronyms indicate the state of the mobile phase. There are two exceptions—IEC and EC—where the entire acronym refers to the interaction with the stationary phase.
2. The second letters of the acronyms indicate the state of the stationary phase.
3. In case the second letter indicates a liquid stationary phase, the packing material of the column is a liquid adsorbed on an inert solid. A bonded stationary phase indicates an organic substance bonded to a solid.

Table 14.2 Classification of chromatography

| <i>Category</i> | <i>Class</i> | <i>Acronym</i> |
|----------------------------|------------------------------------|----------------|
| Liquid chromatography (LC) | Liquid-solid chromatography | LSC |
| | Liquid-liquid chromatography | LLC |
| | Liquid-bonded phase chromatography | LBC |
| | Ion-exchange chromatography | IEC |
| | Exclusion chromatography | EC |
| Gas chromatography (GC) | Gas-solid chromatography | GSC |
| | Gas-liquid chromatography | GLC |
| | Gas-bonded phase chromatography | GBC |

4. The interaction between the analyte and the stationary phase is given in Table 14.3.

Table 14.3 Interaction between the analyte and stationary phase

| <i>Stationary phase</i> | <i>Interaction</i> |
|-------------------------|--------------------------------|
| Solid | Adsorption |
| Liquid | Partitioning |
| Bonded phase | Adsorption and/or partitioning |
| Ion-exchange | Ion-exchange reaction |
| Exclusion | Partitioning |

5. In IEC, the sample ions get separated by selective exchange with counter-ions of the stationary phase.
6. Exclusion chromatography (EC)⁵ relies on exclusion packing as the stationary phase brings about a classification of molecules depending mostly on molecular size and geometry.

Introductory Theories

Plate theory

The first successful theory of chromatographic process of separation, widely known as the plate theory, was proposed by AJP Martin and RLM Syngé⁶ in 1941.

It is well known that separations can be achieved in a fractional distillation column through several individual distillation stages. These stages—known as theoretical plates—are defined as the sections of the distillation column over which the vapour in its lower boundary is in thermodynamic equilibrium with the liquid in its upper boundary.

The Martin and Syngé plate model assumes that the chromatographic column, like the distillation column, contains a large number of separate layers or theoretical plates. Separate

⁵aka *gel-permeation chromatography*.

⁶Nobel prize winners in Chemistry for the invention of chromatograph.

equilibrations of the sample between the stationary and mobile phases occur in these plates. The analyte moves down the column by transfer of equilibrated mobile phase from one plate to the next.

These plates do not actually exist within a chromatographic column. They are imaginary plates which help estimate a column efficiency through their number N , and the height equivalent (HETP) H (Fig. 14.16).

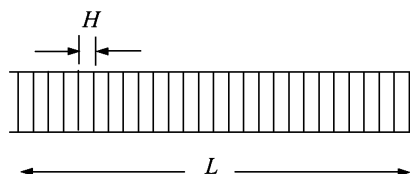


Fig. 14.16 Hypothetical division of the column in plates.

Thus, if L is the length of the column,

$$L = NH \quad (14.8)$$

Now, suppose a solute is placed in the first plate and then a small fraction of it, x , is transferred to the next plate. Then the contents of the first and second plates are $(1-x)$ and x respectively. Next, a second transfer of x fractions of the contents of the two plates takes place. Their revised contents as well as those of total four transfers are shown in Table 14.4. It is clear from the

Table 14.4 Results of transfer of x fraction from each plate for four transfers

| Transfer number (N) | Plate number (r) | Plate content | Systematic representation |
|----------------------------|-------------------------|---|---------------------------|
| 1 | 0 | $(1-x)$ | ${}^1C_0(1-x)$ |
| | 1 | x | 1C_1x |
| 2 | 0 | $(1-x) - (1-x)x = (1-x)^2$ | ${}^2C_0(1-x)^2x^0$ |
| | 1 | $(x-x^2) + (1-x)x = 2(1-x)x$ | ${}^2C_1(1-x)^1x^1$ |
| | 2 | x^2 | ${}^2C_2(1-x)^0x^2$ |
| 3 | 0 | $(1-x)^2 - (1-x)^2x = (1-x)^3$ | ${}^3C_0(1-x)^3x^0$ |
| | 1 | $2(1-x)x - 2(1-x)x^2 + (1-x)^2x = 3(1-x)^2x$ | ${}^3C_1(1-x)^2x^1$ |
| | 2 | $x^2 - x^3 + 2(1-x)x^2 = 3(1-x)x^2$ | ${}^3C_2(1-x)^1x^2$ |
| | 3 | x^3 | ${}^3C_3(1-x)^0x^3$ |
| 4 | 0 | $(1-x)^3 - x(1-x)^3 = (1-x)^4$ | ${}^4C_0(1-x)^4x^0$ |
| | 1 | $3(1-x)^2x - 3(1-x)^2x^2 + x(1-x)^3 = 4(1-x)^3x$ | ${}^4C_1(1-x)^3x^1$ |
| | 2 | $3(1-x)x^2 - 3(1-x)x^3 + 3(1-x)^2x^2 = 6(1-x)^2x^2$ | ${}^4C_2(1-x)^2x^2$ |
| | 3 | $x^3 - x^4 + 3(1-x)x^3 = 4(1-x)x^3$ | ${}^4C_3(1-x)^1x^3$ |
| | 4 | x^4 | ${}^4C_4(1-x)^0x^4$ |

systematic representation column of the Table that after N such transfers, the amount of solute on the r th plate, $q(r)$ is given by the binomial distribution⁷ function expression

$$q(r) = {}^N C_r (1-x)^{(N-r)} x^r \quad (14.9)$$

⁷ aka Bernoulli distribution.

This distribution is considered as the discrete form of Gaussian distribution. It is interesting to note that the division of the chromatographic column into plates automatically gives rise to a Gaussian profile of the solute migration (Fig. 14.17). The mean of the binomial distribution

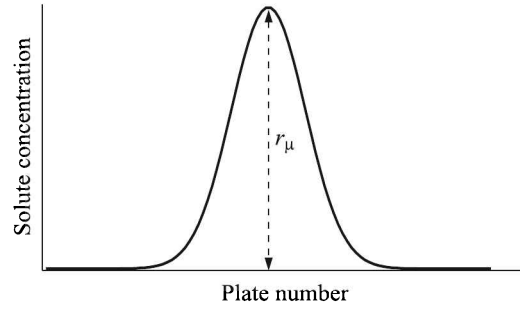


Fig. 14.17 The Gaussian profile of solute migration in a column.

μ is given by⁸

$$\mu = Nx \quad (14.10)$$

This corresponds to the maximum of the peak height, r_μ because of the symmetrical nature of the distribution. So,

$$r_\mu = Nx \quad (14.11)$$

The variance σ^2 is given by

$$\begin{aligned} \sigma^2 &= N.(1-x).x \\ &= N.x \quad [\because 1-x \simeq 1, x \text{ being small}] \end{aligned} \quad (14.12)$$

Consequently, the standard deviation σ is given by

$$\sigma = \sqrt{Nx} = \sqrt{r_\mu} \quad (14.13)$$

We note here that N and x are both numbers and therefore, we need to multiply σ by the HETP H to convert it to length. Thus

$$\sigma = H\sqrt{r_\mu} \quad (14.14)$$

Now, the variance, being equal to r_μ , can be expressed as L/H , where L represents the distance migrated by the solute. Inserting this value in Eq. (14.14), we get

$$\sigma = \sqrt{LH} \quad (14.15)$$

It is obvious from Eq. (14.15) that the zones broaden (i.e. standard deviations increase) by the square root of the distance migrated.

Combining Eqs. (14.8) and (14.15), we get

$$N = \frac{L^2}{\sigma^2} \quad (14.16)$$

⁸See, for example, *Theory and Problems of Statistics*, MR Spiegel, Schaum, New York, p. 122.

Now, if t_R is the time required to obtain the maximum of the solute peak (*retention time*) and t_σ is the standard deviation of the peak maximum in units of time (Fig. 14.18), we can write

$$\frac{\sigma}{L} = \frac{t_\sigma}{t_R} \quad (14.17)$$

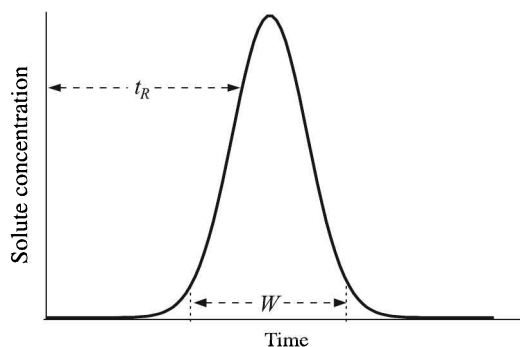


Fig. 14.18 Retention time and peak width.

We know from the properties of the Gaussian distribution that nearly 96% of the area under the curve lies within $\pm 2\sigma$. So, if W is the peak width at the baseline, in units of time

$$W = 4t_\sigma \quad (14.18)$$

Combining Eqs. (14.16), (14.17) and (14.18), we get

$$\begin{aligned} N &= \frac{L^2}{\sigma^2} = \frac{t_R^2}{t_\sigma^2} \\ &= 16 \left(\frac{t_R}{W} \right)^2 \end{aligned} \quad (14.19)$$

Equation (14.19) gives us a way to figure out the number of theoretical plates in terms of measurable quantities such as retention time and peak width. The following points emerge out of Eq. (14.19):

1. The more the number of plates, the less the peak width, i.e. the sharper the peak.
2. The higher the plate number, the more the retention time.
3. The number of theoretical plates for a column is not fixed. It varies from eluent to eluent as t_R and W are different for them.

Since it is difficult to measure accurately the beginning and end of a peak, it is a common practice to use the width at half height. Then,

$$N = 5.54 \left(\frac{t_R}{W_{1/2}} \right)^2 \quad (14.20)$$

Next, we would like to obtain an expression for the resolution factor R which is a measure of how two neighbouring peaks are completely separated from each other. To do that, we need to define a few terms.

Retention volume V_R . Suppose a solute is injected into the mobile phase. The detector at the end of the column will first detect the arrival of the mobile phase and after some time, that of the solute (Fig. 14.19).

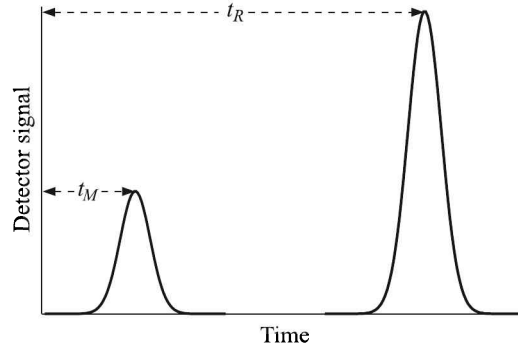


Fig. 14.19 Chromatogram showing arrival of the mobile phase and the solute.

The volume of the mobile phase required to convey the maximum of a solute band to the detector from the point of injection is called the *retention volume* V_R . Obviously,

$$V_R = t_R \cdot Q_a \quad (14.21)$$

where Q_a is the flow rate. If d is the inner diameter of the column and ϕ is the fraction of the column volume occupied by the stationary phase, then the flow rate is given by

$$Q_a = \frac{\pi d^2}{4} \cdot (1 - \phi) \cdot u_M \quad (14.22)$$

where u_M is the linear velocity of the mobile phase. Obviously,

$$u_M = \frac{L}{t_M}$$

The value of ϕ varies normally between 0.55 and 0.65.

Partition coefficient K . The analyte, on being conveyed to the chromatography column, distributes itself between the stationary and mobile phases. In other words, an analyte is in thermodynamic equilibrium between the two phases. The equilibrium constant K , called the *partition coefficient*, is defined as

$$K = \frac{C_S}{C_M} \quad (14.23)$$

where C_M is the concentration of the solute in the mobile phase within the column

C_S is the concentration of the solute in the stationary phase within the column.

Assuming a symmetrical peak, when the peak maximum arrives at the exit point of the column, the amount of solute eluted equals that remaining distributed between the stationary and mobile phases within the column. Thus,

$$V_R C_M = V_S C_S + V_M C_M \quad (14.24)$$

where V_M is the volume of the mobile phase within the column
 V_S is the volume of the stationary phase within the column.

From Eqs. (14.23) and (14.24), we get

$$V_R = V_M + KV_S \quad (14.25)$$

Equation (14.25) holds good for partition columns though for adsorption columns V_S needs to be replaced with the surface area of the adsorbent A_S .

Partition ratio k' . The partition ratio (also called the 'capacity factor') is defined as

$$\begin{aligned} k' &= \frac{\text{moles of the solute in the stationary phase within the column}}{\text{moles of the solute in the mobile phase within the column}} \\ &= \frac{C_S V_S}{C_M V_M} = K \frac{V_S}{V_M} \\ &= K\beta \end{aligned} \quad (14.26)$$

where β is called the *volumetric phase ratio*.

Now, we observe that if u_S is the average linear velocity of the solute

$$\begin{aligned} \frac{u_S}{u_M} &= \frac{\text{number of moles of the solute in the mobile phase within the column}}{\text{total number of moles of the solute within the column}} \\ &= \frac{C_M V_M}{C_M V_M + C_S V_S} \\ &= \frac{1}{1 + k'} \end{aligned}$$

or

$$\frac{L/t_R}{L/t_M} = \frac{1}{1 + k'}$$

or

$$t_R = (1 + k')t_M \quad (14.27)$$

So, the solute fraction in the mobile phase is given by

$$\frac{C_M V_M}{C_M V_M + C_S V_S} = \frac{1}{1 + k'}$$

and, that in the stationary phase is given by

$$\frac{C_S V_S}{C_M V_M + C_S V_S} = \frac{k'}{1 + k'}$$

These fractions also indicate the times the solute spends in the mobile and stationary phases.

In case there are two solutes 1 and 2 in the mobile phase and their adjusted retention times⁹ are t'_{R1} and t'_{R2} , the *selectivity factor* (or *relative retention factor*) α is defined as

$$\begin{aligned}\alpha &= \frac{t'_{R2}}{t'_{R1}} = \frac{t_{R2} - t_M}{t_{R1} - t_M} \\ &= \frac{k'_2}{k'_1} && \text{[From Eq. (14.27)]} \\ &= \frac{K_2}{K_1} && \text{[From Eq. (14.26)]}\end{aligned}$$

Resolution R . The resolution is a measure that indicates how two adjacent peaks are completely separated from each other. From the geometry of Fig. 14.20, it is defined as

$$\begin{aligned}R &= \frac{t_{R2} - t_{R1}}{\frac{W_1 + W_2}{2}} \\ &= \frac{\sqrt{N}}{2} \cdot \frac{k'_2 - k'_1}{2 + k'_2 + k'_1} && \text{[From Eqs. (14.19) and (14.27)]} \quad (14.28)\end{aligned}$$

Equation (14.28), however, is an approximated expression. The exact resolution equation¹⁰

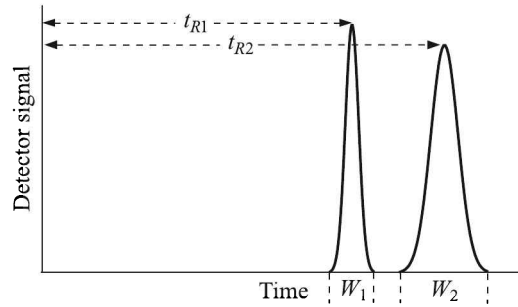


Fig. 14.20 Resolution of two adjacent peaks.

is

$$R = \frac{\sqrt{N}}{4} \cdot \frac{\alpha - 1}{\alpha} \cdot \frac{k'_2}{1 + k'_2} \quad (14.29)$$

The baseline resolution is achieved when $R = 1.5$. From Eq. (14.29) we observe that to obtain a high resolution, all the terms have to be maximised.

The first factor implies, N can be increased by increasing the length of the column [see Eq. (14.19)]. But that may lead to an increase in the retention time and band broadening. Alternatively, the HETP can be reduced by reducing the size of the stationary phase particles.

The selectivity factor can be manipulated to increase resolution. This can be done by suitably adjusting

⁹Adjusted retention time = (individual retention time – mobile phase retention time).

¹⁰Known as *Purnell equation*.

1. Mobile phase composition
2. Column temperature
3. Stationary phase composition

The partition ratio, k'_2 , can be altered by changing the temperature in GC and composition of the mobile phase in LC.

We consider a few examples to make ourselves familiar with the implications of the plate theory.

Example 14.1

A chromatography column with a length of 10.3 cm and inner diameter of 4.61 mm is packed with a stationary phase that occupies 61.0% of the volume. If the volumetric flow rate is 1.3 mL/min, find

- (a) the linear flow rate in cm/min
- (b) how long it takes for the solvent to pass through the column
- (c) the retention time for a solute with a capacity factor of 10.0

Solution

Given, $d = 4.61 \text{ mm} = 0.461 \text{ cm}$. $L = 10.3 \text{ cm}$. $Q_a = 1.3 \text{ mL/min} = 1.3 \text{ cm}^3/\text{min}$. $\alpha = 61.0\% = 0.61$. $k' = 10.0$

Therefore, from Eq. (14.22)

$$(a) \quad u_M = \frac{Q_a}{(\pi d^2/4) \cdot (1 - \alpha)} = \frac{4(1.3)}{\pi(0.461)^2(0.39)} = 17.36 \text{ cm/min}$$

$$(b) \quad t_M = \frac{L}{u_M} = \frac{10.3}{17.36} \text{ min} = 0.593 \text{ min.}$$

$$(c) \quad t_R = (1 + k')t_M = (11)(0.593) \text{ min} = 6.527 \text{ min.}$$

Example 14.2

Determine the partition ratio, the number of theoretical plates, and the HETP for the following analyses:

| Solute | t_R (min) | $W_{1/2}$ (min) |
|---------|-------------|-----------------|
| Air | 1.5 | |
| Benzene | 7.45 | 1.05 |
| Toluene | 10.6 | 1.45 |

Given, column length = 10 m and flow rate = 30 mL/min.

Solution

Separate calculations are given for the two solutes.

Benzene

$$t_R = (1 + k')t_M$$

\therefore

$$k' = \frac{t_R}{t_M} - 1 = \frac{7.45}{1.05} - 1 = 6.09$$

$$N = 5.54 \left(\frac{t_R}{W_{1/2}} \right)^2 = 5.54 \left(\frac{7.45}{1.05} \right)^2 = 279$$

$$H = \frac{L}{N} = \frac{1000}{279} = 3.58 \text{ cm}$$

Toluene

$$t_R = (1 + k')t_M$$

∴

$$k' = \frac{10.6}{1.5} - 1 = 6.07$$

$$N = 5.54 \left(\frac{10.6}{1.45} \right)^2 = 296$$

$$H = \frac{1000}{296} = 3.38$$

Note: Though the same column was used for the analysis, N and H are different for the two solutes.

Rate theory

The plate theory, though useful, assumes that equilibration between phases is infinitely fast. Also, it neglects solute diffusion and flow paths. A more realistic description of the state of affair is given by the van Deemter equation

$$\text{HETP} = A + \frac{B}{u} + Cu \quad (14.30)$$

where u is the average velocity of the mobile phase. A , B and C are constants which take care of eddy diffusion, longitudinal diffusion and resistance to mass transfer respectively.

Gas Chromatograph Set-up

A chromatograph, in its rudimentary form, is shown schematically in Fig. 14.21. It consists of the following essential components:

1. Carrier gas supply
2. Pressure regulator and flow monitor of the mobile phase
3. Sample injector
4. Chromatographic column consisting of a packing material
5. Temperature controlled chamber or oven
6. Detector
7. Recorder

A carrier gas—normally dry N_2 , CO_2 , Ar or He—is used to carry the sample gas through the column. The carrier gas must be chemically inert. The choice of carrier gas often depends upon the type of detector which is used. The carrier gas system also contains a molecular sieve to remove water and other impurities. A pressure regulator and flow monitor are incorporated

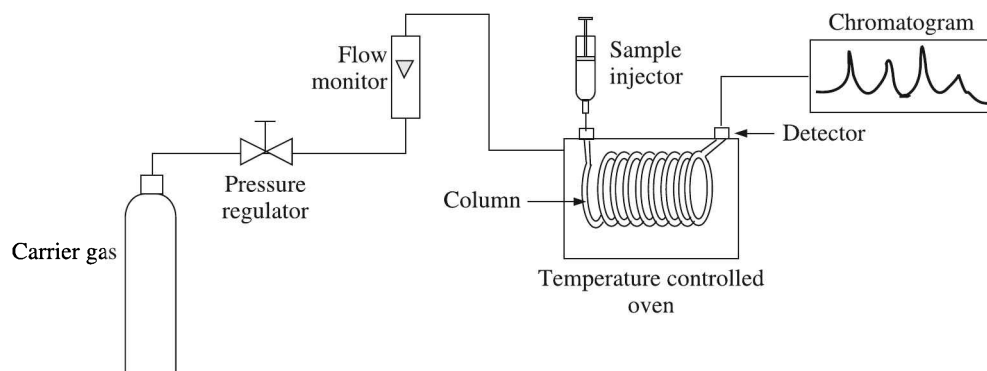


Fig. 14.21 Schematic diagram of a gas chromatograph.

in the supply line to maintain a steady flow of the carrier gas. The sample is injected by a syringe through a self-sealing silicone rubber diaphragm which constitutes the injection port. The sample, carried along with the carrier gas through the column, gets separated into components. The column is maintained at a constant temperature. Components, coming out of the column at different times (called *elution time*) are sensed by the detector that generates an electrical signal proportional to the concentration of the component. The detector signal, when fed to a potentiometric recorder, plots the chromatogram. The chromatogram is used to identify different components of the mixture and their concentrations.

If the sample is in form of gas, its injection by a syringe may cause problem. There, a bypass system, as shown in Fig. 14.22 may be used.

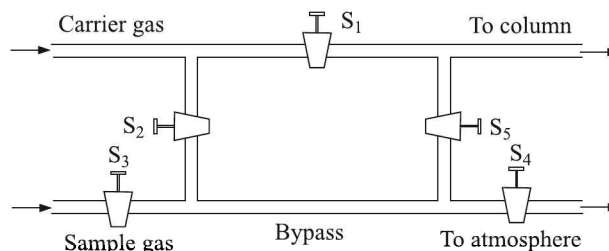


Fig. 14.22 Gas injection bypass system.

The bypass volume is initially flushed with the sample gas by opening S_3 and S_4 stopcocks. Next, S_3 , S_4 , S_1 are closed and S_2 , S_5 are opened when the carrier gas flows through the bypass to carry the sample to the column.

Columns

Columns are of two types

1. Packed
2. Capillary (aka *open tubular*)

Packed columns. Packed columns contain a finely divided, inert, solid support material (commonly based on diatomaceous earth¹¹) coated with liquid stationary phase. Most packed columns are 1.5 to 10 m in length and have an internal diameter of 2 to 4 mm.

Open tubular or capillary columns. Capillary columns have an internal diameter of a few tenths of a millimetre. They can be one of two types

1. Wall-coated open tubular (WCOT)
2. Support-coated open tubular (SCOT)

A WCOT consists of a capillary tube whose walls are coated with liquid stationary phase while in a SCOT, the inner wall of the capillary is lined with a thin layer of support material such as diatomaceous earth, onto which the stationary phase has been adsorbed. SCOT columns are generally less efficient than WCOTs. Both types of capillary column are more efficient than packed columns. In seventies, a new type of WCOT column, called the 'fused silica open tubular' (FSOT) column, was devised. It consists of a fused silica tube, the inner wall of which is chemically bonded with a stationary phase and the outer wall is given a polyimide coating. These columns are not only flexible enough to be wound into coils but also they have very low reactivity.

Detectors

Various detectors are available. Some of them are:

- Katharometer (or, thermal conductivity detector, TCD)
- Flame ionisation detector (FID)
- Electron capture detector (ECD)
- Cross-section ionisation detector
- Discharge ionisation detector

Katharometer¹² (or thermal conductivity detector, TCD)

The TCD is a widely used inexpensive and rugged detector. Thermal conductivities of gases differ from each other. This phenomenon is utilised in the TCD cell which is a resistor enclosed in an insulated container through which a gas may flow. The resistor gets heated as the current passes through it and it settles to a certain temperature by conducting heat through the ambient gas. Two such TCD cells are placed on the two arms of a Wheatstone bridge (Fig. 14.23).

The bridge remains balanced as long as the carrier gas flows through both TCD cells. The moment a different gas enters the sample cell, its resistance changes because of temperature change of the TCD cell caused by a gas of different thermal conductivity. This change in resistance throws the bridge out of balance and a voltage results across A and B which is monitored by a potentiometric recorder.

¹¹A fine, powdered siliceous earth, composed of skeletons of diatoms (minute, unicellular or colonial algae), used in industry as a filler, filtering agent, absorbent, clarifier and insulator. Also called *diatomite* or *kieselguhr*.— *Great Illustrated Dictionary*, The Reader's Digest Association Ltd, London (1985).

¹²Sometimes spelled as 'catherometer'.

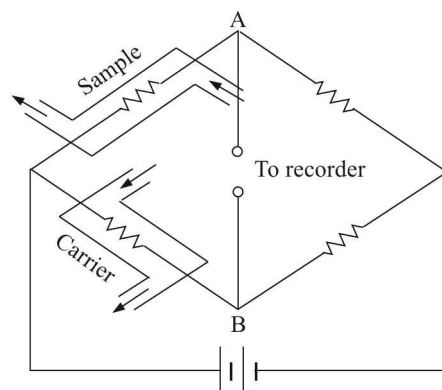


Fig. 14.23 TCD cells in Wheatstone bridge.

Flame ionisation detector (FID)

In an FID, hydrogen is mixed with the column effluent and burnt. An air supply supports combustion. The oxygen-rich hydrogen flame produces ionised fragments of organic molecules. Ions are collected by applying an electric field of about 300 V between the collector and the jet (Fig. 14.24).

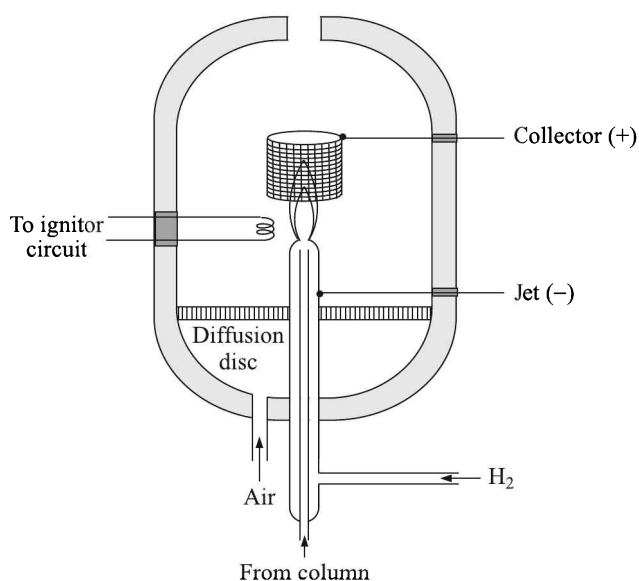


Fig. 14.24 Flame ionisation detector.

The current across the electrodes remains constant as long as the inert carrier gas passes the flame. However, if the vapour of a compound passes the flame, its molecules break into fragments by the hot flame. The ionised fragments, when collected by the collector, increase the current between the electrodes. The magnitude of the current is proportional to the carbon content of the organic molecules in the chromatograph column effluent. Actually, it is not the

carbon number, but effective carbon number (ECN) of a compound that determines the FID current. Table 14.5 lists the ECNs of a few compounds.

Table 14.5 ECN contributions of a few compounds

| <i>Atom</i> | <i>Coming from compound</i> | <i>ECN</i> |
|-------------|-----------------------------|------------|
| C | Aliphatic | 1.0 |
| | Aromatic | 1.0 |
| | Olefin | 0.95 |
| | Acetylene | 1.30 |
| O | Ether | -1.00 |
| | Primary alcohol | -0.60 |

The limitations of FID are:

1. It normally cannot detect noble gases, and H_2S , SO_2 , CO , CO_2 , NH_3 among others. However, some FIDs are provided with a methaniser which can convert CO , CO_2 to methane with the help of suitable catalysts. Then these gases can be detected.
2. Emerging components from the column get destroyed during detection.

Nevertheless, it is a very sensitive detector that possesses a high dynamic range ($\sim 10^7$) and very low (10^{-12} g/L) detection limits. These advantages have made it a very popular GC detector.

Electron capture detector (ECD)

Certain radioactive sources like tritium or nickel-63 emit electrons by β -decay. In this cell, one foil containing such radioactive source forms the cathode and the gas inlet pipe to the detector cell forms the anode (Fig. 14.25).

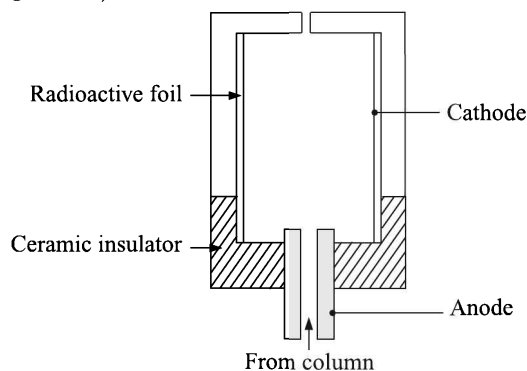


Fig. 14.25 Electron capture detector.

A low voltage, between 10 and 100 V, is applied between the cathode and anode. This produces a base current owing to the generation of low energy (thermal) electrons by the interaction of high energy β -electrons and the carrier gas atoms. Compounds eluting from the GC column have an affinity for thermal electrons and they capture them. This reduces the base current, thereby producing a chromatogram.

This detector is highly selective, having maximum response for halogenated compounds. Ne and He are the best carrier gases for this detector. However, this detector is not linear at higher concentrations.

Cross-section ionisation detector

Applied mostly for the analysis of mine air, cross-section ionisation detector is very useful. It is precise, robust, has a linear response over a wide range of concentration and somewhat insensitive to minor variation of the flow rate of the carrier gas.

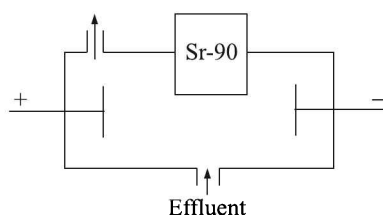


Fig. 14.26 Cross-section ionisation detector.

A radiation source, strontium-90, mounted in the detection unit, produces ion-pairs by interacting with organic compounds in the effluent gas. A rather high voltage, between 300 and 1000 V, applied between electrodes placed in the chamber, collects the ions (Fig. 14.26). The ionisation probability of an atom, called ionisation cross-section, depends on many factors, one of which is how fast the outer electrons are bound to the atom. Anyway, depending on their ionisation cross-section, the atoms produce response to this detector. Hence this name.

Hydrogen or helium is usually the carrier gas in this detection, which is non-destructive, though not very sensitive ($\sim 10^{-7}$ g/L).

Discharge ionisation detector

In contrast to the use of radioactive sources to produce ionisation of the eluting gas from the GC, the newer method is to generate a helium plasma cloud by a low-current arc which may be pulsed. The helium atom, while coming back to the ground state, gives off high energy photon that ionises most of the compounds or gases. Ions are collected by electrodes to produce a current like any other ionisation detector.

Thus, this detector not only detects organic compounds, but also CO, CO₂, N₂, O₂, H₂S, H₂, NO_x. It indeed is a universal, non-destructive, highly sensitive detector which has a wide dynamic range ($\sim 10^5$).

High Performance (Pressure) Liquid Chromatograph

For samples which come in liquid form and which cannot be vaporised for any reason, high performance liquid chromatograph (HPLC) comes in handy.

Identical in principle to gas chromatograph, chemical components of a liquid mixture are separated by the HPLC as the mixture is forced through a column by applying pressure of about 1000 psig (7 MPa). Instead of a carrier gas, here a liquid solvent is used. A schematic set-up is shown in Fig. 14.27.

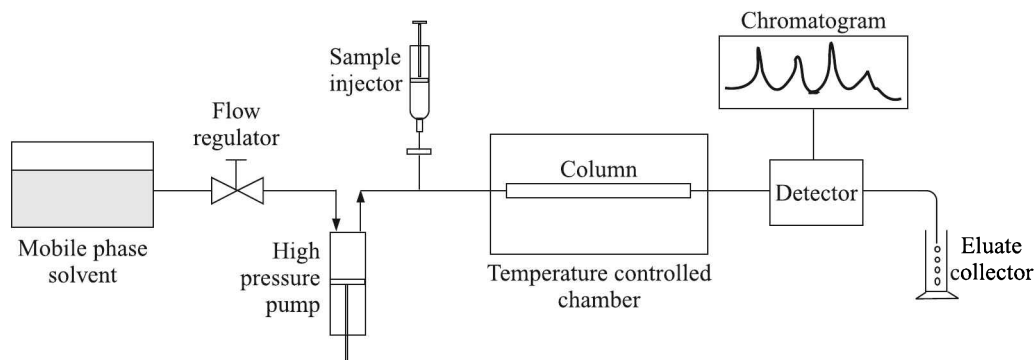


Fig. 14.27 HPLC set-up.

It is important to maintain an accurate solvent flow at a high pressure in HPLC. So, the pump constitutes an important part of the equipment. Columns, about 1 m in length and 2 mm in diameter, are packed with smaller particles in HPLC. Typical particle size is 20 μm .

UV-visible spectrophotometers or refractive index based detectors are used here.

14.3 Mass Spectrometer

A mass spectrometer analyses samples by producing their ions and separating them in the gas phase according to their mass-to-charge ratio (m/Ze). A mass spectrometric analysis is basically made up of the steps shown in Fig. 14.28.

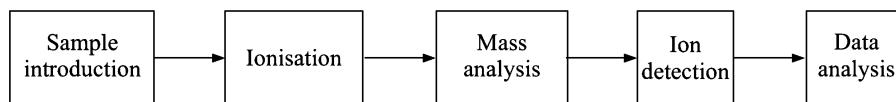


Fig. 14.28 Basic structure of mass spectrometry.

During the flight from the ion source to the detector in the analyser, mass spectrometers operate at vacuum level pressures to prevent collisions of ions with residual gas molecules. The vacuum should be such that the mean free path of an ion is longer than the distance from the source to the detector. For example, at a pressure of 5×10^{-5} Torr, the mean free path of an ion is approximately 1 m. Thus, it is necessary to go through a large pressure drop to introduce a sample into a mass spectrometer. There are several methods for doing that. Gas samples are directly connected to the instrument via a reservoir and controlled by a needle valve, the input to the instrument being metered by a pressure gauge. Liquid and solid samples are introduced through a septum inlet or a vacuum-lock system.

Though samples may be introduced in gas, liquid or solid states, in the latter two cases volatilisation must be done either prior to, or accompanying ionisation. Ranging from simple electron (impact) ionisation (EI) and chemical ionisation (CI) to a variety of desorption ionisation techniques with acronyms such as PD, FAB, ES and MALD, are available to produce charged molecules in the gas phase. Let us consider some of the common ionisation methods.

Ionisation Methods

Electron impact

Electron (impact) ionisation process involves the interaction of the gaseous sample with an electron beam generated by a heated filament in the ion source (Fig. 14.29).

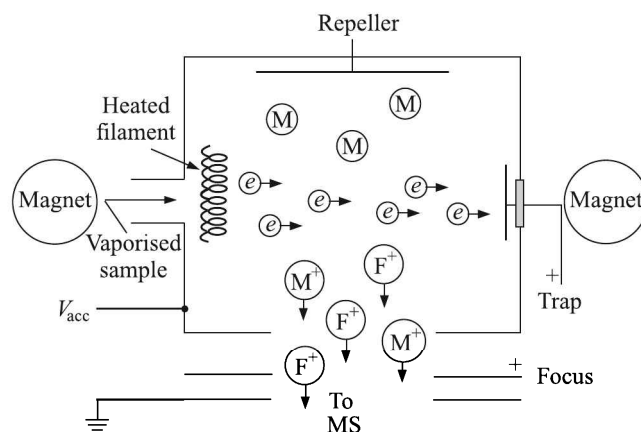
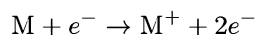


Fig. 14.29 Schematic diagram of electron (impact) ionisation. M: neutral particle, M^+ : molecular ion, F^+ : fragment ion, e^- : electron.

The electron energy, defined by the potential difference between the filament and the source housing, is usually set at 70 eV ($\sim 1.12 \times 10^{-17}$ J). The electron beam is kept focussed by a magnetic field across the ion source and is collected by a trap. Bombarded with a 70 eV electron, the gaseous molecule may lose one of its electrons to become a positively charged ion,



where M^+ indicates the molecular ion. Carrying an unpaired electron, it can occupy various excited electronic and vibrational states. If these excited states are sufficiently energetic, bonds will break and fragment ions (F^+) and neutral particles (M) will be formed. Extensive fragmentation of most of the molecules will occur when impacted by 70 eV electrons.

All ions are subsequently accelerated out of the ion source by an electric field produced by the potential difference applied between the ion source and a grounded electrode. The *repeller* serves to define the field within the ion source.

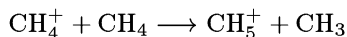
The 70 eV EI may affect the sensitivity of determination of masses in some cases because the signal arising from an analyte can be spread over many fragment ions. By choosing an electron energy close to the ionisation potential of the neutral molecule (typically 10 to 12 eV for simple organic molecules), the fragmentation can be reduced.

Chemical ionisation

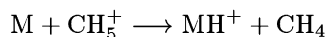
Chemical ionisation (CI) takes place when the molecule of interest is allowed to interact with a reactive ionised reagent species. Many such reactants are gaseous proton donors¹³. For

¹³aka *Bronsted acids*, because according to the theory of Bronsted, proton donors are acids and proton acceptors are bases.

example, EI from methane generates some of the most widely used reactant species. Initially, the ion CH_4^+ is formed. Then the following reaction produces the proton donor CH_5^+ :



If a neutral molecule M in the source has a higher proton affinity than CH_4 , the proton donor CH_5^+ will donate the proton H^+ to M to form the protonated species MH^+ in an exothermic reaction:



The construction of CI is very similar to that of EI except that it is more gastight so that it can retain the reactant gas at higher pressures in order to favour ion/molecule reactions. The pressure inside the CI ion source is typically $\sim 0.1\text{--}1$ Torr.

Plasma desorption ionisation

Plasma desorption¹⁴ (PD) technique uses ^{252}Cf fission fragments to desorb large molecules, like proteins, from a target. A droplet of the sample solution is applied to the target which is made of a thin aluminium foil and often covered with a layer of nitrocellulose. Proteins are adsorbed to nitrocellulose owing to hydrophobic interactions. Alternatively, the sample may be electrosprayed directly onto Ni or Al foil. The latter technique is more effective for smaller peptides.

Two atomic particles are produced by the ^{252}Cf fission reaction, one causing desorption of the analyte and the other providing the start signal for the time-of-flight measurement (Fig. 14.30).

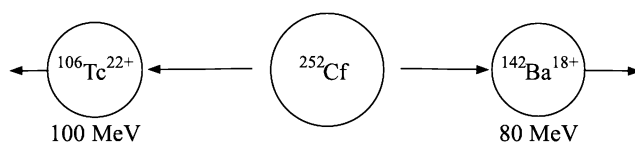


Fig. 14.30 A possible fission process used in PD ionisation.

A time-of-flight mass analyser, described later, is generally used for ion separation. The general arrangement of the instrument is shown in Fig. 14.31.

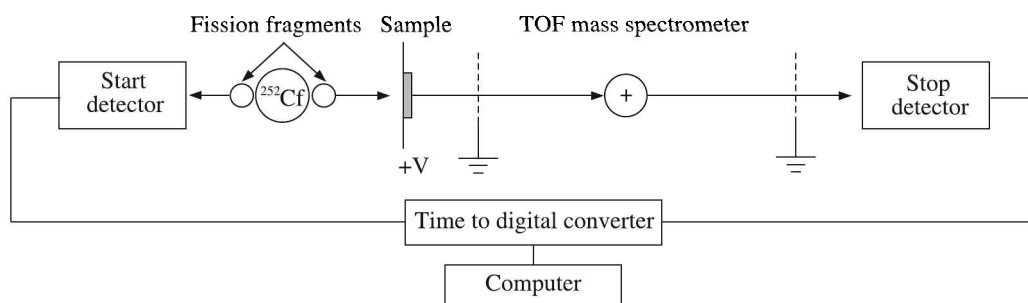


Fig. 14.31 Schematic arrangement for plasma desorption ionisation technique.

¹⁴Means 'changing from an adsorbed state on a surface to a gaseous or liquid state'.

Like other desorption techniques, PD suffers from

1. Cluster formation
2. Multiple charging and
3. Suppression effects, i.e. the inability to ionise some molecules due to the presence of other compounds present.

The technique is particularly suitable for peptides and small proteins and it is simple, that means it does not require to be an expert in mass spectrometry to interpret the results.

Fast atom bombardment ionisation

In fast atom bombardment (FAB) ionisation, a solid sample is bombarded with a high-energy (8–10 keV) beam of neutral atoms, typically Xe or Ar, causing desorption and ionisation. It is used for large biological molecules that are difficult to get into the gas phase. The atomic beam is produced by accelerating ions from an ion source through a charge-exchange cell. The ions pick up an electron in collisions with neutral atoms to form a beam of high energy atoms.

Ions (e.g. Cs^+) are used as the bombarding particle in a similar technique termed liquid secondary ion mass spectrometry (LSIMS). How the molecular ion is formed is depicted in Fig. 14.32. It involves several different mechanisms including ejection of preformed ions.

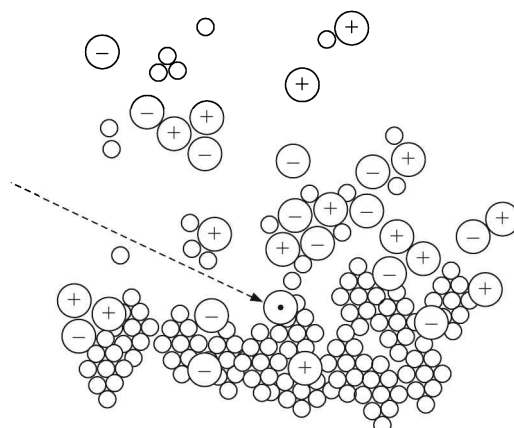


Fig. 14.32 Schematic presentation of FAB process. \oplus cation, \ominus anion, \odot Xe/ Cs^+ , \circ solvent. The arrow indicates the trajectory of the bombarding atom/ion.

FAB is a soft ionisation technique, which means it yields minimal fragmentation. It works well for polar and thermally labile compounds. In a typical FAB analysis, the sample is usually dissolved in an appropriate matrix—a viscous solvent, for example, glycerol—in order to keep the sample in the liquid state as it enters the high vacuum ion source. The matrix also reduces damage to the analyte caused by the high energy bombarding particle.

The conversion of liquid sample to gaseous ions on bombardment is believed to be caused by the sputtering process which basically redistributes momentum from the high energy particle by way of a cascade of collisions within the matrix. The formation of protonated molecular ions $(\text{M} + \text{H})^+$ or other cationised species such as $(\text{M} + \text{Na})^+$ in the positive ion mode, and $(\text{M} - \text{H})^-$ in the negative ion mode, are attributed to both gas phase reactions and solution

chemistry. Doubly-charged ion species and dimeric cluster ions of the analyte are occasionally observed.

The advantages and disadvantages of FAB are given below:

| <i>Advantages</i> | <i>Disadvantages</i> |
|--|---|
| 1. It is easy and fast to operate. | 1. It requires a high concentration of the organic liquid matrix (typically 80 to 95% glycerol) which lowers its sensitivity. |
| 2. The spectra are simple to interpret. | 2. Matrix cluster ions can, in some cases, dominate the mass spectrum. |
| 3. The source itself is easily retrofitted on most mass spectrometers. | 3. In some cases, the matrix also directly reacts with the analyte, forming radical anions or causing reduction of the analyte. |
| | 4. The desorption process also produces a great many sputtered neutral molecules in addition to ions. |

In continuous-flow FAB (CF-FAB), a probe delivers the sample solution to the target at flow rates of up to $\sim 10 \mu\text{L}/\text{min}$. This reduction of the organic matrix results in an increase in the signal-to-noise ratio due to lower chemical background. A major advantage of CF-FAB is its usefulness for (i) flow-injection analysis, (ii) on-line reaction monitoring, and (iii) coupling to HPLC. Now FAB is widely used on quadrupole and sector instruments.

Matrix assisted laser desorption (MALD) ionisation

Based on laser ionisation, MALD is a method of vaporising and ionising large biological molecules such as proteins or DNA fragments. The biological molecules are dispersed in a solid matrix such as nicotinic acid. A UV-laser pulse ablates the matrix which absorbs energy at the wavelength of the laser. Along with the matrix, the large molecules are also ablated and desorbed into the gas phase in an ionised form so that they can be extracted into a mass spectrometer.

The mechanism of how a combination of desorption and ionisation occurs in MALD is still under investigation. One model proposes that the laser energy absorbed by the matrix, typically $\sim 10^6 \text{ watts}/\text{cm}^2$, leads to intense heating and generation of a plume of ejected material that rapidly expands and undergoes cooling. The generation of ions is believed to result from ion/molecule reactions in the gas phase. Generally, the $[\text{M} + \text{H}]^+$ ion, or $[\text{M} + \text{Na}]^+$, $[\text{M} + \text{K}]^+$ etc., are preferentially formed in the positive ion mode, and $[\text{M} - \text{H}]^-$ ion in the negative ion mode. However, the technique also generates disturbing low intensity singly- and multiply-charged clusters of the analyte that tend to complicate the spectrum. Mass resolution is the highest when the used laser power is close to the threshold level required to produce ions from the solid sample.

Depending on the analyser, MALD can be used to determine the molecular weight of molecules up to 500 kDa¹⁵, routinely 5 to 100 kDa of polymers, biomolecules, complexes, enzymes.

¹⁵ 1 Da (dalton) = 1 u (unified atomic mass unit, see Appendix F, Table F.7). In biochemistry and molecular biology literature (particularly in reference to proteins), the *dalton* unit is used, with the symbol Da. Because proteins are large molecules, they are typically referred to in kilodaltons, or *kDa*.

The MALD technique is most commonly coupled with a TOF analyser which offers rather low resolution and accuracy¹⁶. But the combination is popular because it is easy to handle. To get very accurate data, the MALD can be coupled to a Fourier-transform mass spectrometer though it is expensive, difficult to handle and has a low dynamical range.

Electrospray (ES) ionisation

The electrospray source consists of a fine capillary tube through which the sample solution is sprayed into a strong electric field in the presence of a flow of warm nitrogen to assist desolvation. The droplets carry charge when they exit the capillary and as the solvent vaporises, the droplets disappear leaving highly charged analyte molecules (Fig. 14.33).

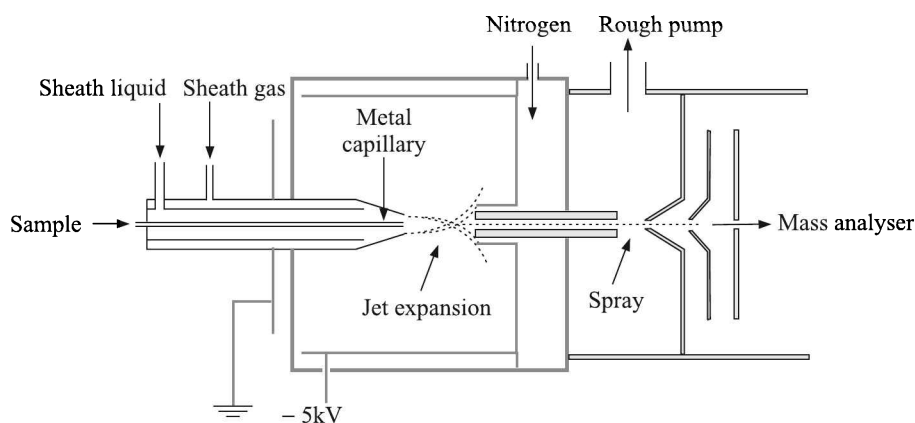


Fig. 14.33 Schematic presentation of the ES ionisation.

One mechanism suggests that ion formation results from an ion evaporation process. The electrostatic dispersion of the liquid ejected from the capillary tip produces a spray of droplets. Aided by the heated bath gas (usually nitrogen), the droplets undergo declustering, losing solvent molecules in the process and eventually producing individual ions.

Another mechanism suggests that owing to the desolvation, the droplets become smaller in size. This increases the charge density on the droplet surface that eventually causes a coulombic explosion producing individual ions.

Whatever the mechanism, ions are formed at atmospheric pressure and enter a cone shaped orifice, which acts as a first vacuum stage where they undergo a free jet expansion. A skimmer then samples the ions and guides them to the mass spectrometer.

ES ionisation is the method of choice for proteins, oligonucleotides and metal complexes. However, the sample must be soluble in low boiling solvents such as acetonitrile, methyl alcohol, methyl chloride, water, etc. and stable at very low concentrations ($\sim 10^{-2}$ mol/L).

Another atmospheric pressure ionisation technique, termed ion spray or atmospheric pressure chemical ionisation (APCI), basically works in a manner similar to ES. The ion source is similar to the ES ion source. In addition to the electrohydrodynamic spraying process, a plasma is created by a corona-discharge needle at the end of the metal capillary. In this plasma proton transfer reactions and, to a small amount, fragmentation can occur.

¹⁶A mass accuracy of $\pm 0.01\%$ (± 1 Da at a molecular mass of 10,000) is the best that can be achieved under favourable conditions with today's commercially available instruments.

Although both techniques give qualitatively the same results, ES and APCI appear to have their unique advantages in specific applications. APCI, for example, has been used more successfully to study non-covalent interactions, probably attributed to its ease of use with 100% aqueous solutions.

Thermospray ionisation

Thermospray ionisation is used for the coupling of HPLC at conventional flow rates (0.5 – 1.5 ml/min) to a mass spectrometer. The effluent from the HPLC column is fed to a hot chamber through a heated stainless steel tube of 0.10 – 0.15 mm inner diameter where it vaporises under reduced pressure. As a result a high velocity jet comprising small droplets is produced. The droplets vaporise further due to the hot gas in this low pressure region of the ion source (Fig. 14.34).

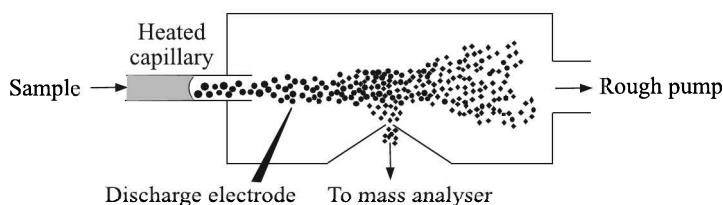


Fig. 14.34 Schematic presentation of the thermospray arrangement.

Polar or charged species and volatile buffers are required for ionisation. For semi-volatile samples a filament arrangement is used while for highly aqueous effluents a discharge device is used. The temperature of the vaporiser requires to be adjusted for a given solvent composition. Ions are drawn into the mass analyser by electric fields where they enter through an orifice of about 0.5 mm diameter. Considered as a soft ionisation technique, thermospray produces only limited fragmentation of the analyte.

Mass Analysers

Mass analysers can be divided into three categories, namely

1. Scanning mass analysers
2. Time of flight (TOF) mass analysers
3. Trapped ion mass analysers

Scanning mass analysers

A scanning mass analysis is analogous to optical spectroscopy where one starts with visible light which is composed of different wavelengths of light that are present at different intensities. A dispersive device, such as prism, breaks the light into its different wavelengths, and a suitably placed slit determines which wavelength will reach the detector. The different wavelengths are then swept (or, scanned) across the detector slit and their light intensities are recorded as a function of wavelength.

In scanning mass analysis, one starts with a mixture of ions having different mass-to-charge ratios (m/Z_e) and different relative abundances. An electromagnetic field disperses the ions according to their m/Z_e ratios, and a slit selects which m/Z_e will reach the detector. The

different m/Ze ratios are then scanned across the detector slit and the ion current is recorded as a function of mass.

Scanning mass analysers are of two types

1. Magnetic sector type
2. Quadrupole type

Magnetic sector-type. Consider a mixture of ions of different masses, all possessing the same velocity v , leaving the container A vertically upwards (Fig. 14.35). Suppose, they were all accelerated by an electrostatic field V_a .

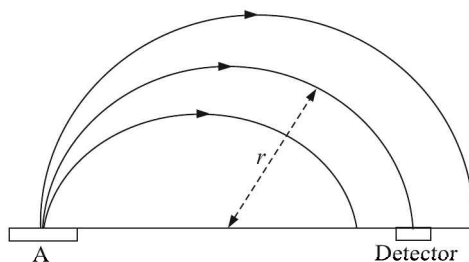


Fig. 14.35 Deflection of ions in a magnetic field.

Equating the KE and PE of one ion of mass m and charge Ze we get

$$\frac{1}{2}mv^2 = ZeV_a \quad (14.31)$$

$$\Rightarrow v = \sqrt{\frac{2ZeV_a}{m}} \quad (14.32)$$

Now, if a magnetic field of intensity B is applied at right angles to the direction of motion of the ion, i.e., perpendicular to this page, it will experience a tangential force $BZev$ which will make the ion describe a circle of radius r such that the centripetal force will be balanced by the centrifugal force as

$$\frac{mv^2}{r} = BZev$$

$$\Rightarrow r = \frac{mv}{BZe} \quad (14.33)$$

Substituting the value of v from Eq. (14.32) in Eq. (14.33), we get

$$r = \frac{m}{BZe} \sqrt{\frac{2ZeV_a}{m}}$$

$$\text{or } \frac{m}{Ze} = \frac{B^2 r^2}{2V_a} \quad (14.34)$$

We observe from Eq. (14.34) that for a fixed B and V_a , an ion of given m/Ze ratio will describe a circular path of a particular r . By suitably changing either B or V_a or both,

another ion of ratio m/Ze may be made to describe the path of same radius. This principle is utilised in the construction of sector-type mass analyser which was first fabricated by Aston at Cambridge University, UK in 1920.

In fact, a magnetic sector alone will separate ions according to their m/Ze ratio [Fig. 14.36 (a)].

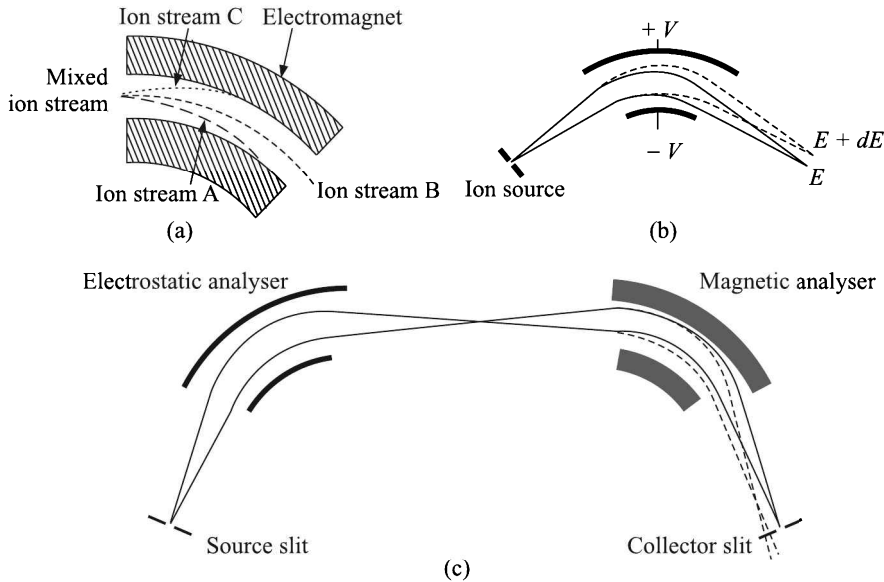


Fig. 14.36 (a) Magnetic sector type, (b) Electrostatic sector type, and (c) Double-focussing mass analysers.

However, ions leaving the ion source do not all have exactly the same energy and therefore do not have exactly the same velocity. This will limit the resolution. To achieve better resolution, it is necessary to add an electrostatic sector that focuses ions according to their kinetic energy. Like the magnetic sector, the electrostatic sector analyser (ESA) applies a force perpendicular to the direction of ion motion, and therefore has the form of an arc. The ESA results in overall directional focussing, but ions of different energy will still have different foci [Fig. 14.36 (b)]. However, if the ESA is so designed that the dispersion of the ions due to their velocity spread is exactly equal and opposite to that of the magnetic sector, the result of the combination is zero net velocity dispersion, i.e. ions of the same m/Ze but different velocities are focussed at the same point. The equation for the ESA radius can be derived by equating the electrical force ZeV to the centripetal force:

$$ZeV = \frac{mv^2}{r}$$

or

$$r = \frac{mv^2}{ZeV} = 2 \frac{V_a}{V} \quad [\text{by applying Eq. (14.31)}]$$

A combination of an ESA and magnetic sector [Fig. 14.36(c)], then, focuses both direction and energy and is called *double-focussing* for that reason. Such an apparatus is capable of a mass resolving power exceeding 100,000.

The resolving power R is defined as

$$R = \frac{m}{\Delta m} \quad (14.35)$$

where Δm is the mass difference between two neighbouring masses, m and $(m + \Delta m)$, of equal intensity with signal overlap of 10%. A resolving power of 100,000 means, one can clearly distinguish between ions of mass 100.000 Da and 100.001 Da, or 100000 Da and 100001 Da, which corresponds to 10 ppm. Such accurate mass measurement at low mass can help one determine the empirical chemical formula of an unknown ion, by determining the compatible combinations of carbon, hydrogen, nitrogen and other atoms at the measured exact mass within the experimental uncertainty. A schematic diagram of a double-focussing mass spectrometer which uses the Mattauch-Herzog geometry, is given in Fig. 14.37.

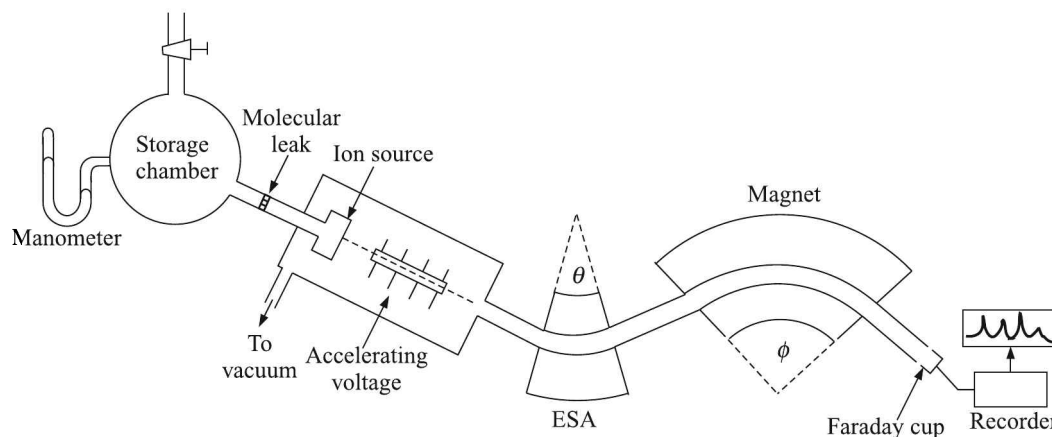


Fig. 14.37 Double-focussing mass spectrometer of Mattauch-Herzog geometry. The electric sector angle $\theta = \pi/4\sqrt{2}$, and the magnetic sector angle $\phi = \pi/2$.

Sample in gaseous form is transferred to a glass bulb attached to a mercury manometer at a pressure of about 50 mm of Hg. The sample is drawn at a steady rate through a molecular leak to the ionising chamber by the high vacuum ($\sim 10^{-6}$ Torr) maintained there. The sample may be ionised by any of the discussed ionisation methods except those which produce pulses.

Ion beam is then electrostatically collimated and accelerated by a series of slits kept at higher and higher potentials. Thereafter, the ions are deflected through a tandem arrangement of an electrostatic analyser and then a magnetic analyser. This arrangement focuses the ions of the same m/Ze ratio, but having different initial velocities and directions, on to the Faraday cup collector. The generated ion current is amplified and fed to a recorder. By sweeping the magnetic field through a variation of current of the electromagnet, the mass spectrum of the sample can be generated.

The electric sector is usually held constant at a value which passes only ions having the specific kinetic energy. Therefore, the most commonly varied parameter is B , the magnetic field strength. The magnetic field is usually scanned exponentially or linearly to obtain the mass spectrum. A magnetic field scan can be used to cover a wide range of m/Ze ratios with a sensitivity that is essentially independent of the m/Ze ratio.

Alternatively, B may be held constant and V_a scanned. The electric sector potential tracks the accelerating voltage. Since the electric field does not suffer from hysteresis, the relationship

between m/Z_e ratio and accelerating voltage remains linear. The disadvantage of V_a -scan is that the sensitivity is roughly proportional to the m/Z_e ratio.

Double-focussing magnetic sector mass analysers are the “classical” models against which other mass analysers are usually compared. Their advantages and disadvantages are:

| <i>Advantages</i> | <i>Disadvantages</i> |
|--------------------------------------|---|
| 1. Very high reproducibility | 1. Big machines, costlier than other mass analysers |
| 2. High dynamic range and resolution | 2. Not suitable for pulsed ionisation input like MALD |
| 3. High sensitivity | |

Quadrupole-type. Quadrupole mass analysers consist of an ion source, a quadrupole mass filter and an electric lens system to focus ion in the quadrupole filter. A schematic diagram of the machine is shown in Fig. 14.38.

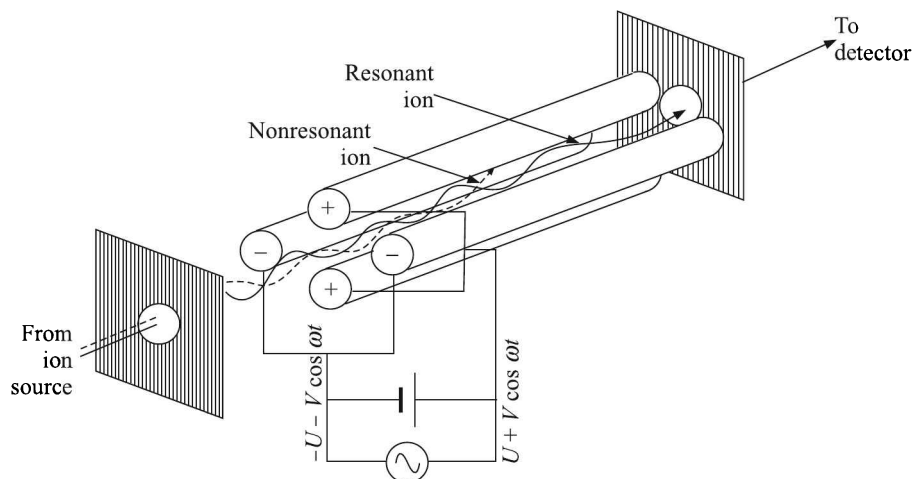


Fig. 14.38 Quadrupole mass analyser.

The quadrupole m/Z_e filter consists of four precision crafted cylindrical rods arranged in an orthogonal array. Opposite rod pairs are shorted and each of such shorted pairs acts as an electrode. The positive pair is given a positive dc bias along with an RF voltage, Similarly, the negative pair is given a negative dc bias along with an RF voltage that is 180° out of phase with the positive pair. The varying electromagnetic field creates a condition that separates ions of different m/Z_e ratio, The separation is shown in diagram form in Fig. 14.39.

The potential ϕ_0 applied to opposite pairs of rods is given by

$$\pm\phi_0 = U + V \cos \omega t$$

where U is a dc voltage and $V \cos \omega t$, the time-dependent RF voltage in which V is the amplitude and ω , the radio frequency.

At given values of U , V and ω , only certain ions will have stable trajectories through the quadrupole. The range of ions of different m/Z_e values, capable of passing through the mass filter, depends on the ratio of U/V . All other ions will have unstable trajectories, that means

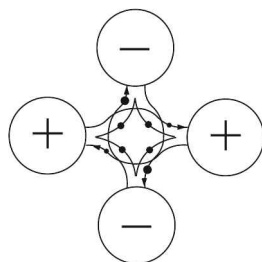


Fig. 14.39 Quadrupole filter action. Smaller masses (small dots) get collected at the anode, bigger ones (big dots) at the cathode and tuned masses (intermediate-size dots) pass through.

they will have large amplitudes in x - or y -direction and will be lost. The equation of motion for a singly-charged particle can be expressed as a Mathieu equation from which one can define expressions for the Mathieu parameters a_u and q_u as,

$$a_u = a_x = -a_y = \frac{4eU}{m\omega^2 r_0^2}$$

$$q_u = q_x = -q_y = \frac{2eV}{m\omega^2 r_0^2}$$

where (m/e) is the mass-to-charge ratio of the ion

r_0 is the half the distance between two opposite rods

There is no z parameter, because the ac field acts only in the xy -plane, z being along the axis of the linear quadrupole. Scanning the mass range on a quadrupole means changing the values U and V at a constant ratio $a_u/q_u = 2U/V$, while keeping the radio frequency ω fixed.

The operation of a quadrupole mass analyser is usually treated in terms of a stability diagram (Fig. 14.40) that relates the applied dc potential U , the RF potential $V(t)$ and the radio frequency ω to a stable vs. unstable ion trajectory through the quadrupole rods.

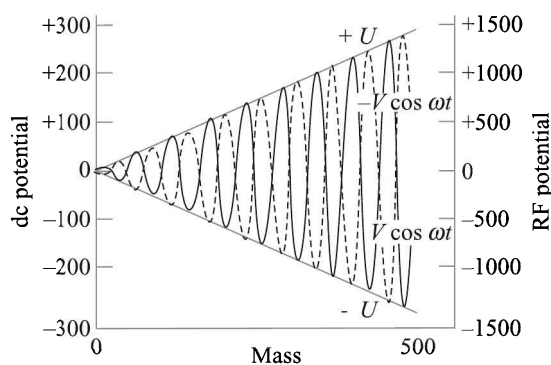


Fig. 14.40 Stability diagram for a quadrupole mass analyser.

If the slope of the scan line is altered, the resolution will change because the number of cycles experienced by an ion within the RF field, which in turn depends on its velocity, depends on this slope. Thus, the resolution will increase with increasing mass, as ions of higher mass

have lower velocity. However, then the number of ions that will reach the detector, i.e. the transmission efficiency, will decrease, because ions of higher masses spend a longer time in the quadrupole. Good quality machining of the quadrupole rods also enhances the resolution.

Apart from their use as a mass filter, quadrupole rods can be used for other purposes as well. A quadrupole, biased with only RF can be used as an ion guide over a wide range of mass while one with the dc bias only acts as an electrostatic lens for ions.

A major advantage of a quadrupole mass filter over a sector instrument is that a low voltage requires to be applied to the ion source so that the KE of the ions is $\sim 5\text{--}10\text{ eV}$, as compared to several keV necessary for a sector instrument. This eliminates high voltage problems and makes interfacing of the mass spectrometer to gas chromatograph (GCMS) and liquid chromatograph (LCMS) easier. Other advantages are its good transmission efficiency, high scan speed, and wide acceptance angle to give high sensitivity.

Save its rather complex electronics, the quadrupole mass spectrometer is lightweight, does not require a stable magnetic field and therefore is not so costly. But, its resolution is not very high and it is also not suitable for use with pulsed ionisation sources.

Time-of-flight (TOF) mass analysers

We have seen from Eq. (14.32) that when subjected to an electrostatic potential V_a , the velocity acquired by an ion is given by

$$v = \sqrt{\frac{2ZeV_a}{m}}$$

If this ion is collected at the end of a tube of length l , the time t required to reach there is

$$\frac{l}{t} = \sqrt{\frac{2ZeV_a}{m}} \quad (14.36)$$

$$\Rightarrow t = l\sqrt{\frac{m}{2ZeV_a}} \quad (14.37)$$

Equation (14.37) shows that ions of different m/Z ratio will arrive at different times at the end of the tube, high mass ions taking longer to reach the detector than low mass ions. This principle is utilised to construct time-of-flight mass analysers, a schematic diagram of which is presented in Fig. 14.41.

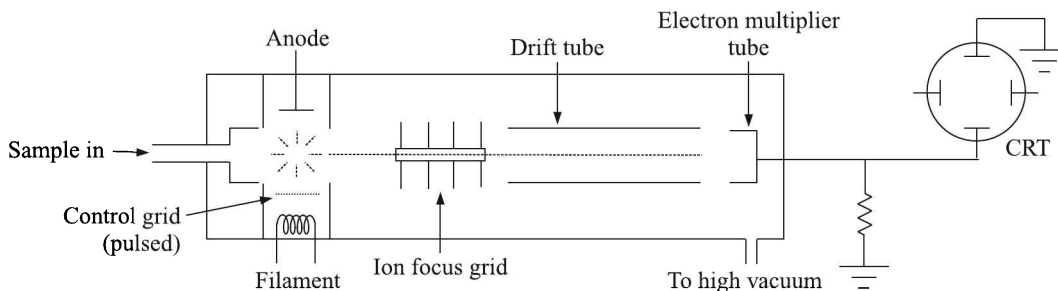


Fig. 14.41 Time-of-flight mass analyser.

An equation relating the flight time of an ion with its m/Ze value can also be derived as

$$t = a\sqrt{\frac{m}{Ze}} + b$$

where a and b are constants for a given set of instrument conditions, and are determined experimentally from flight times of ions of known masses.

In the TOF mass spectrometer shown in Fig. 14.41, the sample is ionised by pulsed electron bombardment having 20,000 to 35,000 pulses/s, focussed and accelerated by a grid of electrodes and then fed to the drift tube. Ions of different m/Ze ratios separate in this tube. Upon exiting, the ions are detected by an electron multiplier tube, the output of which is fed to the vertical deflection plates of a CRT while the horizontal sync is the same as the pulse applied to generate ions.

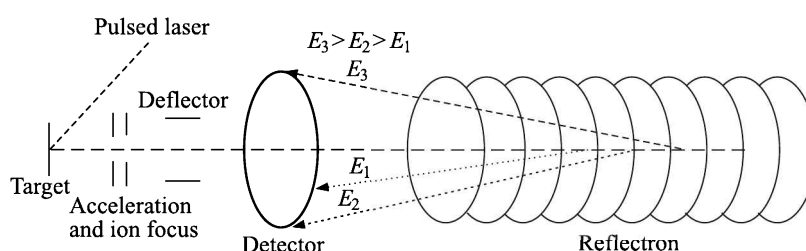


Fig. 14.42 A linear-field reflectron.

Several ionisation techniques are suitable for TOF mass analysers by which ions are generated or ejected from the ion source over very short periods of time. This can be achieved with a laser pulse, ^{252}Cf fission fragments, and introduction of ions from continuous ionisation sources (EI, ES, FAB, etc.) with pulsed deflection or pulsed extraction. This pulse may also give the start signal for the data acquisition.

It follows from Eq. 14.37 that m/Ze is proportional to t^2 , which yields the formula for the resolving power as

$$\frac{m}{\Delta m} = \frac{1}{2} \frac{t}{\Delta t}$$

Linear TOF MS instruments are capable of attaining a resolution of 1 part per 1000. A factor affecting instrument resolving power is the time resolution of the detector and electronic circuitry, which now can handle a Δt of a few nanoseconds. A major limitation in achieving higher resolution is the consequence of the spread in time, space and kinetic energy of the initial ion packet. The ions leaving the ion source of a TOF mass analyser have neither exactly the same starting times nor exactly the same kinetic energies. Various TOF mass analyser designs have been developed to compensate for these differences. A *reflectron* is an ion optic device in which ions in a TOF mass analyser pass through a “mirror” and their flights are reversed.

Reflectron. A linear-field reflectron (Fig. 14.42) allows ions with greater kinetic energies to penetrate deeper into the reflectron (an electrostatic repeller field aka ion mirror) than ions with smaller kinetic energies. The ions that penetrate deeper will take longer to return to the detector. If a packet of ions of a given m/Ze ratio contains ions with varying kinetic energies,

then the reflectron will decrease the spread in the ion flight times and therefore, improve the resolution of the TOF mass analyser.

A curved-field reflectron ensures that the ideal detector position for the TOF mass analyser does not vary with m/Z ratio. This also results in improved resolution for TOF mass analysers.

Reflectron TOF analysers are capable of having a resolving power of over 1 part per 10,000.

The TOF mass spectrometer records the entire mass spectrum at one cycle of the pulse whereas a conventional mass spectrometer detects one mass at a time. Secondly, it does not require the critical mechanical alignment as well as a stable magnetic field. Its electronic circuitry for time determination, however, is a bit complicated.

Trapped-ion mass analysers

Trapped-ion mass analysers can be divided into two types

1. Dynamic
2. Static

Quadrupole ion trap mass analysers belong to the former type while ion cyclotron resonance and orbitrap mass analysers to the latter. Both store ions in the trap and manipulate them by using dc and RF electric fields in a series of carefully timed events. This enables trapped ion mass analysers to achieve high resolution and high sensitivity through MS/MS coupling.

Quadrupole ion trap. Based on the same principle as the quadrupole mass filter, the quadrupole field in the quadrupole ion trap (QIT) is generated within a three-dimensional trap which consists of a ring electrode and two end caps as shown in Fig. 14.43(b).

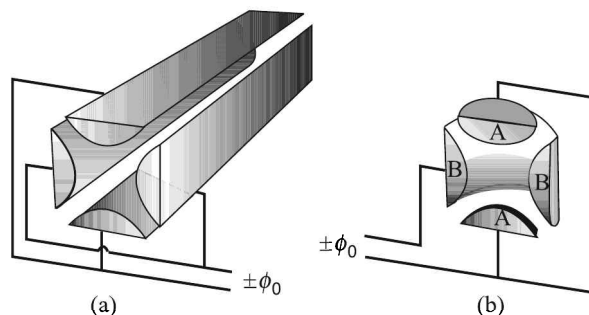


Fig. 14.43 Schematic diagrams of the set up of a quadrupole ion trap: (a) linear form, and (b) ring electrode form.

The QIT has two configurations: the three dimensional form mentioned above and the linear form made of four parallel electrodes [Fig. 14.43(a)]. The advantage of these designs is in their simplicity, though it is not easy to visualise how they work. The motions of a single ion in the trap are shown schematically in Fig. 14.44. These motions are described by the Mathieu equations which can only be solved numerically and the results can be displayed by computer simulations.

The 3D trap generally consists of two hyperbolic metal electrodes and a hyperbolic ring electrode halfway between the other two electrodes. The ions are trapped in the space between

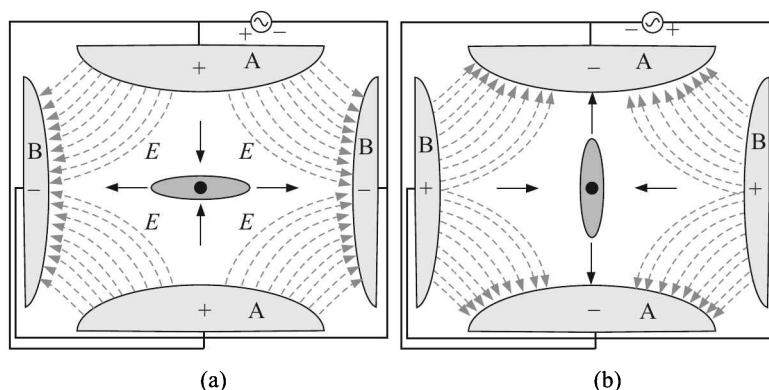


Fig. 14.44 Diagrams (a) and (b) show two states during an ac cycle. Black circle indicates a particle of positive charge, surrounded by a cloud (grey) of similarly charged particles. The electric field E (broken lines) is generated by a quadrupole of endcaps (A) and a ring electrode (B).

these three electrodes by ac and dc electric fields. The radio frequency ac voltage oscillates between the two hyperbolic metal end cap electrodes if ion excitation is desired and the driving ac voltage is applied to the ring electrode. The ions are first pulled up and down axially while being pushed in radially. Then they are pulled out radially and pushed in axially (from the top and bottom). In this way the ions move in a complex motion that generally involves the cloud of ions being long and narrow and then short and wide, back and forth, oscillating between the two states.

As is the case with quadrupole mass filters, the quadrupole field is closest to the theoretical ideal in the centre of the trap. For this reason, in addition to the sample, a moderator gas like helium is often introduced into the trap to dampen the oscillations of the ions and hence concentrate them in the centre of the trap.

There are many methods of mass/charge separation and isolation from the QIT. The most commonly used method is the mass instability mode in which the RF potential is ramped so that the orbit of ions with a mass greater than a certain mass m are stable while ions with mass m become unstable and are ejected onto a detector.

Ions may also be ejected by the resonance excitation method. Then a supplementary oscillatory excitation voltage is applied to the endcap electrodes and the trapping voltage amplitude and/or excitation voltage frequency is varied to bring ions into a resonance condition in order of their mass/charge ratio.

The cylindrical ion trap is a derivative of the quadrupole ion trap.

In the original design¹⁷, the potential ϕ_0 was applied to the ring electrode and $-\phi_0$ to the end caps. With this arrangement, ions were detected by resonance techniques.

As the word 'trap' implies, the QIT can store ions over a long period of time making it possible to study gas phase reactions. Also, the ion trap has excellent MS/MS capabilities. The mass range of commercial instruments is 650 kDa, scanned at over 5 kDa/s. By reducing the scan speed to 0.015 Da/s, a resolving power of 1.2×10^7 FWHM¹⁸ can be achieved.

¹⁷The 3D quadrupole ion trap was invented by Wolfgang Paul who shared the Nobel Prize in Physics in 1989 for this work. For this reason, often it is referred to as Paul trap.

¹⁸Full width at half maximum.

The very high sensitivity of the ion trap is because of its ability to detect all ions that are formed. However, space charge effects reduce the accuracy of mass assignment. Even though ion/molecule reactions take place within the trap, the EI spectra generally compare well with that acquired on quadrupole mass filters. The applications potential of the ion trap is very great because of its size, speed, sensitivity, MS/MS capabilities and compatibility with most ionisation techniques.

Ion cyclotron resonance. The ion cyclotron resonance (ICR) mass analyser consists of three pairs of parallel plates arranged in the form of a cube. These pairs are used for trapping, excitation and detection of ions as shown in Fig. 14.45. The cell is placed in a strong magnetic field \mathbf{B} which is perpendicular to the trapping plates.

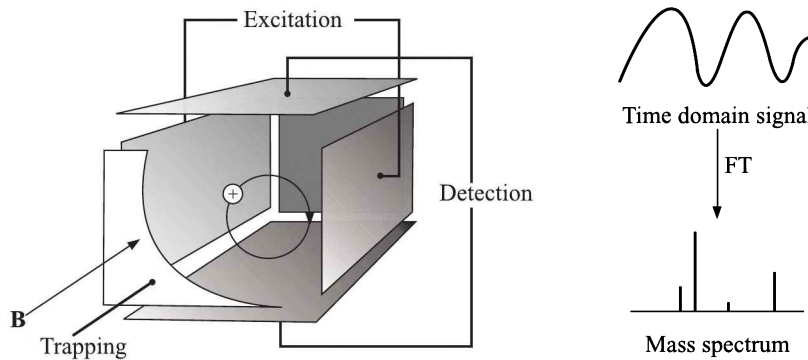


Fig. 14.45 Schematic diagram of ion cyclotron resonance analyser.

A trapped ion, having an initial velocity v will rotate in a circular path of radius r given by [see Eq. (14.33)]

$$r = \frac{mv}{ZeB}$$

The frequency of rotation ω_c , called the *cyclotron frequency*, is then given by

$$\omega_c = \frac{v}{r} = \frac{ZeB}{m} \quad (14.38)$$

Equation (14.38) is only an approximate relation because actually a quadrupolar electrical field is used to trap the ions in the axial direction. As a result of the axial electrical trapping, an axial oscillation of frequency ω_t , given by Eq. (14.3), within the trap is produced.

$$\omega_t = \sqrt{\frac{Ze\alpha}{m}}$$

where α , a constant akin to the spring constant of a harmonic oscillator, varies with the applied voltage and trap dimensions. The cyclotron frequency gets reduced owing to the electric field and the resulting axial harmonic oscillation, and a second radial motion, called the *magnetron motion* that occurs at the magnetron frequency, is introduced. The expressions for modified cyclotron and magnetron frequencies are given by

$$\omega_{\pm} = \frac{\omega_c}{2} \pm \sqrt{\left(\frac{\omega_c}{2}\right)^2 - \left(\frac{\omega_t}{2}\right)^2}$$

where ω_+ is the reduced cyclotron frequency, and ω_- is the magnetron frequency. However, ω_+ is what is measured in the ICR.

The ions remain trapped within the cell owing to the application of a dc voltage to the trapping plates. They can remain trapped for hours, provided the ambient pressure is $< 10^{-8}$ Torr. We note from Eq. (14.38) that ω is inversely proportional to the mass of an ion. Because of the initial spatial distribution of ions, an excitation pulse is applied prior to detection, so that all ions of the same ω absorb energy and move together coherently. This then induces a current in the detector plates (image current) that is proportional to the numbers of ions and m/Ze values of the excited ions. Since the frequency of an ion's cycling is determined by its mass to charge ratio, this can be deconvoluted by performing a Fourier transform of the signal.

An outstanding feature of ICR analyser is its extremely high resolving power. A resolution of over 2×10^6 has been achieved using electrospray ionisation. The other important feature of ICR analyser is its MS/MS capability which makes it highly suitable for basic studies in gas phase chemical research.

Orbitrap. The orbitrap is the most recently introduced¹⁹ mass analyser. Here ions are electrostatically trapped in orbits around a central, spindle-shaped electrode.

The orbitrap is a modification of the ion trap developed by Kingdon in 1923. It consisted of a thin charged wire for confining charged particles. Ions are attracted toward the wire, but their angular momentum causes them to spiral around the wire in trajectories that have a low probability of hitting the wire.

In 1981, Knight introduced a modified outer electrode that included an axial quadrupole term that confines the ions on the trap axis. But, neither the Kingdon nor the Knight traps were reported to produce mass spectra because it was not known how to introduce ions in the electrostatic field and keep them confined without their zooming past.

Only recently Alexander Makarov and others figured out that the field must not be static when the ion is introduced—a potential barrier stopping the ions before they reach the electrode can be created by lowering the central electrode voltage when the ions are entering.

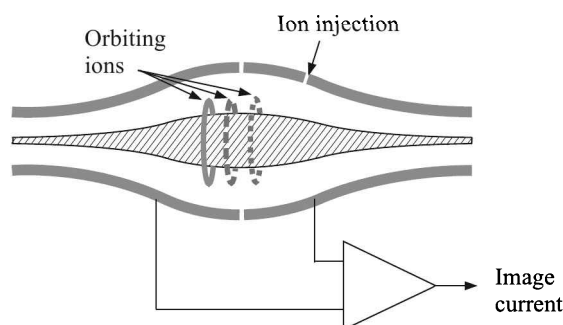


Fig. 14.46 Orbitrap electrode configuration.

In an orbitrap (Fig. 14.46), ions, injected tangentially into the electric field between the electrodes, are trapped because their electrostatic attraction to the inner electrode is balanced by centrifugal forces. Thus, ions cycle around the central electrode in rings. In addition, the

¹⁹Commercially available since 2005 from Thermo Scientific see www.thermo.com/orbitrap.

ions also move back and forth along the axis of the central electrode. Therefore, ions of a specific m/Ze ratio move in rings which oscillate along the central spindle. The frequency of these harmonic oscillations is independent of the ion velocity and is inversely proportional to the square root of the m/Ze ratio. The image current of the ion oscillation is sensed and Fourier transformed to determine relative abundance of different ions, as is done in the ICR mass analyser.

The mass accuracy (1–2 ppm), resolving power (up to 200,000) and dynamic range (~ 5000) of orbitraps are all rather high.

Ion Detectors

Mass analyser of a mass spectrometer separates bunches or streams of ions according to their individual mass-to-charge (m/Ze) ratio. The next stage of all mass spectrometers—apart from ICR and orbitrap, which by themselves are combined mass analysers and detectors—require an ion detector.

The earliest ion detectors like those of Aston, etc. consisted of photographic plates located at the end of the mass analyser. All ions of a given m/Ze would impact at the same place on the photographic plate making a spot. The darkness of the spot would indicate the intensity of that particular m/Ze .

Now, the choice of a detector depends on the design of the instrument and the type of analysis it is required to perform. The detector generates a signal from incident ions either by generating secondary electrons, which are further amplified or by inducing a current generated by a moving charge (similar to ICR and orbitrap). The most common types of ion detectors used nowadays are described here.

Faraday cup

A Faraday cup can just be a metal cup that is placed in the path of the ion beam. An attached electrometer measures the ion-beam current.

Alternatively, it may consist of a dynode. The incident ion strikes the dynode electrode (Fig. 14.47) which is made of a secondary electron emitting material like CsSb, GaP or BeO. As a result, the surface emits electrons and induces a current which is amplified and recorded.

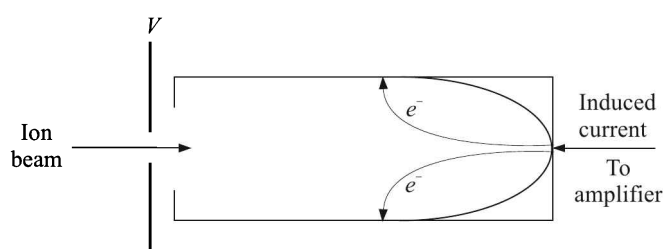


Fig. 14.47 Schematic of a Faraday cup ion detector.

Being suitable for use in an analogue mode, the Faraday cup is less sensitive than other detectors that are capable of operating in pulse-counting mode. However, it is very robust and is ideally suited for isotope analysis.

Electron multiplier tube

Electron multiplier tubes (EMT) are probably the most common means of detecting ions, especially when positive and negative ions need to be detected on the same instrument.

The EMT concept is similar to the alternative design of the Faraday cup. One type of EMT [Fig. 14.48 (a)] consists of a series of dynodes maintained at increasing potentials resulting in a cascaded amplification. The other type—*channeltron*²⁰, Fig. 14.48 (b)—consists of a cornucopia (horn) shaped specially formulated lead silicate glass dynode that exhibits property of electrical conductivity as well as secondary electron emission.

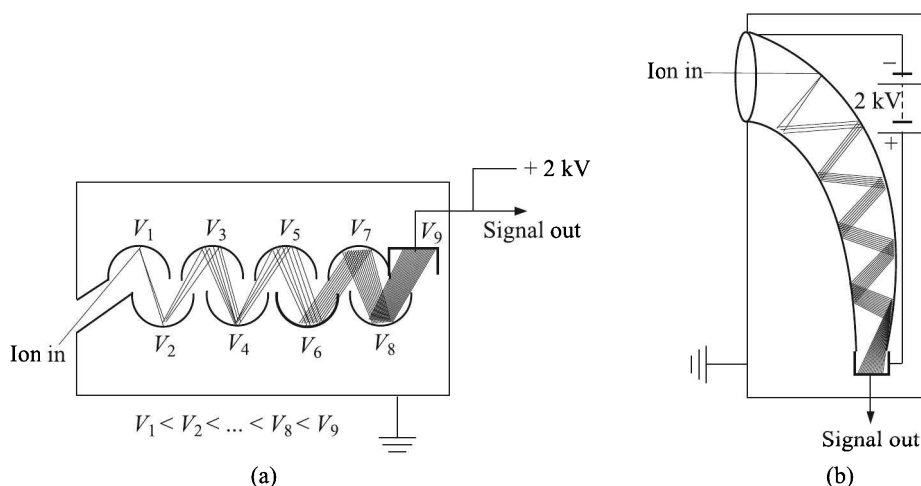


Fig. 14.48 Schematics of (a) linear EMT, and (b) Channeltron.

In both the cases, ions strike the initial amplification dynode surface ejecting secondary electrons which are then attracted either to the following series of dynodes, or into the continuous dynode where more secondary electrons are ejected in a repetitive process ultimately resulting in a cascade of electrons. The typical amplification of an EMT is $\sim 10^6$.

Since EMTs multiply the ion current, they can be used in analogue or digital mode.

Daly detector

Named after its inventor NR Daly, this detector consists of

- (i) a metal "doorknob" that emits secondary electrons when struck by an ion
- (ii) a scintillator, and
- (iii) a photomultiplier tube

as shown in Fig. 14.49.

A high voltage between the doorknob and the scintillator accelerates the electrons onto the phosphor screen where they are converted to photons. These photons are detected by the photomultiplier.

²⁰A registered trademark of Burle (formerly Galileo) see www.burle.com.

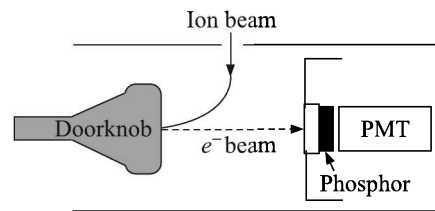


Fig. 14.49 Schematic of a Daly detector.

The advantage of the Daly detector is that the photomultiplier can be separated by a window which lets the photons through from the high vacuum of the mass spectrometer. This prevents an otherwise possible contamination and extends detector life span. In case of TOF mass analysers, the Daly detector allows for a higher acceleration after the field free region of the flight tube, thus improving the sensitivity for high mass ions.

Microchannel plate

A microchannel plate (MCP) consists of an array of 10^4 to 10^7 miniature electron multipliers oriented parallel to each other [Fig. 14.50(a)]. Typical channel diameters are 10 to 100 μm and the length to diameter (l/d) ratio of a channel is between 40:1 and 175:1.

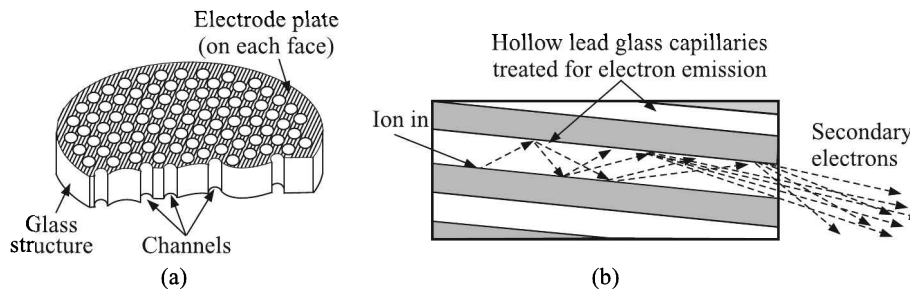


Fig. 14.50 (a) Cutaway view of a microchannel plate and (b) avalanche effect in a channel.

Channel axes are typically slanted at a small angle ($\sim 8^\circ$) to the MCP input surface. Like the channeltron, an ion that strikes a channel creates an avalanche of secondary electrons [Fig. 14.50(b)]. This cascading effect creates a gain. Gains of 10^6 to 10^8 are achievable. The governing physical parameter which determines gain is the l/d ratio. The higher the ratio, the higher the gain.

The channel matrix is usually formed by drawing, etching, or firing in hydrogen a matrix of lead glass capillaries that optimise secondary electron emission characteristic of each channel and render channel walls semiconducting properties so as to allow charge replenishment from an external voltage source.

The electrons exit the channels on the opposite side where they are detected by additional means, often simply a single metal anode measuring total current. In some applications, like X-ray imaging, each channel is monitored independently by CCD or phosphor in combination with photomultiplier tube.

Most modern MCP detectors consist of two microchannel plates with angled channels rotated 180° from each other producing a chevron (V-like) shape. In a chevron MCP the

electrons that exit the first plate start the cascade in the next plate. The advantage of the chevron MCP over the straight channel one is more gain at a given voltage. The two MCPs can either be pressed together or have a small gap between them to spread the charge across multiple channels.

MCPs have been used in a wider range of particle and photon detection systems perhaps more than any other kind of detector.

14.4 Infrared Analyser

Electromagnetic wavelengths from 0.78 to 1000 μm are considered to constitute infrared radiation. In IR spectroscopy, wavelength is often measured in *wave numbers*. Wave number ν_w is the number of waves per centimetre and is given by

$$\nu_w(\text{in cm}^{-1}) = \frac{1}{\lambda(\text{in cm})}$$

It is customary to divide the IR region into three sections as shown in Table 14.6.

Table 14.6 Sections of the infrared region

| <i>Section</i> | <i>Wavelength range</i> (μm) | <i>Wavenumber range</i> (cm^{-1}) |
|----------------|--|---|
| Near | 0.78 to 2.5 | 12800 to 4000 |
| Middle | > 2.5 to 50 | < 4000 to 200 |
| Far | > 50 to 1000 | < 200 to 10 |

Before we discuss the methods of IR analysis, let us consider what are the causes of infrared radiation.

Theoretical Background

The energy of a molecule comprises contributions from three processes:

1. Excitation of its electrons to higher energy levels
2. Vibrational motion of the molecule
3. Rotational motion of the molecule

Vibrational and rotational energy levels, like electronic ones, are also quantised. The three types of transitions are characterised by different amounts of energy. While electronic excitations occur at around 5 eV, vibrational and rotational excitations occur at around 0.1 eV and 0.005 eV respectively. From the Planck relation

$$E = h\nu$$

where $h = 6.63 \times 10^{-34}$ J-s, it is easy to see that the corresponding three wavelengths are nearly 2500 \AA , 12 μm and 250 μm respectively. Thus, the molecular electronic excitation takes place in the ultraviolet, the vibrational in the mid-IR and the rotational in the far IR regions of the electromagnetic wave spectrum. The following analysis will show that the rotational and vibrational motions of a simple molecule actually generate radiations of the stated wavelengths.

Rotational spectrum

In quantum mechanics, the rotation of a molecule is quantised. This means that its angular momentum and rotational energy can only assume certain fixed values. These values are simply related to the moment of inertia I of the molecule. In general, for any molecule there are three moments of inertia— I_a , I_b and I_c —about three mutually perpendicular axes a , b and c with the origin at the centre of mass of the system. However, for a linear molecule $I_a \ll I_b = I_c$. We consider a linear molecule in our discussions on the rotational spectra.

Let us consider a linear diatomic molecule comprising atoms of masses m_1 and m_2 . The molecule is rotating as a whole about the axis passing through its centre of gravity (Fig. 14.51).

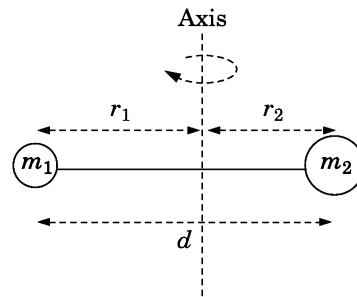


Fig. 14.51 Linear diatomic molecule.

If r_1 is the distance of mass m_1 from the axis of rotation

r_2 is the distance of mass m_2 from the axis of rotation

d is the distance between m_1 and m_2

ω is the angular velocity of rotation

I is the moment of inertia of the molecule

we have

$$d = r_1 + r_2$$

$$m_1 r_1 = m_2 r_2 \quad \Rightarrow \quad r_2 = \frac{m_1}{m_2} r_1$$

$$\therefore d = r_1 \left(1 + \frac{m_1}{m_2} \right) \quad (14.39)$$

Now,

$$I = m_1 r_1^2 + m_2 r_2^2$$

$$= m_1 r_1^2 \left(1 + \frac{m_1}{m_2} \right)$$

$$= \frac{m_1 m_2}{m_1 + m_2} d^2 \quad \text{[from Eq. (14.39)]}$$

$$= \mu d^2 \quad \text{where } \mu = \frac{m_1 m_2}{m_1 + m_2} \quad (14.40)$$

Equation (14.40) shows that the diatomic molecule behaves as a single atom of *reduced mass* μ . The quantised angular momentum of the rotating molecule is given by

$$p_\phi = I\omega = \frac{h}{2\pi}n_r \quad n_r = 0, 1, 2 \dots$$

$$\Rightarrow \omega = \frac{h}{2\pi I}n_r \quad (14.41)$$

The corresponding kinetic energy E is given by

$$E = \frac{1}{2}I\omega^2 = \frac{h^2}{8\pi^2 I}n_r^2 \quad [\text{from Eq. (14.41)}] \quad (14.42)$$

However, Eq. (14.42), which has been derived from the quantum theory, turns out to be

$$E = \frac{h^2}{8\pi^2 I}n_r(n_r + 1)$$

according to quantum mechanics.

Now, if E_{r1} is the energy corresponding to the quantum state n_{r1} of the molecule, and E_{r2} corresponding to n_{r2} when a quantum of energy $h\nu_r$ is emitted or absorbed, we have from the Planck relation

$$h\nu_r = E_{r1} - E_{r2} = \frac{h^2}{8\pi^2 I}[n_{r1}(n_{r1} + 1) - n_{r2}(n_{r2} + 1)]$$

$$= \frac{h^2}{8\pi^2 I}(n_{r1} - n_{r2})(n_{r1} + n_{r2} + 1) \quad (14.43)$$

If $(n_{r1} - n_{r2}) = \pm 1$, the positive sign corresponding to absorption and the negative to emission, we have from Eq. (14.43)

$$h\nu_r = \frac{h^2}{4\pi^2 I}(n_{r1} \pm 1)$$

$$= \frac{h^2}{4\pi^2 I}(n_r \pm 1) \quad [\text{dropping the subscript}]$$

$$\Rightarrow \nu_r = \frac{h}{4\pi^2 I}(n_r \pm 1) \quad (14.44)$$

Equation (14.44) predicts a spectrum of equally spaced lines having a frequency interval of

$$d\nu_r = \frac{h}{4\pi^2 I}$$

The moment of inertia of HCl is $\sim 2.7 \times 10^{-40}$ g-cm² as found out from other methods. If we calculate the wavelength of its rotational line for any transition to the next rotational state ($dn_r = 1$), we find

$$\lambda = \frac{c}{\nu} = \frac{4\pi^2 I c}{h}$$

$$= \frac{(4\pi^2)(2.7 \times 10^{-40})(3 \times 10^{10})}{6.5 \times 10^{-27}}$$

$$= 163 \mu\text{m}$$

The order of magnitude of the wavelength agrees well with what we have discussed before. In this case, the wavelength is in the far IR region. In most of the cases, however, rotational spectra fall in the microwave region.

In reality, rotational spectroscopy is practical only in the gas phase where the rotational motion is quantised. In solids or liquids the rotational motion is usually quenched due to collisions.

Vibrational spectrum

A molecular vibration occurs when atoms in a molecule are in periodic motion while the molecule as a whole has constant translational and rotational motions. In general, a molecule made of N atoms has $(3N - 6)$ normal modes of vibration, though for linear molecules the number is $(3N - 5)$ because the vibration about its molecular axis cannot be observed. From this it follows that a linear diatomic molecule has only one normal mode of vibration.

To a first approximation, the motion in a normal vibration can be described as a simple harmonic motion. According to quantum mechanics, the vibrational energy of a harmonic oscillator is given by

$$E_v = \left(n_v + \frac{1}{2}\right)h\nu_0$$

where $n_v = 0, 1, 2, \dots$ and ν_0 is the natural frequency of vibration. We know, for a harmonic oscillator

$$\nu_0 = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (14.45)$$

where k is the bond strength and μ is the reduced mass. The average value of k for single bonds is $\sim 5 \times 10^5$ dyne/cm. For double bonds, it is $\sim 10 \times 10^5$ dyne/cm and so on. So, for the transition from one vibrational state to the next one (i.e. $dn_v = 1$) we have

$$dE_v = h\nu_0 \quad (14.46)$$

Substituting the value of ν_0 from Eq. (14.45), we get from Eq. (14.46) on rearranging terms

$$\lambda = \frac{c}{\nu_0} = 2\pi c \sqrt{\frac{\mu}{k}}$$

where c is the velocity of light. For the HCl molecule, its value turns out to be

$$\begin{aligned} \mu &= \frac{m_{\text{H}}m_{\text{Cl}}}{m_{\text{H}} + m_{\text{Cl}}} = \frac{(1 \times 1.67 \times 10^{-24})(35.5 \times 1.67 \times 10^{-24})}{(1 + 35.5)(1.67 \times 10^{-24})} \\ &= \frac{35.5}{36.5} \times 1.67 \times 10^{-24} \text{ g} \\ &= 1.624 \times 10^{-24} \text{ g} \end{aligned}$$

$$\begin{aligned} \therefore \lambda &= (2\pi)(3 \times 10^{10}) \sqrt{\frac{1.624 \times 10^{-24}}{5 \times 10^5}} \\ &= 3.397 \text{ } \mu\text{m} \end{aligned}$$

The order of magnitude of the wavelength of vibrational line of HCl agrees well with what we discussed before. These transitions typically require an energy that corresponds to the IR region of the spectrum.

By the way, from what we have discussed it is clear that apart from identifying compounds, the measurement of rotational spectrum can also provide us information about bond length (d) and moment of inertia (I) of molecules while that of vibrational spectrum can help us find the bond strength (or force constant) between atoms of molecules.

In IR spectroscopy, only vibrational and rotational changes of state caused by an incident radiation are observed. Except homonuclear diatomics — O₂, N₂, H₂, Cl₂, etc. — and monatomics — He, Ne, etc. — all other molecules, *possessing dipole moments*, interact with the IR radiation and offer a few fingerprint absorption spectra. The strength of absorption at those frequencies is a measure of the concentration of the species. However, IR analysis alone may sometimes be misleading except when the component molecules of the sample have significantly different atomic groupings. Similar molecules, such as series of homologous hydrocarbons, have very similar IR spectra. The part of the spectra between 2.5 and 15 μm (4000 and 670 cm^{-1}) offers best discrimination between molecules and is, therefore, known as the *fingerprint region*.

Beer-Lambert law

IR analysers are mostly used for gas and liquid samples. Quantitative estimates of concentration are based on the Beer-Lambert law which states

$$A = abc = \log \frac{I_R}{I_S} \quad (14.47)$$

where A is the absorbance

a is the absorption coefficient of the pure component of interest

b is the path length through the sample

c is the concentration of the absorbing species

I_R is the intensity of the IR that passed through the reference cell

I_S is the intensity of the IR that passed through the sample cell

The law can be derived as follows. Suppose a parallel beam of radiation of intensity I traverses through an infinitesimally small distance dx of an absorber and thereby suffers a loss of intensity $-dI$. Then,

$$-dI = k' I dx$$

where k' is a constant that depends on the wavelength of the incident radiation. On rearranging Eq. (14.47), integrating it over the length b of the absorber, and assuming that the concentration of the absorber remains constant, we get

$$\begin{aligned} - \int_{I_R}^{I_S} \frac{dI}{I} &= k' \int_0^b dx \\ \Rightarrow \ln \left(\frac{I_R}{I_S} \right) &= k'b \quad [c \text{ constant}] \end{aligned} \quad (14.48)$$

Equation (14.48) is known as *Lambert's law*. Next, we assume that the length of the absorber remains constant, while concentration of absorbent molecules that interact with the radiation changes from 0 to c to yield I_S for an incident I_R . Then,

$$\begin{aligned} -dI &= k'' I dc \\ \Rightarrow -\int_{I_R}^{I_S} \frac{dI}{I} &= k'' \int_0^c dc \\ \Rightarrow \ln\left(\frac{I_R}{I_S}\right) &= k'' c \quad [b \text{ constant}] \end{aligned} \quad (14.49)$$

where k'' is a constant. Equation (14.49) is known as *Beer's law*.

Applying the law of variation to Eqs. (14.48) and (14.49), when both b and c vary, we get

$$\ln\left(\frac{I_R}{I_S}\right) = kbc$$

Changing the base of logarithm to 10, we get

$$\log\left(\frac{I_R}{I_S}\right) = abc \quad (14.50)$$

The absorbance A is defined as

$$A = \log\left(\frac{I_R}{I_S}\right) \quad (14.51)$$

Hence, from Eqs. (14.50) and (14.51) we get Eq. (14.47) which is the Beer-Lambert law. However, more often than not, it is referred to as the Beer law.

The transmittance T is defined as

$$T = \frac{I_S}{I_R}$$

So,

$$A = \log\left(\frac{1}{T}\right) = -\log T \quad (14.52)$$

From Eqs. (14.47) and (14.52) it is clear that the plots of absorbance and transmittance should look like those given in Fig. 14.52.

It is seen from Fig. 14.52(a) that the Beer-Lambert law does not hold good for concentrations beyond 10^{-3}M . The deviation occurs because absorbent's refractive index, which more or less remains constant at low concentrations, starts varying at higher concentrations.

Analyser Types

Analysers are generally classified as

1. Non-dispersive and
2. Dispersive

Non-dispersive analyses are generally done in process industries as well as in environmental measurements.

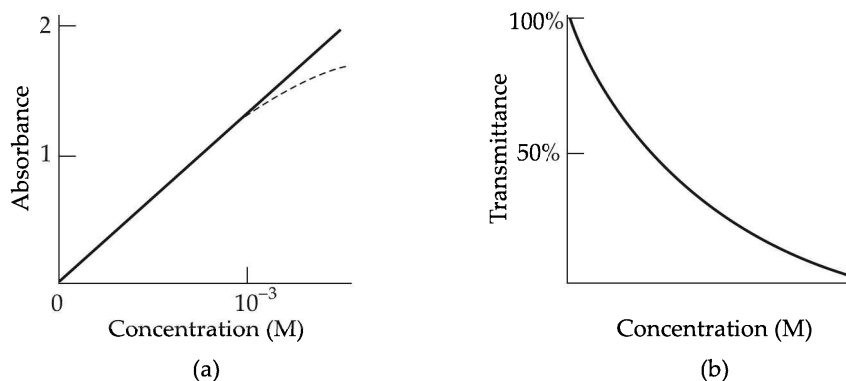


Fig. 14.52 Plots of (a) absorbance vs. concentration, and (b) transmittance vs. concentration. The broken line in (a) indicates the deviation from the Beer-Lambert law.

Non-dispersive infrared (NDIR) analysis

NDIRs can also be of two types

1. Single-beam
2. Dual-beam

We consider a dual-beam NDIR (Fig. 14.53), single-beam instrument being very similar except that it does not use a reference tube and the corresponding optical path.

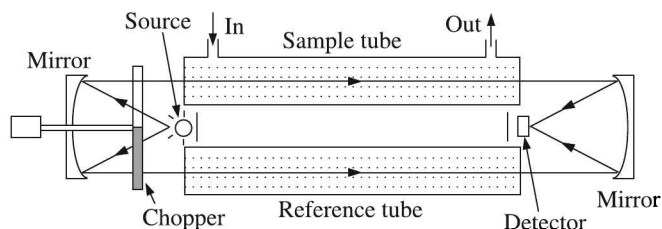


Fig. 14.53 Dual-beam NDIR.

It may be noted that concave mirrors, rather than convex lenses, have been used to concentrate the IR beam. The reason is that ordinary glass does not transmit IR beyond 3000 nm and therefore not suitable for far IR optics. Crystalline NaCl or NaI transmit IR up to 15000 nm and therefore may be used for construction of lenses. But they are hygroscopic and therefore, need to be maintained in controlled humidity. Mirrors do not suffer from this limitation.

In this configuration the IR radiation is chopped such that it alternately falls on the sample and reference tubes. Chopping also reduces noise bandwidth through synchronous demodulation at the detector. The reference tube may contain a non-absorbing gas or other gases of the sample which are not of interest. If the sample contains the species of interest, the detector will show a reduction in intensity.

Dispersive IR analysis

Dispersive (or normal) IR spectrophotometers incorporate a monochromator over and above other components of an NDIR. A monochromator is a dispersive device—a prism or grating—that helps isolate different wavelengths of the radiation from the source. Monochromators use various combinations of mirrors to increase path lengths of rays after dispersion, so that dispersed lines are well separated from each other. In general, diffraction gratings offer more separation between lines though lines separated by prisms have more intensity. One mounting, called *Littrow mounting* (Fig. 14.54), utilises a mirror to disperse the ray twice by the same prism, thus increasing separation between the lines.

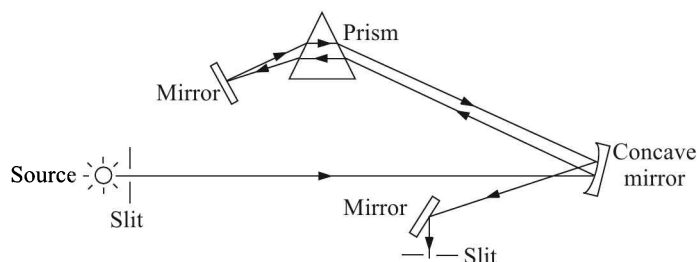


Fig. 14.54 Littrow mounting monochromator.

With the help of monochromators, normal IR spectrophotometers can scan a sample over the entire range of IR spectrum to find its IR absorption lines. We now discuss the source and detector of IR analysers.

IR Source

IR sources are inert solids that are electrically heated from 1500 to 2000 K when they emit a continuous spectrum between 1 and 20 μm . Three such sources are common:

1. Globar
2. Nernst filament
3. Nichrome wire

Globar

The globar²¹ rod is a silicon carbide rod of nearly 5 cm in length and 0.5 cm in diameter. When raised to a temperature of around 1500 K by passing an electrical current through it, the globar emits radiation of 0.6 μm to 26 μm wavelength. It needs water cooling at the central region to prevent arcing and burn-out.

Nernst filament

The Nernst filament is made by fusing oxides of zirconium and yttrium. Taken in a rod form of 1 mm to 2 mm diameter and 20 mm to 30 mm length, its ends are sealed with platinum wires to allow passage of current through it. When raised to a temperature of about 2200 K, it emits radiation almost in the same range as that of the globar rod. Since its resistance decreases at higher temperatures, it needs a ballast resistance in series with it.

²¹Derived from 'glow bar'.

Nichrome wire

Tightly wound nichrome²² wire when raised to a temperature of 1100 K by passing electric current through it, emits IR radiation between 4 μm and 15 μm .

IR Detectors

Detectors can be of various types of which we will consider only three, namely

1. Microphone-type
2. Thermal-type
3. Solid state-type

Microphone-type detector

Microphone-type detectors are basically capacitive d/p sensors. They consist of two absorption chambers separated by a diaphragm and a fixed perforated plate (Fig. 14.55).

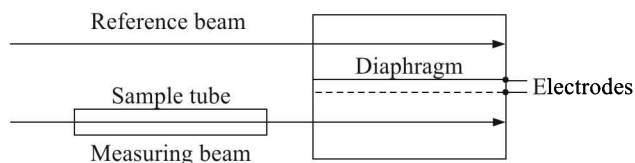


Fig. 14.55 Microphone-type detector.

The diaphragm and the perforated plate form two plates of a capacitor. The chambers are filled with the sample gas. When the IR beam gets absorbed by its passage through the sample tube, a lower intensity beam hits the corresponding chamber in the detector. This results in a temperature difference and hence a pressure difference between the two chambers. The corresponding deflection of the diaphragm produces a change in the capacitance which is measured by a suitable method.

Thermal-type detector

Thermal devices either measure temperature of an IR absorber (black) directly or produce expansion of gas in an enclosure which is sensed. Thermocouple or bolometer belongs to the first category. It has to be noted that the change of temperature of the black absorber is rather small and it may produce an electrical output of 1 μV from the thermocouple. Therefore, extreme precautions have to be taken to shield the detector and the cables from thermal or electromagnetic noises.

Production of expansion of gas is resorted to in Golay pneumatic cells (Fig. 14.56). Here, the absorber, thermally insulated from the enclosure, heats the enclosed gas when the IR beam strikes it.

The expanded gas deflects the mirrored diaphragm. The deflection is enhanced by what is called an optical lever and sensed by a photocell.

²²The common heater wire.

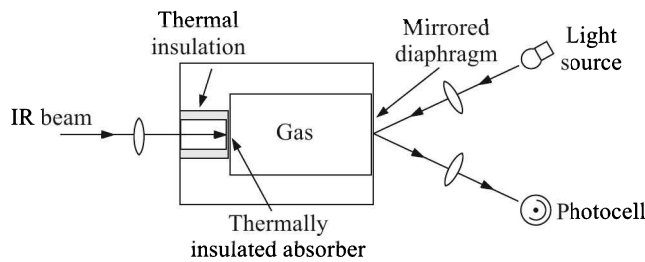


Fig. 14.56 Golay pneumatic cell.

Bolometer consists of a blackened platinum strip in an evacuated glass vessel. IR irradiation increases the temperature and therefore, the resistance of the strip. This change in resistance is measured by a Wheatstone bridge. It is, therefore, one form of RTD.

Solid state detectors

Solid state detectors are either photoelectric or photoconductive devices. Indium antimonide, lead sulphide, cadmium telluride, etc. are used—each having different range of IR sensitivity—to generate photoelectricity.

PbS and PbSe work as good photoconductors in the NIR region. They are more sensitive than thermal detectors and work excellently at higher chopping frequencies.

Finally, it needs to be mentioned that tunable diode laser detectors are now available. They have very high spectral resolution and good power output. But they require liquid nitrogen temperature environment and other related accessories which can be better maintained at research laboratories than process industries.

Fourier Transform Infrared Analysis

In a normal IR analyser, a dispersive device—prism or grating—is used to select the radiation of a particular wavelength from the IR spectrum, irradiate the sample with it and then its transmittance or absorbance is recorded. In this way the sample is scanned over the entire IR region and the corresponding IR transmittance or absorbance spectrum of the sample is obtained. This, indeed, is a slow process.

A Fourier transform infrared (FTIR) spectrometer, on the other hand, helps one obtain the same spectrum in a very short period—a few seconds—and with a better signal-to-noise ratio.

This is achieved by producing an interferogram pulse of the incident radiation, passing the interferogram through the sample, detecting the resultant signal and then performing a fast Fourier transform (FFT) of the detected signal to produce the required IR spectrum. The process is schematically shown in Fig. 14.57.

Principle of Fourier transform spectrogram

The principle of the Fourier transform can be understood by comparison with the behaviour of a tuning fork. If the fork is exposed to sound waves of varying frequencies, it will resonate only at frequencies which correspond to its fundamental frequency of vibration or overharmonics.

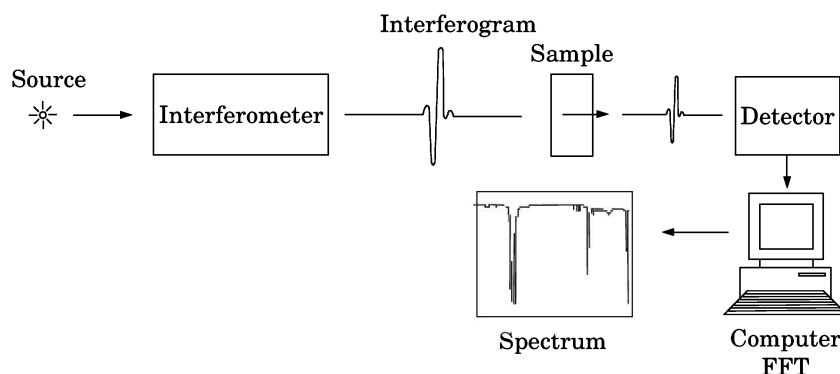


Fig. 14.57 Schematic illustration of the FTIR spectrometer.

However, if the fork is struck it will produce a sound that comprises all of its characteristic frequencies. A Fourier transform of this sound will give us these individual characteristic frequencies. We also note that the Fourier transform of the composite sound is meaningful only if its constituents were produced at the same time, that is, they have temporal coherence.

So, to obtain the Fourier transform of the incident IR we need to generate a pulse comprising all the frequencies of the IR region and that these frequencies should have temporal coherence. This is achieved with the help of an interferometer—mostly, the Michelson interferometer²³. Basically, it consists of two mirrors, F and M, situated at right angles to each other. A beam-splitter, BS, lies in between (Fig. 14.58).

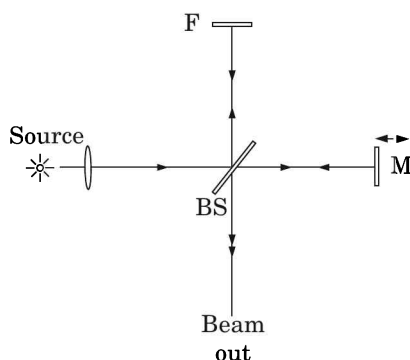


Fig. 14.58 Michelson interferometer arrangement.

Parallel radiation, entering the interferometer, is either transmitted to the movable mirror M, or reflected to the fixed mirror F, on hitting the beam-splitter BS. 50% of the radiation reflected from F and M comes out of the interferometer while the other 50% goes back to the source. The beams of wavelength λ that come out cancel each other if the paths traversed by them differ by $\lambda/2$ or its odd multiple, while they reinforce each other if the path difference is λ or its integral multiple. We know, these are called destructive and constructive interference respectively.

²³see, for example, *Physics, Part II*, Section 43-7, by D Halliday and R Resnick, 2nd Ed, Wiley Eastern.

So, for a monochromatic incident beam, the movement of M will generate the intensity variation of the output beam as shown in Fig. 14.59. This is called an ‘interferogram’.

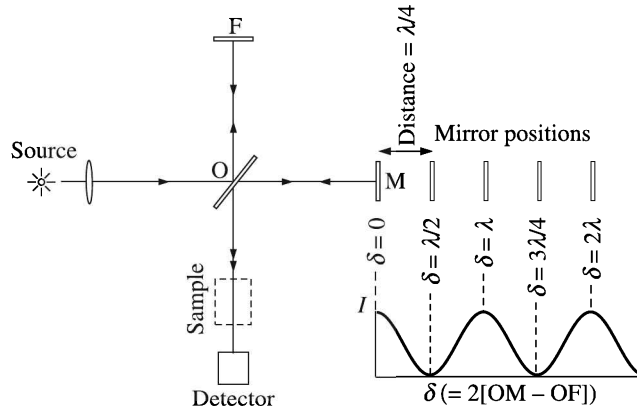


Fig. 14.59 Resulting interferogram for a monochromatic radiation.

If the incident radiation consists of two wavelengths λ_1 and λ_2 , the result will be as shown in Fig. 14.60.

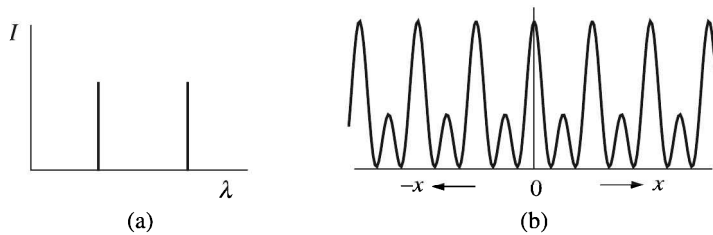


Fig. 14.60 (a) Incident radiation consisting of two wavelengths, and (b) the resulting interferogram.

We observe that at zero path difference, the two waves reinforce. Which means, the waves can be represented by cosine functions that have maxima at 0. So, the resulting waveform can be generated by summing two cosine functions. If the incident radiation is white,²⁴ the output interferogram will be a *centre burst* at zero path difference and a very complex pattern of waves symmetrically dispersed about it (Fig 14.61).

We note that this centre-burst pattern consists of contribution from all the frequencies. And since this is an interferogram, it has had temporal coherence. So, this is our desired pulse which is to be sent through the sample.

After passing through the sample, the resulting interferogram is Fourier transformed to get the individual lines of the spectrum. Now, the question is how to identify the frequency corresponding to a spectral line? This is possible if we know the path difference between the interfering waves.

So, the next problem is to figure out the path difference. This is done with the help of the interferometer itself by feeding it with a source of monochromatic radiation of known wavelength. A helium-neon laser source emitting a radiation of wavelength 632.8 nm is

²⁴That means, it consists of all frequencies.

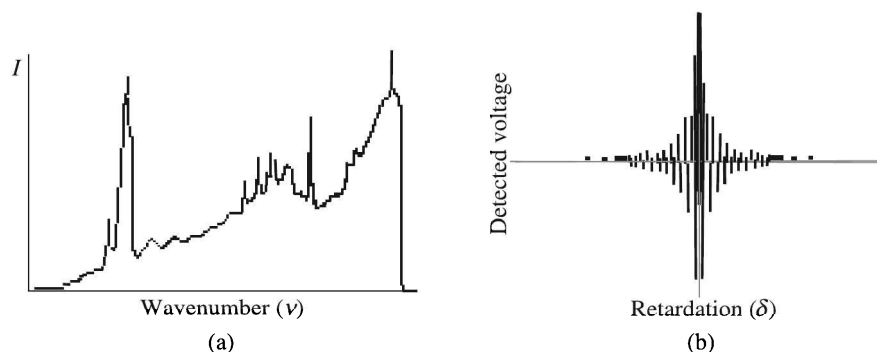


Fig. 14.61 (a) Typical white radiation, and (b) its interferogram.

deployed to generate a strictly cosine variation of intensity for every 632.8 nm of path difference. This reference interferogram is electronically converted to square waves that generate wavelength markers on the spectrum. In order to avoid contamination of the IR interferogram, the He-Ne laser beam is moved out of the optic axis of the interferometer (Fig. 14.62).

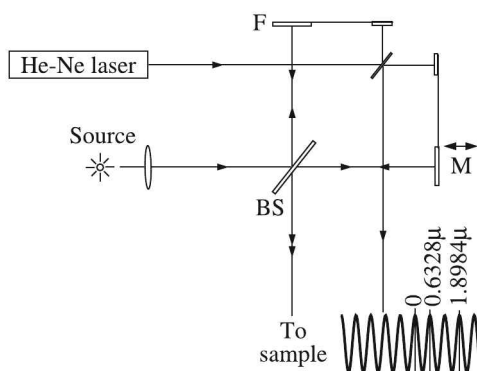


Fig. 14.62 Placement of He-Ne interferometer arrangement.

This is, in a nutshell, the principle of operation of an FTIR spectrometer.

It is to be noted that an FTIR spectrometer is a single beam instrument. Hence, its analysis is susceptible to errors caused by the presence of background contaminants like water vapour, carbon dioxide, etc. in the ambience. So, it is a practice to run an analysis, maybe once in a day, without the sample and record a background spectrum. The ratio of the background spectrum and the raw spectrum of the sample yields the true spectrum of the sample (Fig. 14.63).

Advantages of the FTIR

FTIR spectra possess three prominent advantages—called Fellgett, Jaquinot and Connes advantages—over normal IR spectra.

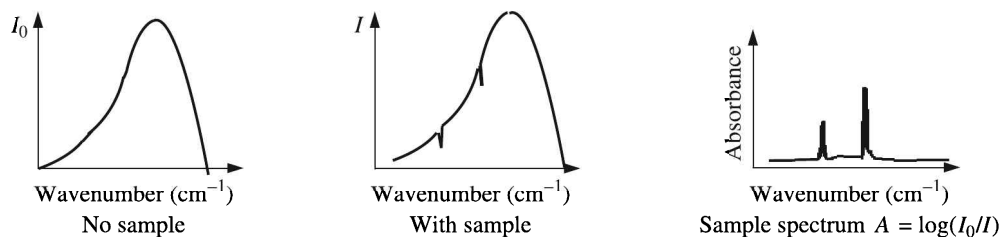


Fig. 14.63 Construction of the sample spectrum eliminating the background.

Fellgett advantage. The Fellgett advantage²⁵ states that if N is the number of elements comprising the resultant spectrum, the FT spectrum produces a gain of $\sim \sqrt{N}$ in the signal-to-noise ratio because of its simultaneous scan of the entire spectrum.

Jaquinot advantage. The Jaquinot advantage relates to higher throughput of the FTIR because unlike dispersive spectroscopy no slits attenuate the energy-starved incident IR beam. Of course, the source could be heated to generate IR of higher intensity. But then, one runs the risk of generating corrosive $(\text{NO})_x$ gases from the combination of nitrogen and oxygen of the atmosphere. The $(\text{NO})_x$ gases not only corrode metal and plastic parts of the instrument but also produce unwanted lines in the spectrum.

Connes advantage. The Connes advantage is that since the FTIR determines frequencies from the path difference, it gives very accurate frequencies in the spectrum and does not require calibration. This enables processing techniques such as spectral subtraction. For example, one may subtract the spectrum of the solvent to obtain that of the solute.

Fellgett disadvantage

There is a Fellgett disadvantage however. It states that since all regions of the spectrum are observed simultaneously, if there is noise in the IR source, it spreads everywhere throughout the FTIR spectrum.

Another point needs be mentioned in this context. All components of the FTIR instrument are fixed except the movable mirror, which must move back and forth smoothly. The movement must be reproducible and free from wobble or shake. These requirements are hard to achieve and so, they call for high precision engineering.

Secondly, be it dispersive, non-dispersive or FTIR spectrometer, its sample cells and optical components—except for mirrors silvered on the front surface—cannot be made of glass or quartz because they absorb infrared radiation. They are made of salts like NaCl, KBr, CsBr, LiF that are transparent to IR. But they are all water soluble and hence they have to be maintained in a moisture-free ambience.

14.5 Atomic Spectrometry

Atomic spectrophotometry is a useful technique for the detection of *elements* in a sample. Three methods namely,

²⁵aka multiplex advantage.

1. Atomic emission
2. Atomic absorption
3. Atomic fluorescence

are used. In order to understand the relationship between these techniques, it is necessary to consider the structure of an atom and the atomic process involved in each technique.

We know, an atom is made up of a nucleus surrounded by electrons. Each element has a specific number of electrons that occupy orbital positions in an orderly and predictable manner around the corresponding nucleus. When the atom is in its lowest energy state, known as the *ground state*, all the electrons occupy their normal orbitals as enunciated in the periodic classification of elements.

If energy of the right magnitude is applied to an atom in its ground state, the atom absorbs it to promote an outer electron to a less stable *excited state*. The excited electron spontaneously decays to its initial state by emitting a photon (Fig. 14.64). Since each element possesses

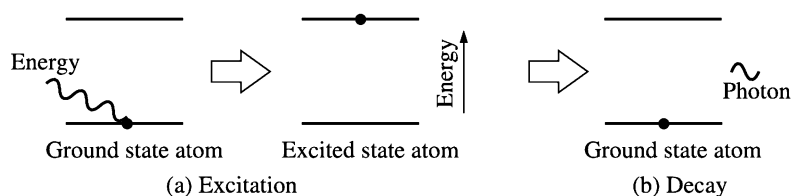


Fig. 14.64 (a) Excitation of an atom from its ground state, and (b) decay to the ground state.

a unique electronic structure, the wavelength of the emitted radiation is characteristic of the individual element. Depending upon the complexity of the electronic configuration of the element, the characteristic radiation may comprise many wavelengths resulting from many electronic transitions.

In all these spectrophotometric studies, individual atoms need to be produced from the sample which normally has the form of a solution of ions. This is done by producing a fine spray of the sample (nebulisation) and subject the spray to heating by a flame of appropriate temperature.

Nebulisation

Nebulisers can be of pneumatic crossflow type or ultrasonic.

Crossflow nebuliser

In the crossflow type nebuliser, two capillary tubes—one vertical for the sample uptake and another horizontal for the gas flow—are held at right angles to each other, as shown in Fig. 14.65.

When gas flows through the horizontal tube, there is a pressure drop in the area owing to the Bernoulli effect. As a result the liquid from the sample bottle is sucked in there producing a spray of aerosols by the collision with flowing gas molecules.

Ultrasonic nebuliser

In an ultrasonic nebuliser, an RF supply between 200 kHz and 10 MHz through a coil drives a piezoelectric crystal to vibrate at that high frequency. The liquid sample coming in contact

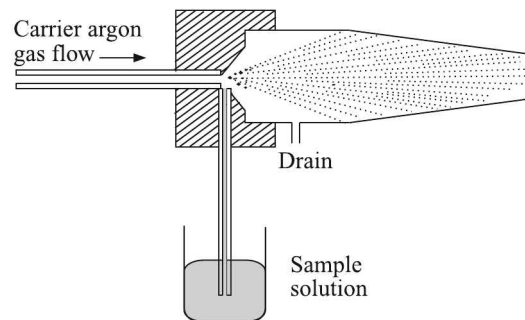


Fig. 14.65 Schematic diagram of a crossflow nebuliser.

with the vibrating crystal breaks down into aerosols. Liquids enter a vessel through a side port and sit on top of a vibrating element. As waves move through the liquid, the liquid begins to be pushed upward making a small fountain. Off the surface of this fountain small particles begin to float above the liquid and appear like smoke (Fig. 14.66). Gravity will have very little affect on these particles produced. To move the particles, a small air flow/gas or vacuum is needed.

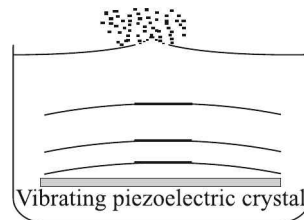


Fig. 14.66 Nebulisation produced by ultrasonic vibration.

When the spray is subjected to heating, it goes through different states and stages such as, desolvation, vaporisation, atomisation, etc. successively. These are depicted in Fig. 14.67 where the metal atom is designated by M , anion by A and heating by filled arrows.

Excitation Sources

The excitation sources are of many types. We describe seven of them, namely

1. Flame excitation source
2. Electrothermal source
3. Arc emission source
4. Spark source
5. Direct current plasma source
6. Inductively coupled plasma source
7. Laser-induced excitation source








| <u>Process</u> | <u>Product</u> | <u>State</u> |
|----------------|---|--------------|
| | $M^+ + e$ | Gas |
| Ionisation |  | |
| | M^* | Gas |
| Excitation |  | |
| | $M^0 + A^0$ | Gas |
| Atomisation |  | |
| | MA | Gas |
| Vaporisation |  | |
| | MA | Liquid |
| Liquefaction |  | |
| | MA | Solid |
| Desolvation |  | |
| | $M^+ + A^-$ | Aerosol |
| Nebulisation |  | |
| | $M^+ + A^-$ | Solution |

Fig. 14.67 Different states and stages of a solution when heated. Filled arrow indicates application of heat.

Flame excitation source

A suitable flame desolvates, liquefies and atomises the sample. Further heating excites the atoms to higher energy levels. Table 14.7 lists some commonly used flames and the temperatures they produce. Figure 14.68 shows the schematic of a burner where the sample solution is led to the burner by aspiration. In some designs, the sample is nebulised, mixed with the fuel and oxidant and then the whole mixture is fed to the burner.

Table 14.7 Common flames and temperatures

| <i>Fuel</i> | <i>Oxidant</i> | <i>Temperature (K)</i> |
|-------------------------------|------------------|------------------------|
| H ₂ | Air | 2000 – 2100 |
| H ₂ | O ₂ | 2600 – 2700 |
| C ₂ H ₂ | Air | 2100 – 2400 |
| C ₂ H ₂ | N ₂ O | 2600 – 2800 |

Despite the development of other variants, the flame atomiser remains the mostly used atomiser. The requirements of a satisfactory flame source are:

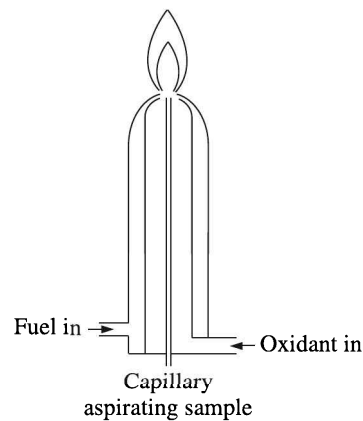


Fig. 14.68 Schematic diagram of a flame source.

1. It should provide the required temperature with a suitable fuel/oxidant ratio. Temperatures high enough to cause ionisation of analyte atoms are undesirable. Standard tables are available to determine a suitable fuel/oxidant mixture for a particular analyte.
2. Spectrum of the flame itself should not interfere with the emission lines of the analyte.

In atomic emission spectrophotometry, C_2H_2 /air flame is used for practically all analytes involving alkali metal elements. For analyses of alkaline earth metals as well as Ga, In, Tl, Cu, Co, Cr, Ni and Mn, the C_2H_2/N_2O flame is used.

A sheath of an inert gas, blown around the outside of the flame, elongates the flame. As a result, the noise is reduced and a wider range of excitation conditions over a small flame volume is provided.

Electrothermal source

In a flame atomiser, the analyte atoms stay in contact with the flame for a short period. Also, the nebulising system wastes some amount of the sample. To take care of these problems, electrically heated devices, such as graphite furnaces (Fig. 14.69) and carbon rod atomisers are used.

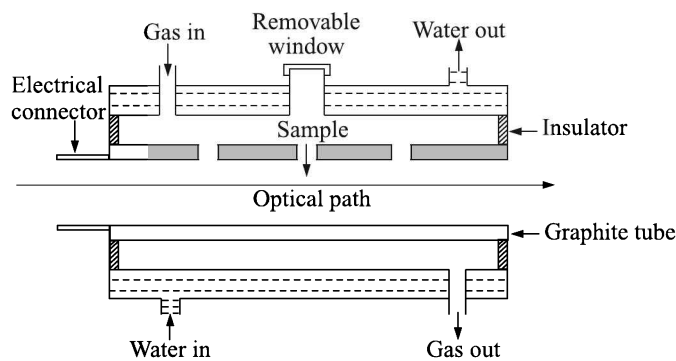


Fig. 14.69 Graphite furnace.

Graphite furnace. The size of the graphite cylinder used in a graphite furnace is about 28 mm (L) \times 8 mm (ϕ). A low voltage, high current power supply delivering ~ 3.6 kW is used to heat the tube. Liquid samples are injected by a microsyringe through the top opening while solid samples are introduced through the tube-end by a special sampling spoon or a tungsten microboat. An inert gas flow through the sample chamber ensures that matrix components vaporised during the ash formation step are quickly removed, leaving no deposits on the inner wall of the tube. Two removable quartz windows at the tube-ends (not shown in the diagram) are used to prevent the entry of air.

Carbon rod atomiser. A carbon rod atomiser is a simple graphite crucible having a hole on the opposite sides of its wall so as to allow incident light to pass through it. The sample, placed inside the crucible, is heated by passing a low voltage, high current through the attached electrodes. A schematic diagram of the atomiser is shown in Fig. 14.70(a).

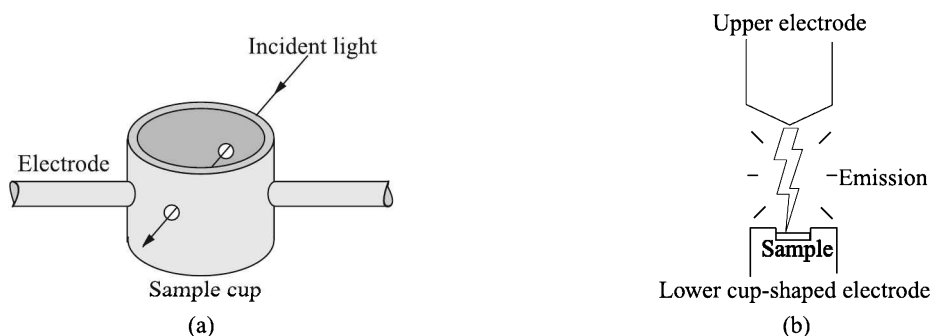


Fig. 14.70 (a) Carbon rod atomiser, and (b) arc emission source.

Arc emission sources

Arc emission sources use an electrical discharge between two electrodes to atomise and excite analyte atoms. The applied direct current and voltages are 5 to 30 A and 10 to 25 V respectively and high purity graphite is used as electrode material. A temperature of 4000 to 6000 K is attained in this way. The graphite electrode has the advantage of withstanding a high temperature because it is a refractory material. Apart from that, it is chemically resistant to acids, etc. and its emission lines are few which do not interfere with the measurements.

Arc emission sources are better suited for qualitative or semi-quantitative analyses of solid samples. Non-conducting samples are ground with graphite powder and placed in a cup-shaped lower electrode [Fig. 14.70(b)].

Spark sources

A spark emission source consists of a sparking stand and a spark generator. The sparking stand is shown in Fig. 14.71. The sample surface is cleaned with an abrasive and is held on the support table by a holding device in such a way that it shuts off the sparking chamber, known as *Petrey's chamber*. The chamber is swept with a steady flow of argon which drives out the aerosol of metal particles left by the discharge.

In *Petrey's chamber*, a tungsten electrode, acting as the cathode, faces the sample, which itself acts as the anode. The spark consists of two phases. In the first phase, a low energy

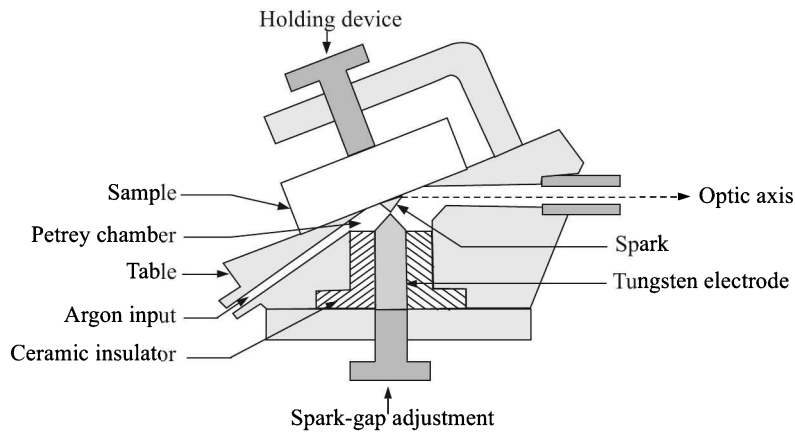


Fig. 14.71 Sparking stand.

discharge is produced when a potential difference of about 10 kV is applied for a few microseconds to ionise argon and create a conducting plasma. No sooner the plasma forms, than the second phase starts. A 300 to 500 V potential difference, delivering about 100 to 400 W, causes the sample to melt and evaporate at the spark point, generating an emission of the characteristic radiation of the sample. The stages are shown in Fig. 14.72. The total duration of both phases of the spark is on the order of a few milliseconds. So, a large number of sparks are necessary to carry-out an analysis.

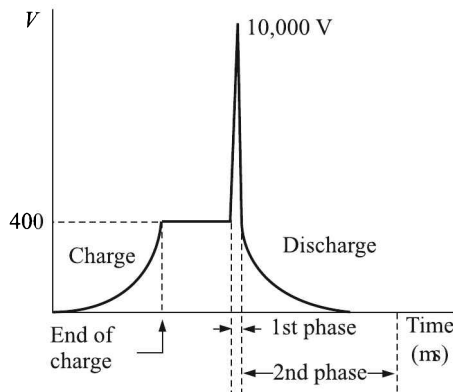


Fig. 14.72 The two phases of spark discharge.

The arc and spark excitation sources are used for semiquantitative or qualitative analyses because of the flicker of the emitted light. They are commonly used for routine analysis of metals, alloys, ores, soils, etc. because the methods are more suitable for handling solid samples. However, they are being replaced with plasma and laser sources in many applications.

Direct current plasma sources

In a dc arc source, if the current is increased, an arc of bigger cross-section results, keeping the energy density and hence the available energy for atomic excitation the same. In other words, the arc temperature does not increase if the current is increased.

However, the arc temperature can be increased if the arc cross-section is not allowed to increase with the increase of the arc current. This is accomplished in a direct current plasma (DCP) source by squeezing the plasma with a high velocity vortex of an inert gas like argon. The flowing inert gas lowers the temperature of the outer edge of the plasma thus inhibiting the formation of ion there. As a result, the cross-section remains small and the current density increases.

Two graphite electrodes and a tungsten cathode form an inverted Y. An argon flow of about 8 L/min surrounds the anodes (Fig. 14.73). To initiate the arc, the electrodes are brought in

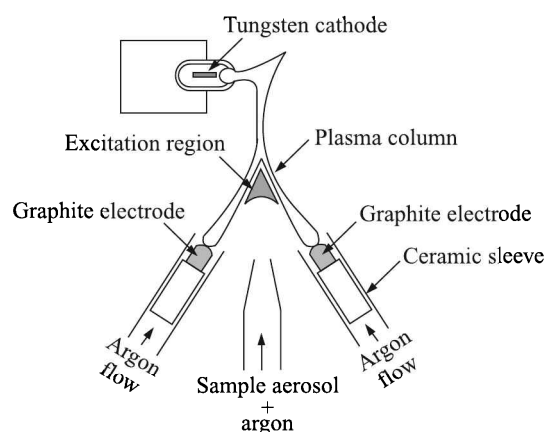


Fig. 14.73 The dc plasma source diagram.

contact and then separated. About 400 V at 7.5 A sustains the plasma once it is ignited. It generates a temperature of 8000 to 10000 K. The Y-configuration is necessary to stabilise the plasma.

A carrier argon gas sprays the nebulised sample at the excitation region which attains a temperature of ~ 6000 K.

The disadvantages of a DCP are (i) the edges of anodes get consumed and require reshaping every couple of hours of continuous operation and (ii) therefore, it is difficult to automate the process.

Inductively coupled plasma source

Deriving its sustaining power from a high-frequency magnetic field, an inductively coupled plasma (ICP) source is a very high temperature (6000 K to 10000 K) excitation source that can efficiently desolvate, vaporise, excite and ionise atoms. A diagram of an ICP is shown in Fig. 14.74.

The basic set-up of an ICP consists of three concentric quartz tubes called, the outer, intermediate and inner loops. At the tip of the outer loop, having a diameter of about 25 mm, a copper coil is wound. Initially, Ar gas passes through the loop and a 27 MHz ac of power level 1 kW to 5 kW flows through the coil. The support gas stream, that enters through the intermediate loop, is then struck with a spark from a Tesla coil that seeds the stream with electrons. These seed electrons quickly interact with the magnetic field generated by the flowing current in the coil and gain sufficient energy to ionise Ar atoms by collisions. Owing

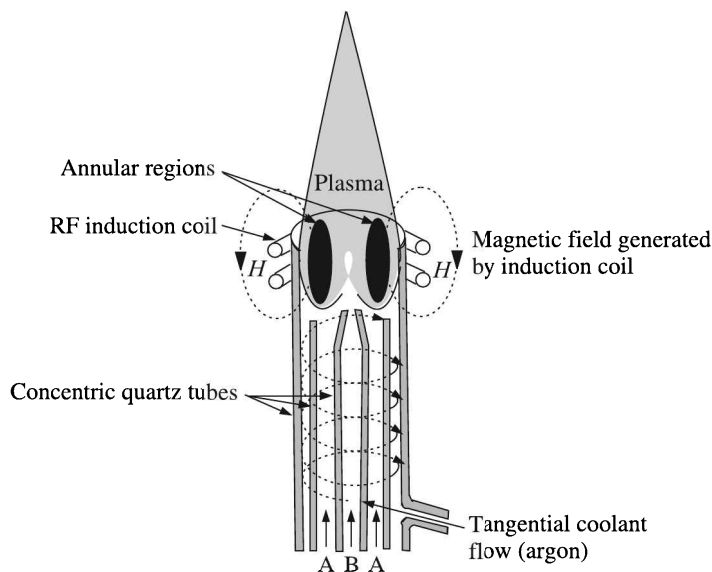


Fig. 14.74 Schematic diagram of ICP. A: support gas (argon), B: sample aerosol + argon.

to the presence of the magnetic field, generated cations as well as electrons move in circles perpendicular to the gas stream. Because it is ac, an instantaneous reversal of the direction of the current reverses the direction of the magnetic field. This makes the cations and electrons move in reverse directions and collide with more Ar atoms to produce further ionisation that releases an intense thermal energy. Thus, a flame-shaped plasma forms near the tip of the outer loop.

A tangential argon coolant flow through the outer loop is necessary to (i) cool the inside of quartz tube walls, and (ii) stabilise the plasma.

At a frequency of 27 MHz, the skin effect occurs which gives the ICP a toroidal (or annular) shape. The shape of the plasma lengthens the resident time of the sample to nearly 2 ms in the interior of the high-temperature zone and thus lowers the detection limit of many elements, the typical range being 1 to 100 ng/L.

Typical flow rates of Ar through the outer, intermediate and inner loops are 15 L/min, 0 to 1 L/min and 1 L/min respectively. At the high temperature of the ICP, no chemical bond survives, causing a complete atomisation of the analyte solution. The extreme temperature generates strong emission lines for most of the elements of the periodic table.

Laser-induced excitation sources

When the light from a pulsed high-energy laser, such as a ruby laser, is focussed on a small area (50 μm diameter) of the sample with the help of a microscope, it can cause dielectric breakdown and create a hot plasma. For solids, the laser also ablates material into the gas phase. The resulting emission is enhanced by using a cross-excitation with a spark discharge between two graphite electrodes as shown in Fig. 14.75. The energy of the laser-created plasma can atomise, excite and ionise analyte species, which can then be detected and quantified by atomic spectrophotometry.

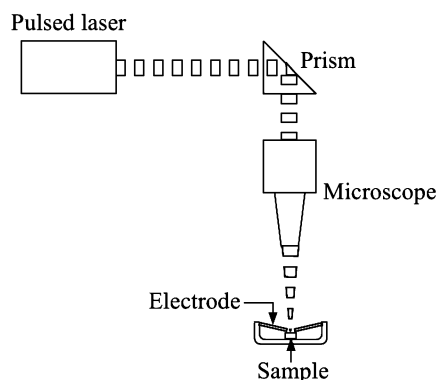


Fig. 14.75 Laser-induced excitation.

Atomic Emission Spectrophotometry

In atomic emission spectrophotometry (OES)²⁶, we need atoms promoted to a higher energy level through excitation by an appropriate source of supply of energy. The excited atoms spontaneously decay to their ground states, emitting radiation of different wavelengths (Fig. 14.76).

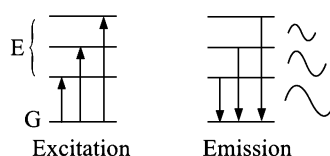


Fig. 14.76 Excitation of an atom and its spontaneous decay to the ground state. E: excited states, G: ground state.

Since the transitions are between distinct atomic energy levels, the emitted spectral lines are sharp. A multi-element sample may give rise to a very congested spectrum which may call for the deployment of a high resolution wavelength selector. However, all the elements in a sample are excited simultaneously and therefore, they can also be detected simultaneously by using a polychromator. This is the advantage of an OES. Schematic arrangements of OES and polychromator are shown in Figs. 14.77 and 14.78 respectively.

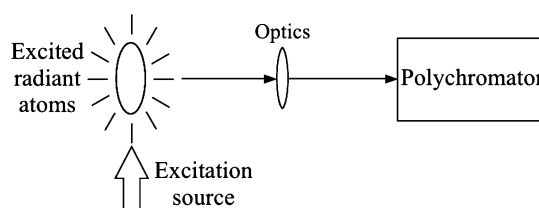


Fig. 14.77 Schematic arrangement of OES.

²⁶Optical emission spectrophotometry, the acronym AES being reserved for Auger electron spectroscopy. The atomic emission spectrophotometry is also called the flame emission spectrophotometry, FES.

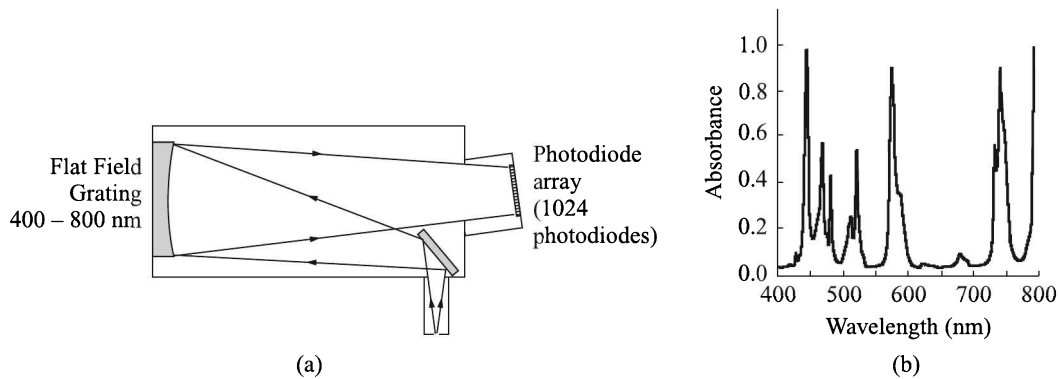


Fig. 14.78 (a) Polychromator, and (b) a recorded spectrum.

The polychromator readout device helps measure the radiant power or intensity of the selected radiation. The intensity is given by

$$I = F_E N_j h \nu \quad (14.53)$$

where I is the intensity of the selected radiation
 F_E is the Einstein transition probability
 h is the Planck's constant
 ν is the frequency of the selected radiation
 N_j is the number of atoms in the excited state j .
 N_j , in turn, is given by

$$N_j = N_0 \frac{P_j}{P_0} e^{-E_j/kT}$$

where N_0 is the number of ground state atoms per unit volume
 P_j is the statistical weight factor of the excited state
 P_0 is the statistical weight factor of the ground state
 E_j is the energy of the excited state
 k is the Boltzmann constant
 T is the absolute temperature.

Equation (14.53) shows that the intensity of the characteristic line is not only proportional to the concentration of the analyte (N_0) but also the ionisation cross-section (P_j/P_0) and other factors of the selected radiation. So, it is difficult to utilise the equation to figure out the concentration of the analyte from the measurement of the intensity of the selected radiation. What is done in practice is that a calibration curve of intensities vs. known concentrations of the analyte is drawn. The concentration of the unknown analyte is found from this calibration curve once its intensity is known.

Wavelengths used in the OES range from the upper part of the vacuum ultraviolet (160 nm) to the limit of the visible light (800 nm). Since borosilicate glass and oxygen in the air absorb light below 310 nm and 200 nm respectively, optical lenses and prisms are fabricated from quartz glass and optical paths are evacuated or filled with a non-absorbing gas such as argon.

OES is a versatile, rapid method for estimation of elements. It finds use in many fields which include pharmaceutical, food, polymer, pesticide and catalyst industries as well as in environmental studies and pollution monitoring of air and water.

Atomic Absorption Spectrophotometry

In atomic absorption spectrophotometry (AAS), light of right wavelength is impinged on a free, ground state atom. The atom absorbs the light to enter an excited state (Fig. 14.79). From this absorption study, one can identify the presence or absence of an element in a sample.

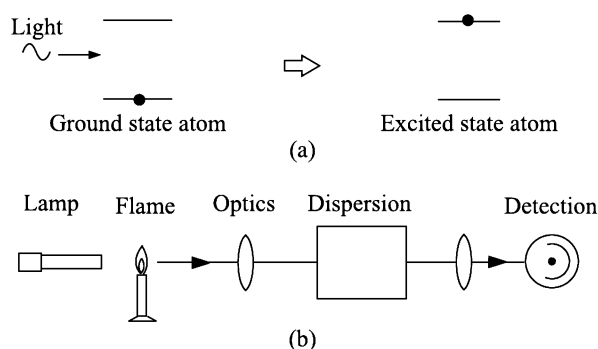


Fig. 14.79 Atomic absorption spectrometry: (a) excitation of atom by light absorption, and (b) spectrometry.

The interest in atomic absorption measurement lies in estimating at the resonant wavelength the amount of light that is absorbed by the cloud of atoms. Obviously, the more is the number of such atoms, the more is the absorption. Therefore, by measuring the strength of absorption, a quantitative determination of the element present can be made. Special light sources and careful selection of wavelength allow this quantitative determination of individual elements.

The required atom cloud is produced by supplying suitable thermal energy to the sample so that chemical compounds break to form free atoms. A solution of the sample is aspirated for this purpose on a flame aligned with the light beam. The ease and speed at which a precise determination can be made have made atomic absorption spectrometry one of the popular methods for estimation of metals. The schematic diagram of a dual-beam AAS is presented in Fig. 14.80.

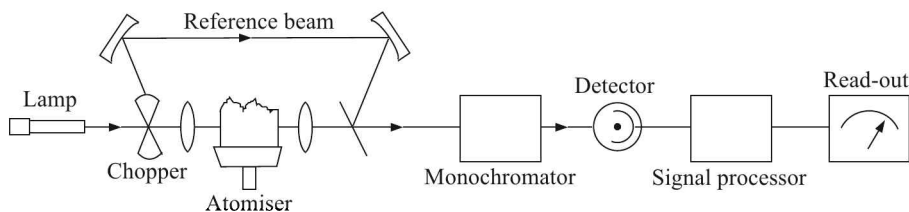


Fig. 14.80 Dual-beam atomic absorption spectrophotometer.

The whole instrumentation can be divided into four basic structural elements

1. Light source (hollow cathode lamp)
2. Atomiser
3. Monochromator
4. Detector and readout system

Light source

An atom absorbs light at specific wavelengths. The light source should produce those specific wavelengths at reasonable intensities. These goals are met with by the hollow cathode lamp whose construction is shown in Fig. 14.81.

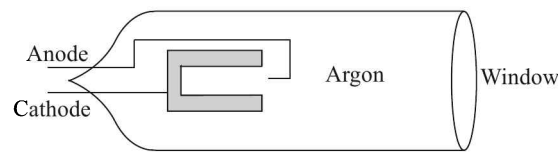


Fig. 14.81 Hollow cathode lamp.

A hollowed-out cylinder, made of the metal whose spectrum is to be produced, constitutes the cathode of the lamp. A glass envelope, filled with an inert gas such as argon or neon, encloses the anode and the cathode. The material of the window should be such that it will not absorb the required wavelengths.

The mechanism of emission from hollow cathode lamps is illustrated in Fig. 14.82. The applied voltage between the electrodes ionises fill gas atoms. Positively charged ions, thus generated, accelerate through the electrical field to collide with the cathode and sputter metal atoms. The sputtered metal atoms collide with further ions to be raised to excited states and thus eventually emit the required wavelengths.

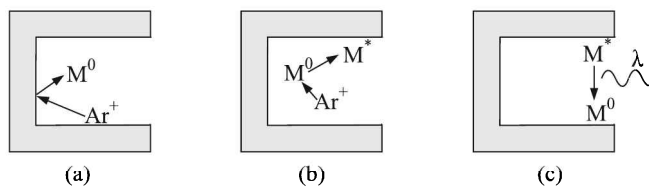


Fig. 14.82 Mechanism of emission from hollow cathode lamp: (a) sputtering, (b) excitation, and (c) emission. M^0 —sputtered metal atom, M^* —excited metal atom.

Hollow cathode lamps have a finite lifetime. The primary reasons are—(i) during sputtering some dislodged metal atoms get deposited elsewhere, (ii) some cathodes vaporise during use, (iii) fill gas atoms get adsorbed on the cathode surface, and (iv) some cathode materials slowly evolve occluded hydrogen when heated. Tantalum “getter” is attached to anode to adsorb hydrogen.

For volatile elements where low intensity and short lamp life is a problem, electrodeless discharge lamps (EDL) are helpful. In an EDL the element is sealed in a quartz capsule. The capsule is placed inside a ceramic cylinder on which an RF coil is wound (Fig. 14.83).

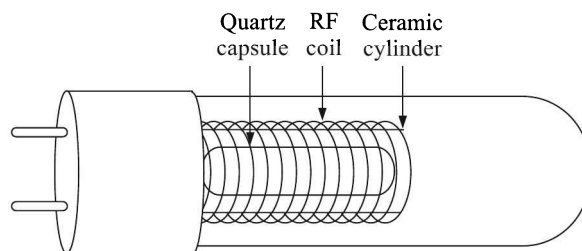


Fig. 14.83 Electrodeless discharge lamp.

The RF field of required power vaporises the element in the capsule to generate the emission. This is basically the induction heating of the element.

Light generated from the lamps is chopped to provide a means of selectively amplifying the light emitted from the source lamp and ignoring emission from the sample cell. Chopping can be done mechanically or by applying a pulsed power to the lamp.

Atomiser

The principle of atomic absorption requires light absorption by “free atoms”. A “free atom” means an atom which is not combined with other atoms. However, elements in the sample are invariably combined with other elements to form molecules. The combination must be broken by some means to free the atoms. This is called atomisation. Of all the methods described at the beginning of this section, the most popular method of atomisation used in AAS is the flame excitation source. Samples are heated to a high temperature by oxy-acetylene (or acetylene-nitrous oxide) flame so that molecules are converted into free atoms. The appearance of an actual nebuliser-cum-burner is shown in Fig. 14.84.

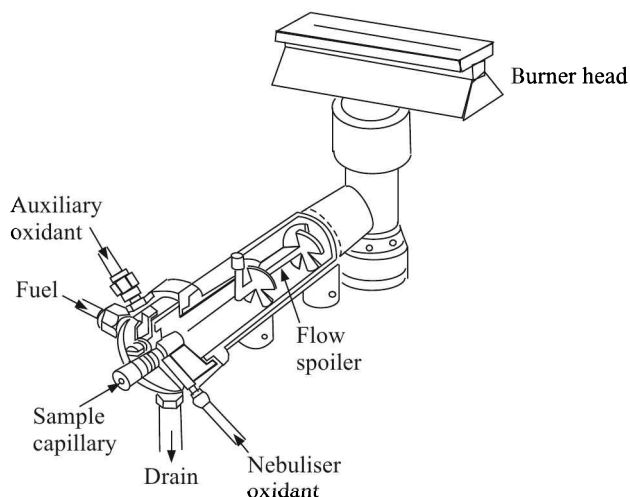


Fig. 14.84 Appearance of an actual nebuliser-cum-burner.

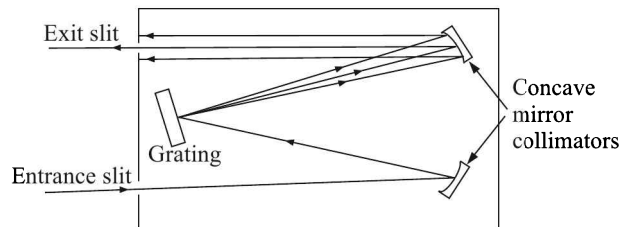
The alternative to the flame method is the electrothermal method, described earlier. The comparison of the flame atomisation method and electrothermal atomisation method is shown in Table 14.8.

Table 14.8 Comparison between flame and electrothermal methods of atomisation

| <i>Characteristic</i> | <i>Flame atomisation</i> | <i>Electrothermal atomisation</i> |
|-----------------------|----------------------------|-----------------------------------|
| Sensitivity | ppm level in the solution | ppb level in the solution |
| Sample volume | ~1 ml for one analysis | 5 to 50 μ l for one analysis |
| Atomising efficiency | about 10% | more than 90% |
| Shape of signal | plateau shape | peak shape |
| Repeatability | 0.5 to 1.0% of the reading | 2.0 to 5.0% of the reading |
| Time for analysis | 10 to 30 s for one sample | 2 to 5 min for one sample |

Monochromator

The monochromator is included as an important device of the optical system of an atomic absorption spectrophotometer. The function of this device is to separate the spectral line of interest from other spectral lines of different wavelengths emitted by the hollow cathode lamp. The desired spectral line is chosen with the preferred wavelength and bandwidth with the help of a grating (Fig. 14.85).

**Fig. 14.85** Grating monochromator.

A grating is a collection of closely spaced parallel slits on a surface. It can be of transmission- or reflective-type and can be designed for different wavelength regions. Generally, most of the instruments are equipped with two reflective-type gratings with a view to covering a wavelength range of 189 to 851 nm which is used in atomic absorption.

Detector and readout systems

The monochromator receives light from the hollow cathode lamp through the flame together with the light emitted from the flame. The reference beam helps to reject the signal arriving at the detector from the flame emission and accept only that coming from the hollow cathode lamp. The detector used almost universally is a photomultiplier tube whose current output corresponds to the intensity of the light falling on its photocathode. This feeds the amplifier and output device, which displays the measured signals.

The simplest output device consists of a moving-coil meter or a pen recorder displaying percentage transmission. Some instruments have a digital display, which provides a direct readout of absorbance values. This, with a provision for curve linearisation, forms the basis for displaying outputs directly in concentration terms, using standard solutions for calibrations. At present, most aspects of instrument control, operation, standardisation and data processing or storage are carried out by a microcomputer or microprocessor built-in into the atomic absorption unit or interfaced to it.

Sensitivity and limit of detection

Even though the Beer-Lambert law is followed in AAS, the absorbance is smaller than it should be owing to emissions by the flame, causing negative deviations from the Beer-Lambert law. The use of a light chopper helps minimise errors arising out of flame emission. Absorbance vs. sample concentration calibration curve is a way to correct for flame emission.

There are certain factors that decrease the sensitivity such as (i) excessive noise in the source, (ii) having a large number of reflecting surfaces, and (iii) turbulence in the optical path. Some ways of minimising noise is by providing a stable flame, using a low noise detection system, and by letting the electronics, gas regulators, chamber, burner reach equilibrium. AAS is a very sensitive analytical instrument that can measure in the ppb range and detect as low as 10^{-12} g/L.

Applications

Considered to be one of the best quantitative methods of analysis for metals, AAS has also been found useful in the analysis of plant materials, biological materials, food and beverages, chemical products, and in environmental studies.

Uses of AAS have led to major discoveries in the field of astrophysics. A considerable amount of our knowledge about the cosmos comes from our empirical understanding of spectral lines emitted by objects in outer space. Studying EMR wavelengths emitted by astronomical objects helped us know the chemical composition of celestial bodies, and their respective velocities. Also densities, temperatures, abundances of elements can be acquired by studying the intensities of spectral lines that are emitted. The Doppler shift of spectral lines toward red (red shift) provides evidence that the universe is expanding.

Atomic Fluorescence Spectrophotometry

In AFS, the radiation from an external source is impinged upon the free analyte atoms generated by an appropriate way. This energisation raises the atoms to an excited state wherefrom they return to the ground state through an emission of radiation. This emitted radiation is analysed, as in OES, to determine the concentration of the element in question.

So, this technique, in essence, incorporates aspects of both atomic absorption and atomic emission. Like AAS, ground state atoms are created and then they are excited by focussing a beam of light into the atomic cloud. But unlike AAS, the emission resulting from the decay of excited atoms is measured rather than light absorbed in the process. The excited atoms may not decay directly from the energy level they were excited to. There may be some internal loss of energy which brings them to a lower intermediate energy state from where they decay to the ground state to emit radiation (Fig. 14.86). This radiation is called 'fluorescence' which characterises an atom.

The intensity of this fluorescence increases with the atomic concentration, providing the basis for quantitative determination. The main advantage of fluorescence detection compared to absorption measurements is the greater sensitivity achievable because the fluorescence signal has a very low background noise.

The source lamp, which may be a hollow cathode lamp or a laser, is out of line (normally perpendicular) with the rest of the optical system so that the detector sees only the fluorescence in the flame and not the light from the lamp itself. Lamps are normally much brighter in AFS

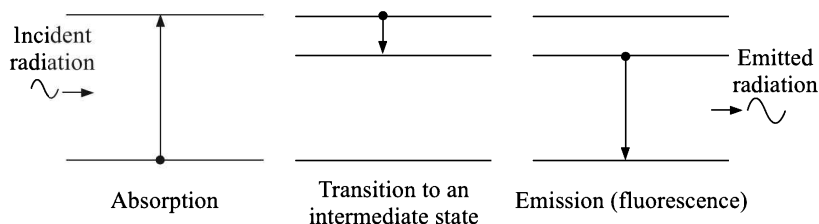


Fig. 14.86 Origin of fluorescence.

than in AAS in order to increase the degree of atomic excitation and provide high fluorescence sensitivities. Figure 14.87 illustrates the technique.

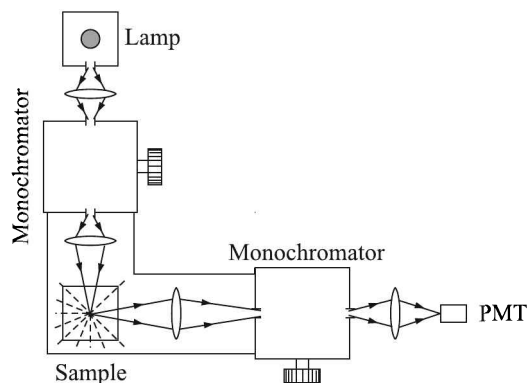


Fig. 14.87 AFS arrangement (seen from the top).

14.6 UV-visible Absorption Spectrophotometer

As the name implies, absorption studies are made in the UV-visible absorption spectrophotometry. Unlike the atomic spectrophotometry, here the sample is taken in the form of a solution and no atomisation is done. Which means, the absorption spectrum of the compound, rather than elements, of the analyte is studied. The ultraviolet (UV) region normally scanned is from 200 to 400 nm and the visible region, from 400 to 800 nm. The essential components of the spectrophotometer is illustrated in Fig. 14.88.

A chosen wavelength of light from a visible or UV light source is isolated with the help of a diffraction grating based monochromator. A half-silvered mirror, in turn, splits this monochromatic beam into two beams of equal intensity. The sample beam passes through a small transparent container (called 'cuvette') containing a solution in a transparent solvent of the compound being studied. The other beam, i.e. the reference beam, passes through an identical cuvette containing only the solvent. The intensities of these light beams are then measured by electronic detectors and compared. The intensity of the reference beam, which should have suffered little or no light absorption, is defined as I_R and the intensity of the sample beam, as I_S . In this way, the spectrometer automatically scans all the component wavelengths over a short period of time.

Absorbance vs. wavelength in the span of the wavelength studied is plotted. Different compounds may have very different absorbance maxima. In order that the detectors receive

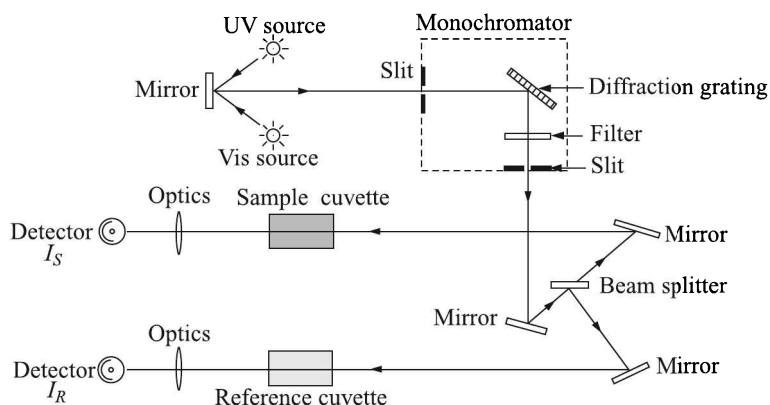


Fig. 14.88 UV-visible absorption spectrophotometer components.

significant amount of light, intensely absorbing compounds must be examined in dilute solution. This necessitates the use of transparent, i.e. non-absorbing solvents. The most commonly used solvents are water, ethanol, hexane and cyclohexane. Solvents of compounds having double or triple bonds, or heavy atoms (e.g. S, Br and I) are generally avoided. Because the absorbance of a sample is proportional to its molar concentration in the sample cuvette, a corrected absorption value known as the *molar absorptivity* is used when comparing the spectra of different compounds. We know that the molar absorptivity ϵ is defined as

$$\epsilon = \frac{A}{bc} \quad (14.54)$$

where A is the absorbance

b is the length (in cm) of the light path through the cuvette

c is the sample concentration in mole/litre.

For example, a solution of 0.249 mg of the unsaturated aldehyde in 95% ethanol (1.42×10^{-5} M) was placed in a 1 cm cuvette for measurement. The resulting absorbance spectrum is shown in Fig. 14.89. Using Eq. (14.54), $\epsilon = 36600$ for the 395 nm peak, and 14000 for the 255 nm peak.

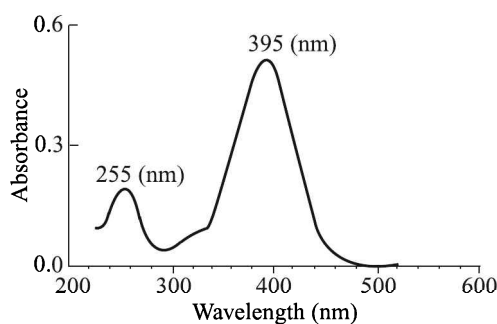


Fig. 14.89 Absorbance spectrum of unsaturated aldehyde in 95% ethanol.

The molar absorptivity can be very high (> 10000) for strongly absorbing compounds while for weakly absorbing ones, ϵ may be as low as 10 to 100.

As regards the instrumentation of the UV-visible spectrophotometer, most of the components, except the sources, are familiar to us. The visible light source is mostly the incandescent tungsten filament lamp. The ultraviolet source is, however, mostly the deuterium discharge lamp.

Deuterium discharge lamp

The schematic diagram of a deuterium discharge lamp is shown in Fig. 14.90.

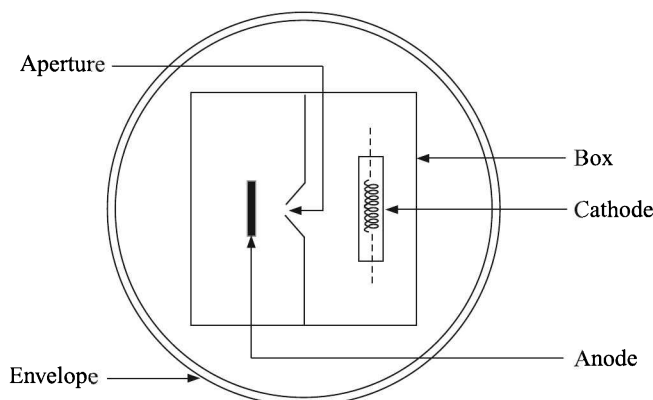


Fig. 14.90 Cross-sectional top view of the deuterium discharge lamp.

A deuterium discharge lamp uses an oxide-coated cathode having a tungsten filament. The directly heated cathode produces an abundance of electrons. The main feature of the lamp is an aperture of 0.6 to 1.5 mm between the cathode and the anode where an intense ball of radiation is created. The aperture is highly polished and shaped like a parabolic reflector to increase the intensity of radiation. The cathode, anode and the aperture are housed within a high purity nickel box to ensure that a uniform discharge occurs only between the cathode and anode. The lamp is filled with deuterium at a low pressure of 0.2 to 0.5 Torr and a low dc voltage of about 40 V is applied between the electrodes though initially about 350 V is to be applied to start the discharge by the heated cathode. Once the discharge starts, the lamp resistance goes down. Therefore, a constant current supply is required to keep the current within about 300 mA. Because normal glass absorbs UV radiation, quartz, UV-glass or magnesium fluoride is used as the envelope (Fig. 14.90).

The deuterium lamp emits radiation between 112 and 900 nm. But its continuous spectrum range is 180 to 370 nm.

14.7 Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a powerful and theoretically complex analytical tool that helps deduce the physical, chemical, electronic and structural information about a molecule. It is the only technique that can provide detailed information

on the exact three-dimensional structure of biological molecules in solution. Also, NMR is one of the techniques that have been used to build quantum computers.

Principle

Subatomic particles (electrons, protons and neutrons) are thought to have spin motion about their axes. In many atoms (such as ^{12}C) opposite spins are paired, such that the nucleus of the atom has no overall spin. However, in some atoms (such as ^1H and ^{13}C) the nucleus does possess an overall spin. The rules for determining the net spin of a nucleus are given in Table 14.9.

Table 14.9 Rules for determining the net spin of a nucleus

| <i>Number of protons</i> | <i>Number of neutrons</i> | <i>Net spin</i> |
|--------------------------|---------------------------|--|
| Even | Even | Zero |
| Odd | Odd | Integral (i.e. 1, 2, 3, ...) |
| Even | Odd | Half-integral (i.e. $\frac{1}{2}$, $\frac{3}{2}$, $\frac{5}{2}$, ...) |
| Odd | Even | Ditto |

The overall spin J is important. A nucleus of spin J will have $2J + 1$ possible orientations such as a nucleus with spin $\frac{1}{2}$ will have 2 possible orientations, namely, spin $\frac{1}{2}$ and spin $-\frac{1}{2}$. In the absence of an external magnetic field, these orientations are of equal energy. If a magnetic field is applied, then the energy levels split. Each level is given a 'magnetic quantum number', m .

The positively-charged spinning nucleus generates a small magnetic field. The corresponding magnetic moment μ is given by

$$\mu = \frac{gJh}{2\pi}$$

where the constant g is called the *gyromagnetic ratio* and is a fundamental nuclear constant which has a different value for every nucleus and h is Planck's constant. The energy of a particular energy level is given by

$$E = -\frac{gh}{2\pi}mB$$

where B is the strength of the magnetic field *at the nucleus*²⁷. The difference in energy between two consecutive levels (i.e. $m = 1$ and 2, say) is

$$\Delta E = \frac{ghB}{2\pi} \quad (14.55)$$

This energy is called the *transition energy*. From Eq. (14.55) we find that

- (a) If the magnetic field B is increased, so is ΔE
- (b) If the gyromagnetic ratio of a nucleus is large, then ΔE is correspondingly large

²⁷It will be somewhat different from the applied field because of shielding by the surrounding electrons which will also generate a magnetic field.

The nucleus of an atom possesses magnetic moment owing to the spin of protons. Therefore, when a sample is placed between the poles of a strong magnet, its nuclei precess²⁸ about respective axes. The motion is somewhat akin to that of a rotating top when its axis of rotation is not vertical [Fig. 14.91(a)].

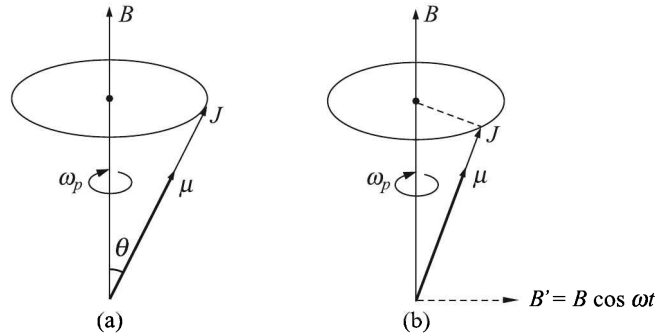


Fig. 14.91 (a) Precession of nucleus with magnetic moment μ and angular momentum J in a magnetic field B , and (b) change of the angle of precession by an RF field.

The frequency of precession is termed the *Larmor frequency*, which is identical to the transition frequency. The potential energy of the precessing nucleus is given by

$$E = -\mu B \cos \theta$$

where θ is the angle between the direction of the applied field and the axis of nuclear rotation.

If a radio frequency electric field is applied to the sample perpendicular to the magnetic field, there may be a resonance between the magnetic fields produced by the RF and the nuclear precession frequency. If energy is absorbed by the nucleus, then the angle of precession θ will change [Fig. 14.91(b)]. For a nucleus of spin $\frac{1}{2}$, absorption of radiation “flips” the magnetic moment so that it *opposes* the applied field (the higher energy state). The phenomenon is known as *nuclear magnetic resonance* (NMR). An increased absorption of the RF by the sample results at the resonance.

NMR Set-up

In continuous-wave NMR, the RF is maintained at a suitable value while the magnetic field strength is varied by changing the current through the electromagnet. A careful selection of the RF in relation to the magnetic field strength can make the device specific for a particular element. We have seen in Section 13.1 how it can be utilised to measure moisture content of a sample.

But now in FT-NMR an RF pulse containing all the frequencies is sent (Fig. 14.92). At the top left of the schematic representation, the magnet of the NMR spectrometer can be seen. In most cases nowadays, it is a magnet having windings of a superconducting wire so that very high fields can be produced. The magnet produces the B_0 field necessary for the NMR experiments. Immediately within the bore of the magnet are the shim coils for homogenising the B_0 field. Within the shim coils is the probe. The probe contains the RF coils for producing

²⁸Called the *Larmor precession*.

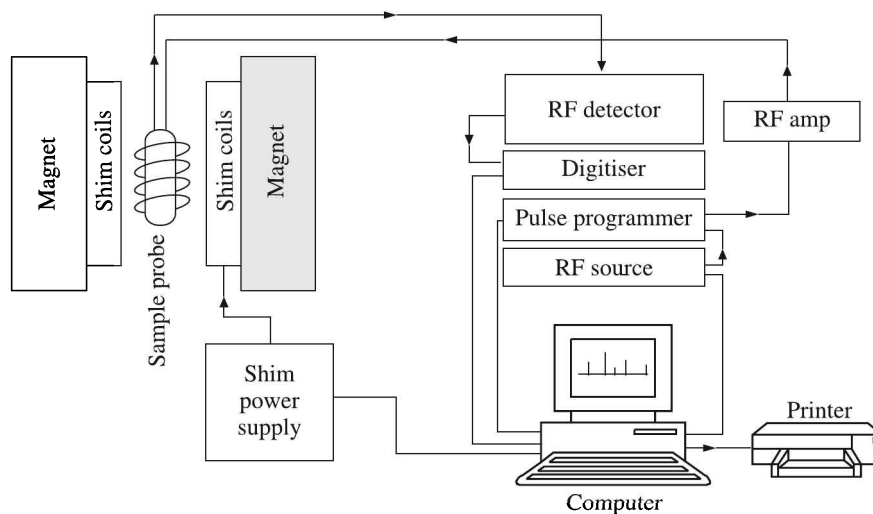


Fig. 14.92 A nuclear magnetic resonance set-up.

the B_1 magnetic field necessary to rotate the spins by 90° or 180° . The RF coil also detects the signal from the spins within the sample. The sample is positioned within the RF coil of the probe. Some probes also contain a set of gradient coils (not shown). These coils produce a gradient in B_0 along the x -, y -, or z -axis. Gradient coils are used for gradient enhanced spectroscopy, diffusion, and NMR microscopy experiments.

The RF generator is the RF frequency source and pulse programmer. The source produces a sine wave of the desired frequency. The pulse programmer sets the width, and in some cases the shape of the RF pulses. The RF amplifier increases the pulse power from milliwatts to tens or hundreds of watts. We will now discuss the magnet, the shim coils, the sample probe, the RF coils, and the RF detector in little more detail.

Magnet

The magnet is one of the most expensive components of the NMR spectrometer system. As already mentioned, most magnets are of the superconducting type. The superconducting wire has a resistance approximately equal to zero when it is cooled to a temperature of -268.95°C (or 4.2 K) by immersing it in liquid helium. Once current is caused to flow in the coil it will continue to do so as long as the coil is kept at liquid helium temperature. Of course, some losses do occur over time due to the infinitesimal resistance of the coil. These losses are on the order of a ppm of the main magnetic field per year.

The length of superconducting wire in the magnet is typically several miles. The wire is wound into a multi-turn solenoid or coil. The coil and liquid helium are kept in a large Dewar. This Dewar is surrounded by a liquid nitrogen (77.4 K) Dewar, which acts as a thermal buffer between the room temperature and the liquid helium.

In order to produce a high resolution NMR spectrum of a sample, we need to have a temporally constant and spatially homogeneous magnetic field. The field strength of the magnet might vary over time due to ageing, movement of magnetic objects near it, and temperature fluctuations. The field lock can compensate for these variations.

The field lock is a separate NMR spectrometer within the spectrometer. This spectrometer is typically tuned to the deuterium NMR resonance frequency. It constantly monitors the resonance frequency of the deuterium signal and makes minor changes in the B_0 magnetic field to keep the resonance frequency constant.

Shim coils

Shim coils correct minor spatial inhomogeneities in the B_0 magnetic field. These inhomogeneities could be caused by the magnet design, materials in the probe, variations in the thickness of the sample tube, sample permeability, and ferromagnetic materials around the magnet. A shim coil is designed to create a small magnetic field which will oppose and cancel out an inhomogeneity in the B_0 magnetic field. Because these variations may exist in a variety of functional forms (linear, parabolic, etc.), many shim coils are needed which can create a variety of opposing fields. By passing the appropriate amount of current through each coil, a homogeneous B_0 magnetic field can be achieved. On most spectrometers, the shim coils are controlled by a computer algorithm that finds the best shim value by maximising the lock signal.

Sample probe

The sample probe holds the sample, sends RF energy into it, and detects the signal emanating from it. Apart from the RF coil, it consists of sample spinner, temperature controller and in some, gradient coils. The RF coil is described below.

The sample spinner rotates the sample tube about its axis in order to achieve a narrower spectral line-width. By spinning the sample about its z -axis, inhomogeneities in the x - and y -directions of the magnetic field are averaged out to yield narrower NMR line-width.

Sometimes it becomes necessary to examine properties of samples as a function of temperature. For this purpose many sample probes have accessories to control the temperature of the sample above and below room temperature by passing warm or cool air or nitrogen from a thermostat over it.

RF coils

RF coils create the B_1 field which rotates the net magnetisation in a pulse sequence. They also detect the transverse magnetisation as it precesses in the xy -plane. Therefore, they should resonate at the Larmor frequency of the nucleus being examined. RF coils incorporate an inductor and a set of capacitive elements. Most RF coils on NMR spectrometers are of the saddle coil design (Fig. 14.93) and there may be one or more RF coils in a probe.

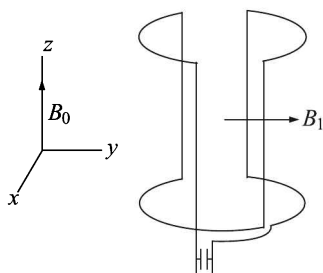


Fig. 14.93 Saddle coil design of the RF coil.

The resonant frequency ν of an RF coil is determined by the inductance L and capacitance C of the tank circuit as

$$\nu = \frac{1}{2\pi\sqrt{LC}} \quad (14.56)$$

However, conductivity and dielectric constant of a sample when placed in an RF coil, affect its resonance frequency. If this frequency is different from the resonance frequency of the nucleus being studied, the coil will not efficiently set up the B_1 field nor efficiently detect the signal from the sample. Furthermore, because of poor detection of the signal, signal-to-noise ratio will be poor.

The B_1 field of the RF coil must not only be perpendicular to the B_0 magnetic field but also has to be homogeneous over the volume of the sample. Or else, spins will be rotated by a distribution of rotation angles to generate strange spectra.

Detector

The detection is carried out by what is called a *quadrature detector*. The quadrature detector is a device in which inputs are frequencies ν and ν_0 while outputs are transverse magnetisation components $M_{x'}$ and $M_{y'}$. Without going into further details, it can be said that the quadrature detector turns the laboratory system of coordinate into a rotating frame of reference.

Fourier Transform Nuclear Magnetic Resonance

Throughout its first few decades from its invention by F Bloch and EM Purcell in 1946, NMR practice utilised continuous-wave (CW) spectroscopy.

The CW technique probes each frequency individually in succession, which is bugged by poor signal-to-noise ratio. Fortunately for NMR in general, signal-to-noise ratio (S/N) can be improved by signal averaging. One way of increasing sensitivity is to record many spectra, and then adding them together. Because noise is random, it adds as the square root of the number of spectra recorded. For example, if one hundred spectra of a compound were recorded and summed, then the noise would increase by a factor of ten, but the signal would increase in magnitude by a factor of one hundred, giving a large increase in sensitivity. However, if this is done using a continuous wave instrument, the time needed to collect the spectra is very large (one scan takes 2 to 8 minutes).

Fourier transform NMR spectroscopy (FT-NMR) can hasten a scan by allowing a range of frequencies to be probed at once. Pioneered by RR Ernst²⁹, this technique has been made more practical with the development of computer control that can not only create an array of frequencies at once to produce a spectrum but also perform the mathematical transformation of the data from the time domain to the frequency domain.

The FT-NMR works by irradiating the sample (held in a static, external magnetic field) with a short square pulse of RF containing all the frequencies in the range of interest because the Fourier decomposition of a square wave contains contributions from all frequencies. The polarised magnets of the nuclei then begin to spin together, creating an RF that is detectable. However, they ultimately decay to their ground state (in the magnetic field) of having a net polarisation vector that aligns with the field. This decay is known as the *free induction decay*

²⁹Nobel Prize winner in Chemistry (1991).

(FID). This time-dependent pattern can be converted into a frequency-dependent pattern of nuclear resonances using Fourier transformation, revealing the NMR spectrum.

Precautions

It is necessary to take some safety measures before using an NMR spectrometer. They are related to the use of strong magnetic fields and cryogenic liquids.

Caution must be taken to keep all ferromagnetic items away from the magnet because high field magnets can pick up and pull them into the magnet. The kinetic energy of an object being sucked into a magnet can smash a Dewar or an electrical connector on a probe. Even a small metal piece can get sucked into the magnet and destroy the homogeneity of the magnetic field.

A person having an implanted pacemaker walking through a strong magnetic field can induce currents in the pacemaker circuitry which will cause it to fail and possibly cause death. Magnetic fields of approximately 50 gauss will erase credit cards and magnetic storage media.

The liquid nitrogen and liquid helium used in NMR spectrometers can cause frostbite, if precaution is not taken while filling the magnet. Also, if the magnet quenches, or stops being a superconductor for some system failure, it will rapidly boil off all its cryogenics to produce nitrogen and helium gases which can cause suffocation in a confined space.

14.8 Electron Spin Resonance Spectrometer

A great number of materials contain unpaired electron spins which give rise to their paramagnetic properties. This non-pairing of electron spins may occur owing to

1. Electrons in unfilled conduction bands or trapped in radiation damaged sites
2. Existence of free radicals
3. Existence of various transition ions, bi-radicals, triplet states in the lattice
4. Impurities in semiconductors, etc

Electron spin resonance (ESR) or electron paramagnetic resonance (EPR)³⁰ is a spectroscopic technique that detects chemical species that have unpaired electrons.

Any spectroscopic technique basically identifies the chemical species being studied. In cases where two or more paramagnetic species coexist, EPR helps one observe simultaneously the spectral lines arising from each. Often definitive identification of the individual species is realised solely from the analysis of the EPR spectrum. Furthermore, EPR spectroscopy is capable of providing molecular structural details inaccessible by any other analytical tool.

After over half a century of its development, EPR remains one of the most sensitive tools to study the chemical (i.e. electronic) nature of matter, because it is able to detect and identify spins with concentrations in the 10^{-9} region. Of course, due to line-width effects, limits of detection are usually somewhat less than this.

We may note here that although in principle it is possible to determine absolute concentrations of spin species using EPR, relative concentrations are usually determined because of the complexity of the procedure involved.

³⁰EPR is preferred since the method comprises other than *spin only* systems. On the other hand, ESR is a widely used term that cannot be dispensed with.

Principle

Apart from its orbital motion, every electron has a spin motion which is quantised and indicated by the quantum number s . This gives rise to a quantised magnetic moment the components of which are indicated by magnetic quantum numbers $m_s = +\frac{1}{2}$ and $m_s = -\frac{1}{2}$. In the presence of an external magnetic field with strength B_0 , the electron's magnetic moment aligns itself either parallel ($m_s = -\frac{1}{2}$) or antiparallel ($m_s = +\frac{1}{2}$) to the field. Each alignment has a specific energy, the parallel alignment corresponding to the lower energy state. The separation between it and the upper state is given by

$$\Delta E = g_e \mu_B B_0 \quad (14.57)$$

where g_e is the Landé g -factor³¹ and μ_B is the Bohr magneton. The Bohr magneton is a physical constant of magnetic moment defined by

$$\mu_B = \frac{eh}{4\pi m_e} \quad (14.58)$$

where h is the Planck constant

e is the electronic charge

m_e is the electron mass

The value of Bohr magneton in SI units is 9.274×10^{-24} joule/tesla. Equation (14.57) implies that the splitting of the energy levels is directly proportional to the strength of the magnetic field, as shown in Fig. 14.94.

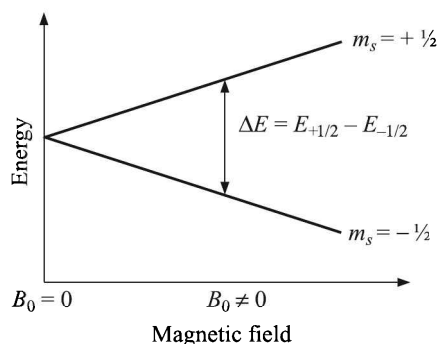


Fig. 14.94 Splitting of energy levels on application of magnetic field.

An unpaired electron can move between the two energy levels by either absorbing or emitting electromagnetic radiation of energy

$$\varepsilon = h\nu \quad (14.59)$$

such that the resonance condition,

$$\varepsilon = \Delta E \quad (14.60)$$

³¹All fundamental particles with spin are characterised by a magnetic moment and a g -factor. The g -factor is a proportionality constant between the magnetic dipole moment and the angular momentum.

is satisfied. Combining Eqs. (14.59) and (14.57) we get the fundamental equation of EPR spectroscopy:

$$h\nu = g_e\mu_B B_0 \quad (14.61)$$

By application of a strong magnetic field B_0 to the material containing paramagnetic species, the individual magnetic moment arising via the electron spin of the unpaired electron can be oriented either parallel or anti-parallel to the applied field. This creates distinct energy levels for the unpaired electrons, making it possible for net absorption of electromagnetic radiation (in the form of microwaves) to occur. The situation referred to as the resonance condition takes place when the magnetic field and the microwave frequency are “just right” (i.e., the energy of the microwaves corresponds to the energy difference ΔE of the pair of involved spin states).

EPR Absorption and Dispersion

EPR spectra of a sample can be generated by either varying the incident microwave frequency while holding the magnetic field constant, or vice versa. However, in practice, the frequency is kept fixed because variation of microwave frequency over a wide range poses a few problems.

Therefore, the sample, which is a collection of paramagnetic centres, such as free radicals, is exposed to microwaves at a fixed frequency. By increasing B_0 , the gap between the $m_s = +1/2$ and $m_s = -1/2$ energy states is widened until it matches the energy of the microwaves, as represented by the double-arrow in Fig. 14.94. At this point the unpaired electrons can move between their two spin states.

The population of paramagnetic centres in states at thermodynamic equilibrium is governed by the Maxwell-Boltzmann distribution as

$$\begin{aligned} \frac{n_{\text{up}}}{n_{\text{lo}}} &= \exp\left(-\frac{E_{\text{up}} - E_{\text{lo}}}{kT}\right) = \exp\left(-\frac{\Delta E}{kT}\right) \\ &= \exp\left(-\frac{h\nu}{kT}\right) \end{aligned} \quad (14.62)$$

where n_{up} is the number of paramagnetic centres having the upper energy
 n_{lo} is the number of paramagnetic centres having the lower energy
 E_{up} is the energy of paramagnetic centres having the upper energy
 E_{lo} is the energy of paramagnetic centres having the lower energy
 k is the Boltzmann constant
 h is the Planck constant
 ν is the frequency of absorption
 T is the absolute temperature of the sample

Consider, an incident X-band microwave frequency ($\nu = 9.75$ GHz) at room temperature (298 K). Then from Eq. (14.62),

$$\begin{aligned} \frac{n_{\text{up}}}{n_{\text{lo}}} &= \exp\left[\frac{(6.63 \times 10^{-34})(9.75 \times 10^9)}{(1.38 \times 10^{-23})(298)}\right] \\ &= 0.9984 \end{aligned}$$

The result indicates that $n_{\text{up}} < n_{\text{lo}}$, i.e. the upper level is less populated than the lower one. That, in turn, indicates that the lower→upper transition is more likely. Which is why there is a net absorption of energy, and it is this absorption which is monitored and converted into a spectrum.

Equation (14.61) permits a large combination of frequency and magnetic field values, but the great majority of EPR measurements are made with microwaves in the 9 GHz to 10 GHz region, with fields corresponding to about 0.35 tesla.

Example 14.3

Consider the case of a free electron, which has $g_e = 2.0023$. Find the magnetic field necessary for the EPR to occur when the applied microwave frequency is 9.75 GHz.

Solution

Given: $\nu = 9.75 \times 10^9 \text{ s}^{-1}$ and $g_e = 2.0023$. We know, $h = 6.63 \times 10^{-34} \text{ J-s}$ and $\mu_B = 9.274 \times 10^{-24} \text{ J-T}^{-1}$. Therefore, from Eq. (14.61)

$$\begin{aligned} B &= \frac{h\nu}{g_e\mu_B} \\ &= \frac{(6.63 \times 10^{-34})(9.75 \times 10^9)}{(2.0023)(9.274 \times 10^{-24})} \\ &= 0.3481 \text{ T} \end{aligned}$$

The appearance of the EPR spectrum corresponding to Example 14.3 is shown in Fig. 14.95 in two forms. The first form shows the actual absorbance while the second one shows its first derivative. It has been observed that though the noise level increases with the derivatives of spectra, they offer more resolution where peaks overlap. Therefore, in practice the second form is reported in most of the measurements. This is called the *derivative spectroscopy*³².

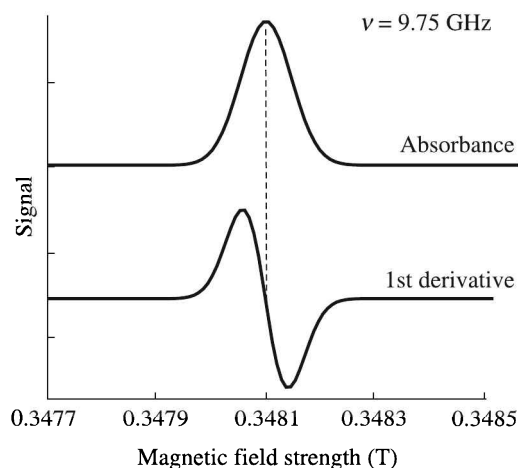


Fig. 14.95 The absorbance and first derivative of EPR spectrum of a free electron.

³²For a good discussion on derivative spectroscopy, see *Manual of spectrofluorometric and spectrophotometric derivative experiments*, Section II, A M Gillespie, CRC Press (1993).

Also, we note that electromagnetic radiation of a much higher frequency is required to bring about a spin resonance with an electron than a nucleus. This is apparent from the resonance equation in NMR spectroscopy which is given by

$$h\nu = g_N \mu_N B_0 \quad (14.63)$$

where

$$\mu_N = \frac{eh}{4\pi m_N}$$

Since, $m_N \gg m_e$, Eq. (14.63) suggests that ν for NMR is much lower than that for EPR. For example, we have seen that the spin resonance for electron occurs at 9.75 GHz for a field of 0.3481 T. For the same field, it is easy to calculate that the proton (i.e. ^1H nucleus, $m_N \sim 1840m_e$) resonance occurs only at about 14.85 MHz.

The g-factor and the internal structure of the paramagnetic centre

The value of g_e , as obtained experimentally through Eq. (14.61), can give information about the electronic structure of the paramagnetic centre.

We know that an unpaired electron spin energy level is split by the magnetic field acting upon it. This magnetic field comprises the applied magnetic field B_0 as well as the local magnetic field to which the unpaired electron's spin magnetic moment is very sensitive. These local magnetic fields often arise from the nuclear magnetic moments of various nuclei that may be present within the bulk medium. Examples of such nuclei are interstitial atoms (or ions) within a crystal or glass matrix, nuclei (such as nitrogen) within the molecular structure that also contains the unpaired electron, and so on. The effective magnetic field B_{eff} acting on the electron, can be written as

$$B_{\text{eff}} = B_0(1 - \sigma) \quad (14.64)$$

where σ , which can be positive or negative, represents the contribution from the local field. Combining Eqs. (14.61) and (14.64), we get

$$h\nu = g_e \mu_B B_0(1 - \sigma) \equiv g \mu_B B_0 \quad (14.65)$$

where

$$g = g_e(1 - \sigma)$$

Equation (14.65) shows that by measuring B_0 and ν in an EPR experiment, actually g is determined. If $g \neq g_e$, it implies that the ratio of the unpaired electron's spin magnetic moment to its angular momentum differs from the free electron value. An electron's spin magnetic moment is constant, having approximately the value of the Bohr magneton. Any change therefrom is an indication that the electron must have gained or lost angular momentum through spin-orbit coupling.

The mechanisms of spin-orbit coupling being well understood, the magnitude of the change in the value of g_e gives information about the nature of the atomic or molecular orbital containing the unpaired electron. Thus, EPR provides a unique means of studying the internal structures in great detail.

Information from hyperfine coupling

It may be thought that all EPR spectra should consist of a single line because it originates from the change in the spin state of an unpaired electron. But in practice, the spectra are

multi-lined because of interaction between the magnetic moment of the unpaired electron and those generated by nearby nuclear spins. This interaction gives rise to additional allowed energy states that produce a multi-lined spectrum. The effect is known as *hyperfine coupling*.

The spacing between the EPR spectral lines indicates the degree of hyperfine coupling for multi-lined spectra. The hyperfine coupling constant of a nucleus is directly proportional to, or, in the simplest cases, equal to the spectral line spacing.

In many cases, the isotropic hyperfine splitting pattern for a radical freely tumbling in a solution (isotropic system) can be predicted. One such case is, for a radical having M equivalent nuclei, each with a spin of I , the number of EPR lines expected is $2MI + 1$. For example, the methyl radical, CH_3 , has three ^1H nuclei each with $I = 1/2$. So the number of expected lines is

$$2MI + 1 = 2 \times 3 \times \frac{1}{2} + 1 = 4$$

Figure 14.96 shows the corresponding simulated first derivative spectrum.

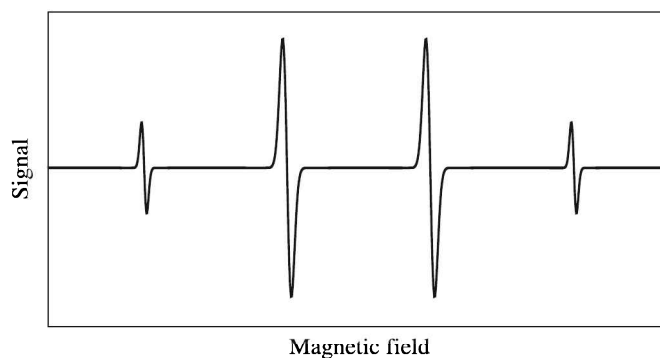


Fig. 14.96 Simulated first derivative EPR spectrum of CH_3 radical.

While it is not very difficult to predict the number of lines in a radical's EPR spectrum, the converse problem of interpreting a complex multi-line EPR spectrum and assigning the various spacings to specific nuclei, requires a good understanding of the subject.

EPR Instrumentation

The EPR instrumentation is very similar to that of the NMR. The differences are:

1. Because of the high frequency involved, a resonant cavity is necessary to contain the sample.
2. To minimise noise, a lock-in amplifier is necessary to amplify the signal.
3. An NMR is necessary to measure the magnetic field precisely.

14.9 X-ray Methods

Discovered in 1895, X-rays are electromagnetic radiation with typical photon energies in the range of 100 eV to 100 keV. They occupy the position between γ -rays and the ultraviolet of the electromagnetic spectrum.

X-rays primarily interact with electrons in atoms. When X-ray photons collide with electrons, some photons from the incident beam are deflected away from their original direction of travel. If the wavelength of these scattered X-rays do not change (that is, X-ray photons do not lose any energy), only the momentum is transferred in the scattering process and the process is called *Thompson scattering*. On the other hand, if due to scattering X-rays transfer some of their energy to the electrons and the scattered X-rays have different wavelength than the incident X-rays, the process is called *Compton scattering*.

Normally, X-ray diffraction studies are based on the Thomson scattering of X-rays as the scattered X-rays carry information about the electron distribution in materials. It has been use in the main areas shown in Table 14.10.

Table 14.10 Main areas of use of Thomson scattering of X-rays

| <i>Area</i> | <i>Use</i> |
|-----------------------------|---|
| Powder diffraction | <ol style="list-style-type: none"> 1. To identify unknown crystalline materials by comparing the diffraction data against a database maintained by the International Centre for Diffraction Data³³. Each crystalline solid has its unique characteristic X-ray powder diffraction pattern which may be used as a <i>fingerprnt</i> for its identification. X-ray diffraction is one of the most important non-destructive characterisation tools used in the materials science. 2. Powder diffraction is also used to study the grain size and strains in crystalline materials. |
| Single crystal diffraction | To determine the structure—i.e. how the atoms pack together in the crystalline state and what the interatomic distance and angle are etc.—of crystalline materials ranging from inorganic solids to complex macromolecules such as DNA. |
| High-resolution diffraction | To characterise epitaxial thin films with regard to their thickness, crystallographic structure and strain. |
| Pole figure analysis | To determine the distribution of orientation of crystals in a crystalline thin-film sample. |

Thomson scattering of monochromatic X-rays is also used to study materials that *do not have a long range order* through methods shown in Table 14.11.

Table 14.11 Other areas of use of Thomson scattering of X-rays

| <i>Method</i> | <i>Use</i> |
|-------------------------------------|--|
| Small angle X-ray scattering (SAXS) | To measure scattering intensity at angles 2θ close to 0° to probe structures in the nm to μm range. |
| X-ray reflectivity | To determine surface roughness, thickness and density of very thin—like monolayer or multilayer—films. |

With the advancement of electronics, it is now possible to study the energy and angle of scattered X-rays that change wavelength. These measurements are useful to probe the electronic band structure of materials by studying

³³Its products and services can be found at www.icdd.com.

1. Compton scattering
2. Raman scattering
3. Resonant inelastic X-ray scattering (RIXS)

We will be concerned with the X-ray diffraction studies based on the Thomson scattering of X-rays from *powdered substances that have a long-range order*. This study can be made by four different methods, namely

1. X-ray diffraction (Bragg reflection) spectrometry
2. Auger emission spectroscopy (AES)
3. X-ray fluorescence studies (XFS)
4. Electron spectroscopy for chemical analysis (ESCA)

But before discussing these methods, let us study how X-rays are produced and how they are detected.

X-ray Generation

We know, transitions of electrons in atoms from higher energy states to lower ones produce electromagnetic radiation. An electromagnetic radiation can also be created by accelerating electrons. For example, oscillation of electrons in an antenna creates radio waves. Also, decelerating (negatively accelerating) electrons can generate EM radiation. When an accelerated electron hits an anode, it generates EM radiation by two processes:

1. Bremsstrahlung³⁴
2. Electronic transitions between shells

Bremsstrahlung. The accelerated electron collides with atoms and slows down, creating radiation of a continuous distribution of wavelengths. The continuous wavelength from a cutoff λ_0 over a range is generated owing to successive collisions of impinging electrons.

Duane and Hunt experimentally determined in 1915 that the relation between λ_0 and excitation voltage V is given by

$$\lambda_0 \propto \frac{1}{V}$$

This empirical law was later interpreted by applying Bohr's theory of spectral lines. The collision of cathode electrons with the target atom may be considered as an initial state made up of a neutral atom and a free electron with kinetic energy eV . If the atom remains neutral after the collision, the final state comprises a neutral atom and an electron with kinetic energy eV' where $V' < V$. Applying Planck's law, we get

$$h\nu = \frac{hc}{\lambda} = e(V - V')$$

or

$$\lambda = \frac{hc}{e(V - V')}$$

where h is Planck's constant, c is the velocity of light, and e is the electronic charge.

³⁴From the German *bremsen*—to brake and *strahlung*—radiation. Thus, the word means *deceleration radiation*.

When $V' = 0$, λ will be minimum. Then,

$$\lambda_0 = \frac{hc}{eV} \quad (14.66)$$

Equation (14.66) is known as the *Duane-Hunt law*. Substituting the values of h , c and e , the working form of Eq. (14.66) becomes

$$\lambda_0(\text{in } \text{\AA}) \cong \frac{12340}{V(\text{in volts})} \quad (14.67)$$

The spectra of generated radiation are shown in Fig. 14.97 for a few values of accelerating voltage.

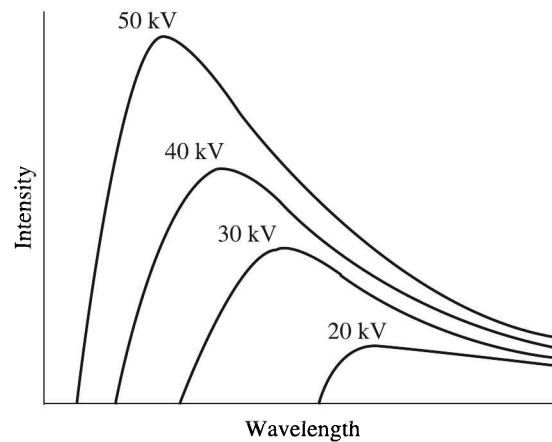


Fig. 14.97 X-ray continuum generated by different accelerating voltages.

It may be seen from the plots that

1. The more the accelerating voltage, the more the intensity of radiation and the shorter the peak wavelength.
2. Also, though not apparent from the plots, the intensity of the radiation increases with the atomic number of the anode element.

Electronic transition between shells. The application of a high voltage between the electrodes held in high vacuum causes sharp electronic transitions from one shell (like K , L , M) to another within the atoms of the anode, resulting in generation of X-rays with definite wavelengths. The K -series of lines are produced when an electron from the innermost K -shell ($n = 1$) is knocked off and electrons from the L - or M -shell drop down to fill the vacancy (Fig. 14.98). The K -series consists of two prominent lines named K_α and K_β .

Similarly, vacancies generated in the L -shell may get filled by electrons from the M - or N -shell to give rise to L -series of lines. All these transitions taken together constitute the fingerprint line spectrum of the material of the anode or a specimen pasted on it. And this fingerprint spectrum is superimposed on the continuous spectrum produced by the bremstrahlung.

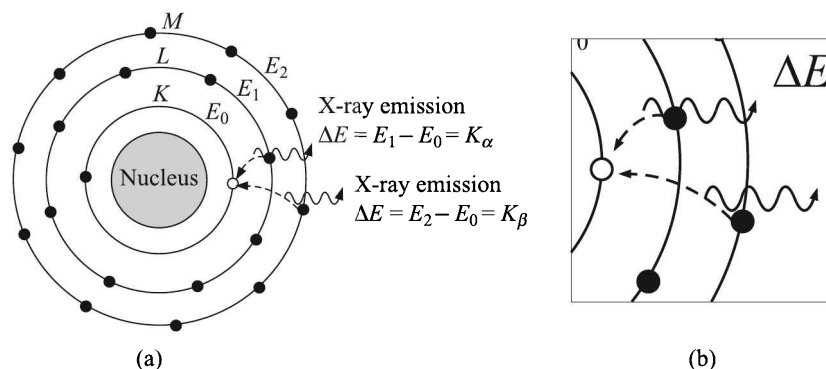


Fig. 14.98 (a) Process of production of X-ray lines of the K -series, and (b) enlarged view.

X-ray tubes

The diagram of an X-ray tube is shown in Fig. 14.99. Common anodes³⁵ used in X-ray tubes include copper and molybdenum, which, with 8 keV and 14 keV accelerating voltages, emit X-rays of wavelengths 1.54 \AA and 0.8 \AA respectively. The anodes are commonly cooled with circulating chilled water so that they do not melt owing to the heat generated. The end of the anode is angled at 5° to 20° perpendicular to the electron current so as to allow some of the X-ray beam to escape through a beryllium window in the perpendicular direction to that of the electron current. A concave mirror-shaped cathode serves as an electrostatic lens that focuses the beam onto a very small spot on the anode.

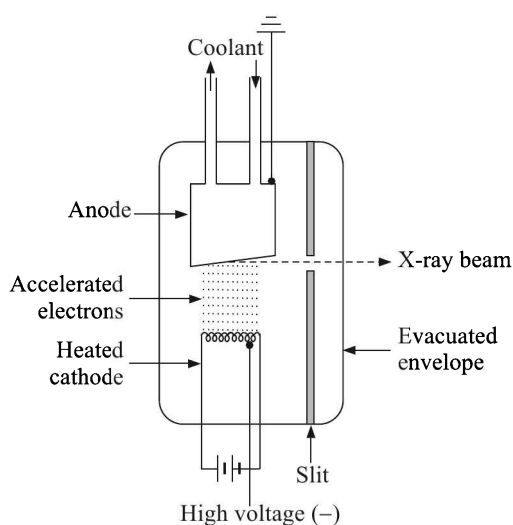


Fig. 14.99 Schematic diagram of an X-ray tube.

But, X-rays of high intensity cannot be generated by this method because X-ray generation is an inefficient process. Only 1% of the incident energy is converted to X-rays and the rest to

³⁵ aka targets.

heat. Thus, a focussed electron beam, when accelerated by a higher voltage, melts the spot on the anode despite cooling.

Rotating anode tube. This problem can be somewhat overcome by using a rotating anode, the construction of which is shown in Fig. 14.100. Typically, the anode is a disc of tungsten or an alloy of tungsten and rhenium. The disc, having a bevelled edge (5° to 20°), rotates at 3000 to 3600 rpm (or 9000 to 10000 rpm). By sweeping the anode past the focal spot, the heat is spread over a larger area, thus increasing the power rating by a factor of about 20.

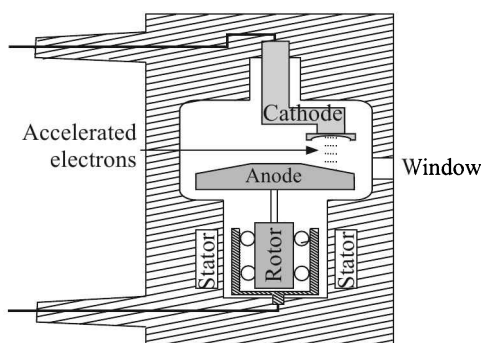


Fig. 14.100 Schematic diagram of a rotating anode tube.

Apart from their limited power rating, X-ray tubes produce an isotropic radiation of which only a small fraction can be extracted through the window. Also, the radiation is not monochromatic. To produce more intense and monochromatic X-rays, one has to have recourse to synchrotron radiation as the X-ray source.

Synchrotron radiation source

It is known that a cyclotron uses a constant magnetic field to turn the particles so that they move in a circular path and an electric field of constant frequency to accelerate the charged particles. A synchrotron is a kind of cyclic particle accelerator in which the magnetic field and the electric field are carefully synchronised with the travelling particle beam. By increasing the magnetic and electric fields appropriately as the particles gain energy, their path can be held constant as they are accelerated. This allows the vacuum container for the particles to be a large thin torus, unlike the disc-shaped chamber of the cyclotron. Actually, it is easier to use some straight sections between the bending magnets and some bent sections within the magnets giving the torus the shape of a round-cornered polygon. A path of large effective radius may thus be constructed using simple straight and curved pipe segments. The shape also allows and requires the use of multiple magnets to bend the particle beam (Fig. 14.101).

When charged particles, in particular electrons or positrons, are accelerated, photons are emitted. At relativistic velocities³⁶ these photons are emitted in a narrow cone in the forward direction, at a tangent to the orbit. In a high energy electron or positron synchrotron, these photons can be emitted with energies of X-rays. This radiation is called *synchrotron radiation*. This radiation is intense (hundreds of thousands of times higher than that of conventional X-ray tubes), monochromatic (i.e. tunable to a desired frequency), intrinsically collimated, pulsed and polarised.

³⁶i.e. when the velocity of moving particles is close to that of light.

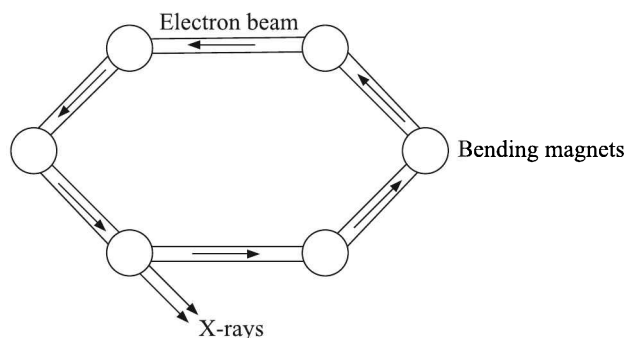


Fig. 14.101 Schematic shape of a synchrotron and X-ray emission therefrom.

The disadvantage is that synchrotrons are bulky instruments and very expensive³⁷.

Monochromators

Nearly monochromatic X-rays can be generated from radiations of ordinary X-ray tubes with the help of what are called *analysers*. These are nothing but crystals of topaz, lithium fluoride, aluminium, sodium chloride, quartz, calcium fluoride, etc. from the planes of which the X-rays are reflected to produce small ranges of radiation. For example, a reflection from the (111)-plane of CaF_2 produces a range of X-rays from 0.466 Å to 6.46 Å. Depending upon the angle that satisfies the Bragg condition, a particular wavelength can be selected (Fig. 14.102).

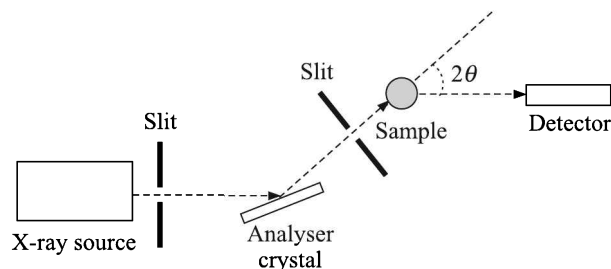


Fig. 14.102 Use of the analyser crystal to produce monochromatic X-rays.

Bragg condition. The Bragg condition is a relation between the incident radiation and the reflected beam from successive planes of the crystal in order that a constructive interference is produced. Reflections of two rays from two consecutive planes of a crystal are shown in Fig. 14.103.

The condition for the constructive interference of these two beams is that the path difference between the two rays should be an integral multiple of the incident wavelength. From the geometry of the figure, it is apparent that the condition is given by Eq. (14.68).

$$n\lambda = 2d \sin \theta \quad (14.68)$$

³⁷There are only about 50 such facilities in the whole world. The only facility in India is located at the Raja Ramanna Centre for Advanced Technology at Indore.

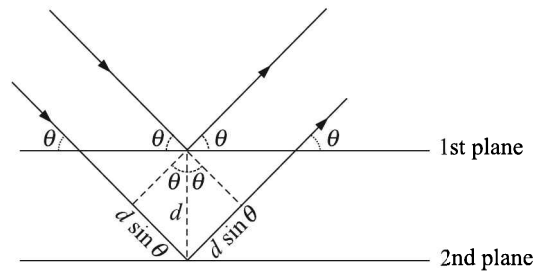


Fig. 14.103 Reflection of X-rays from two consecutive crystal planes.

where λ is the wavelength of the incident X-ray
 θ is the angle of incidence
 d is the lattice spacing of the crystal
 n is the an integer.

Equation (14.68) is called the *Bragg condition*.

X-ray Detection

Apart from special photographic films, X-rays are detected by ionisation chambers, GM counters or scintillation detectors³⁸. Specially made GM counters with thin mica window and a mixture of Ar and $\text{CH}_4/\text{CH}_3\text{OH}/\text{Cl}_2$ gas filling are mostly used to detect X-rays. Nowadays, semiconductor detectors are more common because of their low cost and many advantages.

Also, gas-flow proportional counters and sealed gas detectors are used to detect X-rays.

Si(Li) detector

Si(Li) detector consists of a 3 mm to 5 mm thick silicon junction type *p-i-n* diode with a bias of -1 kV across it. The lithium-drifted centre part forms the non-conducting *i*-layer [Fig. 14.104(a)].

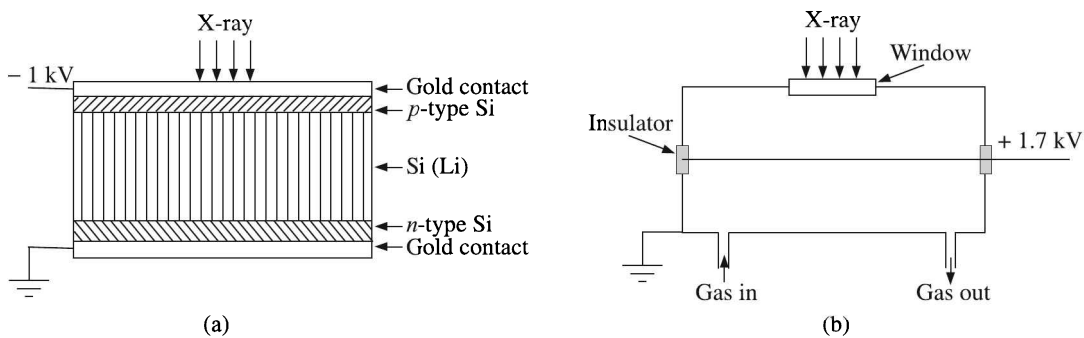


Fig. 14.104 Schematic diagrams of: (a) Si(Li) detector, and (b) gas-flow proportional counter.

³⁸See Section 14.10 at 714.

When an X-ray photon passes through, it causes a swarm of electron-hole pairs to form, and this causes a voltage pulse. To obtain sufficiently low conductivity, and the best resolution, the detector needs to be cooled to liquid-N₂ temperature. Of course, the much more convenient Peltier cooling can be employed, though with some loss of resolution.

Of late, high-purity wafer Si having low conductivity is available. With the Peltier cooling, this provides a cheap and convenient detector, although the liquid-N₂ cooled Si(Li) detector still offers the best energy resolution. Also, CCD detectors have been developed to detect X-rays.

Gas-flow proportional counter

A gas flow proportional counter is used mainly to detect longer wavelengths.

Gas used. A schematic diagram of the gas-flow proportional counter is shown in Fig. 14.104(b). A mixture of 90% argon and 10% methane³⁹ gas flows through it continuously. The incoming X-ray photons ionise argon which produces a measurable pulse owing to the presence of the electric field. The role of methane is to suppress the formation of fluorescent photons caused by the recombination of argon ions with stray electrons.

When multiple detectors are used, the gas is passed through them in series and then led to waste. In case very long wavelengths (over 5 nm) are to be detected, argon may be replaced with neon or helium.

Anode. The anode is typically tungsten or nichrome wire of 20 to 60 μm diameter. The pulse strength is proportional to the ratio of the detector chamber diameter and the wire diameter. Therefore, a fine wire is needed but it must also be strong enough to be maintained under tension so that it remains precisely straight and concentric with the detector.

Window. The window needs to be

1. Thermally conductive so that it can dissipate the generated heat
2. Thin enough to transmit the X-rays effectively
3. Strong enough to minimise diffusion of the detector gas into the high vacuum of the monochromator chamber

Usually beryllium metal, aluminised mylar and aluminised polypropylene are used which satisfy these criteria.

Sealed gas detector

Sealed gas detectors are similar to the gas-flow proportional counters. The gas is usually krypton or xenon at a few atmosphere pressure. They are applied usually to wavelengths in the 0.15 nm to 0.6 nm range. Owing to the problem of manufacturing a thin window capable of withstanding the high pressure difference, they are not applicable to longer wavelengths.

While photographic film is the cheapest and it records the entire diffraction pattern over all angles at a time, its dynamic range is poor (about 250 : 1). Also, it offers a poor angular resolution and energy resolution. So, the data extracted from photography is more of a qualitative nature. GM, proportional or scintillation counters do not suffer from these

³⁹The mixture is known as *P10* in industry.

limitations though they can measure at one angle at a time and therefore, need quite some time to cover the entire angular range. CCD detectors are more like photographic films with the added advantage that they are amenable to computerisation.

X-ray Instrumentation

X-ray tubes need a stabilised high voltage source for the anode and a stabilised current source (~ 10 mA to 100 mA) for heating the filament. The anode also needs a water-cooling arrangement which must maintain a stable temperature. The detector needs a precise goniometer⁴⁰ apart from a data acquisition system. The whole instrumentation is shown schematically in Fig. 14.105.

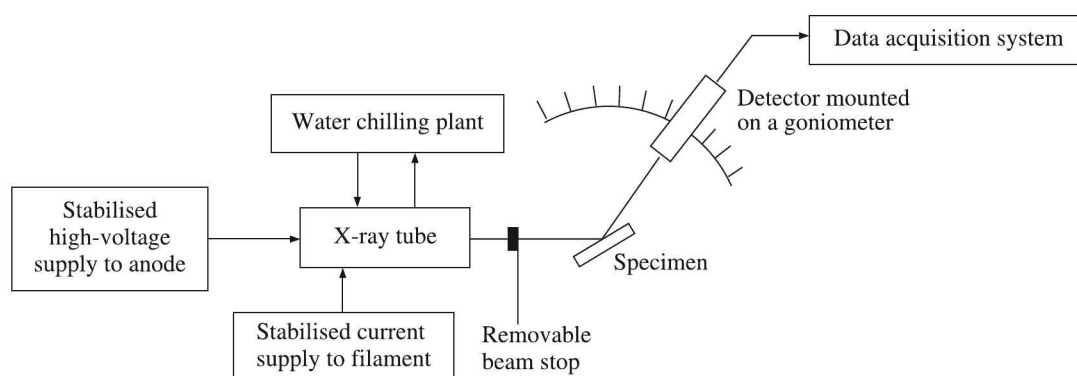


Fig. 14.105 X-ray instrumentation.

We have already stated that X-ray study is a versatile tool for not only materials but also biological specimen. Of them, we will consider only those methods by which analytical studies of materials can be made.

X-ray Diffractometry

For X-ray diffraction (XRD) applications, only short wavelength X-rays (hard X-rays) in the range of a few angstroms to 0.1 \AA (1 keV to 120 keV) are used. Because the wavelength of X-rays is comparable to the size of atoms, they are ideally suited for probing the structural arrangement of atoms and molecules in a wide range of materials. The energetic X-rays can penetrate deep into the materials and provide information about the bulk structure.

X-rays, being incident on crystalline materials, get scattered by atoms. Depending upon the satisfaction of Bragg condition or not, they produce patterns of high or low intensity at different angles. The angle or position of the beam depends on the crystal structure of the material while the intensity depends on the location of atoms in the unit cell of the crystal. Thus, in respect of position and intensity of the scattered beam, no two substances will produce the same pattern. The diffraction pattern is thus a fingerprint of the material. Powders are

⁴⁰ Angle measuring arrangement.

nothing but arbitrarily oriented microcrystals of the material. So, they make no difference in producing a fingerprint pattern.

For perfect, infinite crystals and perfectly collimated beams, the diffraction condition must be satisfied exactly. But no crystal is perfect or infinite and no beam is perfectly collimated. Therefore, because of existence of strains, defects, finite size effects and instrumental resolution, diffraction peaks are broadened. Actually, the scattered intensity is proportional to the square of the Fourier transform of the charge density as given by Eq. (14.69).

$$I(\mathbf{q}) \propto \left| \int \rho(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} d^3r \right|^2 \quad (14.69)$$

where $\rho(\mathbf{r})$ = electron charge density at \mathbf{r}

$$\mathbf{q} = 2k \sin \theta$$

$$k = (2\pi/\lambda)$$

For perfect crystals, $I(\mathbf{q})$ should be δ -functions indicating perfectly sharp scattering while in case of imperfect crystals, the peaks have finite breadth. And for liquids and amorphous solids like glasses, it is a continuous, slowly varying function (Fig. 14.106).

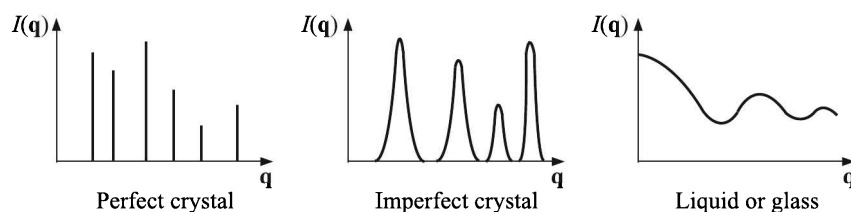


Fig. 14.106 X-ray diffraction patterns for perfect crystal, imperfect crystal and liquid or glass.

From a knowledge of the intensity of diffracted X-rays at different angles in space, the structure of a crystal can be determined by using Eq. (14.69). However, the procedure is rather complicated.

The simplest ways by which X-ray diffraction patterns can be utilised to identify elements are

1. By measuring the frequency of emitted X-rays by the material when used as a cathode
2. By observing the absorption of X-rays by the material

Frequency of emitted radiation

A simple relationship, known as *Moseley's law*, exists between the frequency ν of a particular line emitted by an element and its atomic number Z . The relationship is given by

$$\nu = a(Z - b)^2$$

where a and b are constants that are characteristics of the line under consideration. Moseley experimentally determined the value of a for K_α line to be 82303 cm^{-1} , while from the Bohr's theory of atomic structure it turns out to be 82775 cm^{-1} . Anyway, by knowing a , b and ν , the Z -value of the target material can be figured out.

X-ray absorption study

The relationship between the observed intensity I after absorption and the incident intensity I_0 on the absorbing material is given by

$$I = I_0 \exp \left[- \left(\frac{\mu}{\rho} \right) \rho x \right]$$

where μ is the absorption coefficient of the absorber

ρ is the density of the absorber

x is the thickness of the absorber.

The quantity (μ/ρ) is called the *mass absorption coefficient* and is denoted by m . So,

$$I = I_0 \exp(-m\rho x)$$

A typical m vs. λ curve is shown in Fig. 14.107. The curve clearly shows a few edges which occur owing to dislodging of electrons from corresponding shells as indicated in the figure.

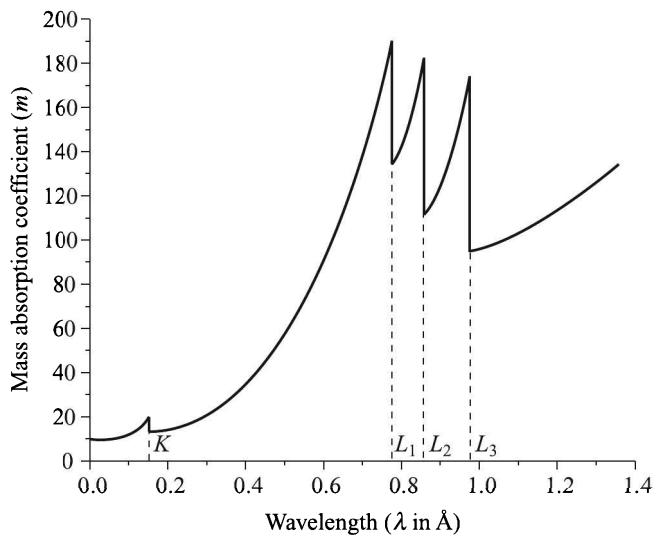


Fig. 14.107 Variation of mass absorption coefficient with wavelength.

Also, it may be observed from Fig. 14.107 that at lower λ (i.e. at higher energies) the number of edges diminishes rapidly and m varies as λ^3 . This corroborates the well-known fact that X-rays become more penetrating and less absorbed at higher energies.

For a constant λ at higher energies, the *Bragg and Pierce law* holds. It is given by

$$\tau_\alpha = CZ^4\lambda^3$$

where τ_α is the atomic absorption coefficient. This law can also be used to identify elements.

Auger Electron Spectroscopy

The Auger electron spectroscopy (AES) is based on the Auger⁴¹ effect which is an electron emission process resulting from transitions of electrons within an excited atom.

When an energetic X-ray photon (or a beam of electrons with energies in the range of 2 keV to 50 keV) is incident on an atom, a core state electron may get removed leaving behind a hole. This emission of electron is called the *photoelectric emission*. As this is an unstable state, the core hole can be filled by an outer shell electron. The electron moving to the lower energy level loses an amount of energy which is equal to the difference of energies of corresponding orbitals. The release (loss) of energy from this non-radiative transition can trigger a second outer shell electron to get emitted from the atom if the transferred energy is greater than its orbital binding energy. This emission is called the *Auger electron emission*. The sequential steps involved in Auger electron emission are illustrated in Fig. 14.108.

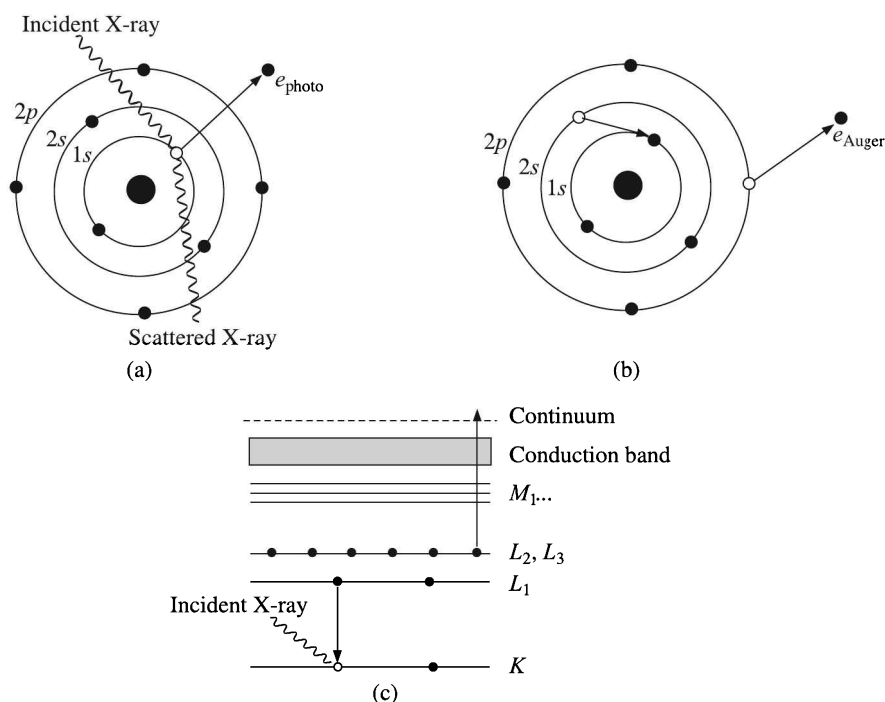


Fig. 14.108 Sequential steps involved in Auger electron emission. (a) An incident X-ray beam creates a hole in the 1s level (photoelectric emission). An electron from the 2s level fills in the 1s hole. (b) The transition energy is imparted to a 2p electron which is emitted (Auger electron emission). (c) The same process is shown in the conventional energy level diagram.

The kinetic energy of the emitted electron will be

$$E_{\text{kin}} = E_{\text{core state}} - E_{\text{o1}} - E_{\text{o2}}$$

where E_{o1} and E_{o2} indicate energies of the first and second outer shells respectively. Quantitative chemical analysis of a sample using AES depends on measuring the yield of Auger electrons during a probing event.

⁴¹Pronounced as "o-zay". Named after the French physicist Pierre Victor Auger (1899–1993).

Auger effect, however, is not the only mechanism available for atomic relaxation. Electrons may as well move to lower energy shells by emitting X-rays. This radiative transition is called the *X-ray fluorescence*. The total transition rate is a sum of those of the non-radiative and radiative processes. The Auger yield ω_A is thus related to the X-ray fluorescence yield ω_X by the relation

$$\omega_A = 1 - \omega_X = 1 - \frac{W_X}{W_A + W_X}$$

where W_X and W_A are X-ray and Auger transition probabilities respectively.

The X-ray fluorescence- and Auger-yield vs. atomic number plots look like Fig. 14.109. It shows that as the atomic number increases, the X-ray fluorescence overtakes the Auger

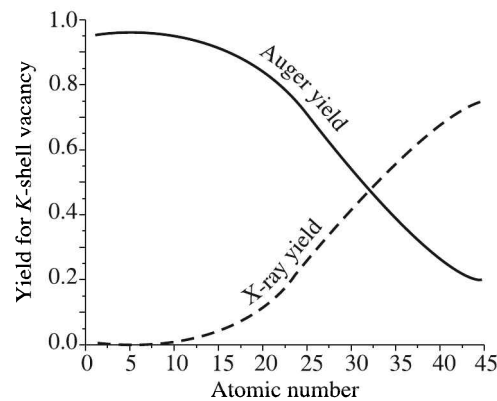


Fig. 14.109 X-ray fluorescence and Auger electron yields vs. atomic number plot for *K*-shell vacancies.

electron emission. Thus, for large Z -values the X-ray yield becomes greater than the Auger yield, indicating an increased difficulty in measuring the Auger peaks for them. Conversely, AES is sensitive to the lighter elements, and unlike X-ray fluorescence, Auger peaks can be detected for elements as light as lithium ($Z = 3$). Actually, lithium represents the lower limit for AES sensitivity since the Auger effect is a *three state* event which means, at least three electrons are necessary for the process.

The electron impact cross-section is a critical quantity that determines the yield of Auger electrons at a detector. An approximate expression for this cross-section is

$$\sigma_{ax}(E) = \frac{1.3 \times 10^{-13} b f(E_p)}{E_p} \text{ cm}^2 \quad (14.70)$$

where b is a scaling factor between 0.25 and 0.35
 E_p is the primary electron beam energy
 $f(E_p)$ is a function of E_p .

Equation (14.70) is actually valid for an isolated atom. To account for the matrix effects, a simple modification can be made as

$$\sigma(E) = \sigma_{ax} \{1 + r_m(E_p, \alpha)\} \text{ cm}^2 \quad (14.71)$$

where r_m is the matrix radius established empirically considering electron interactions with the matrix such as ionisation due to backscattered electrons
 α is the angle subtended by the incident beam to the surface normal.

The total Auger yield can be written as

$$Y(t) = N_x(\Delta t)[\sigma(E, t)](1 - \omega_x) \exp\left(\frac{-t \cos \theta}{\lambda}\right) [I(t)]T \frac{d\Omega}{4\pi} \quad (14.72)$$

where N_x is the number of x atoms per unit volume
 Δt is the thickness of the layer being probed
 λ is the the escape depth of the electron
 θ is the the analyser angle
 $I(t)$ is the flux of excited electron at depth t
 T is the transmittance of the analyser
 $d\Omega$ is the solid angle

Often, all the terms of Eq. (14.72) are not known. So, most analyses compare measured yields with standard values for known compositions. Ratios of the acquired data to standards eliminate interference terms, arising from characteristics of experimental set-up and material parameters. The identification of elements is best done from comparison of pure samples. It also works well for samples of homogeneous binary materials or uniform surface layers.

AES instrumentation

The AES is better utilised for analysing surface materials because the emitted electrons usually have energies ranging from 50 eV to 3 keV. At these energies, electrons have a short mean free path within a solid. Therefore, electrons lying at the depth of within a few nanometres of the target surface can only escape, making AES extremely sensitive to surface species. Owing to the low energy of Auger electrons, AES set-ups need to be run under ultra-high vacuum (UHV) thereby preventing (i) scattering of electrons off the residual gas atoms, and (ii) the formation of a thin gas layer on the surface of the specimen which degrades the analytical performance. The AES instrumentation is shown schematically in Fig. 14.110.

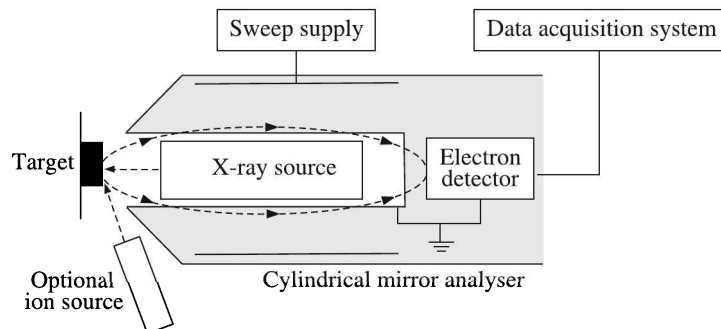


Fig. 14.110 Schematic diagram of an AES set-up.

In this arrangement, a collimated X-ray beam (or a beam of electrons) is incident on a sample and the emitted electrons are deflected onto a cylindrical mirror analyser (CMA). As the voltage applied to the outer cylinder is scanned, the secondary electron distribution is measured by the current output of the analyser. Auger electrons are multiplied in the detection unit and the signal is sent to the data acquisition system. The collected Auger electron current is plotted as a function of energy against the broad secondary electron background spectrum.

The intensity of the Auger peaks is usually small compared to the noise level of the background. Therefore, the AES is often run in a derivative mode which serves to accentuate the peaks by modulating the electron collection current via a small applied AC voltage

$$\Delta V = k \sin \omega t$$

Then, the collection current becomes

$$\begin{aligned} I &= f(V + k \sin \omega t) \\ &\approx I_0 + I'(V + k \sin \omega t) + O(I'') \quad [\text{by Taylor expansion}] \end{aligned}$$

Using the setup in Fig. 14.110, the detected signal at frequency ω will give a value for I or (dN/dE) .

Figure 14.111(a) is an Auger spectrum of Pd metal, generated using a 2.5 keV electron beam to produce the initial core vacancies and hence to stimulate the Auger emission process. The main peaks for palladium occur between 220 and 340 eV. The peaks are situated on a high background which arises from the vast number of so-called secondary electrons generated by a multitude of inelastic scattering processes.

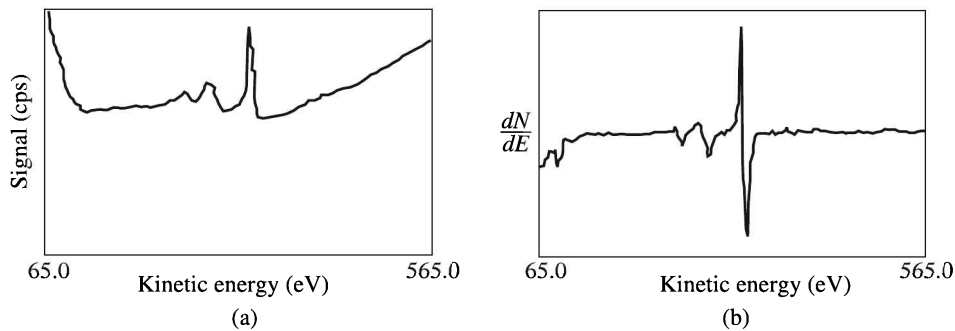


Fig. 14.111 Auger spectrum of Pd metal: (a) direct mode, and (b) derivative mode.

A derivative mode plot, shown in Fig. 14.111(b), accentuates the Auger fine structure which appears as small secondary peaks surrounding the primary Auger peak.

X-ray Fluorescence Spectroscopy

We have already discussed that when a primary X-ray beam from an X-ray tube strikes a sample, it can either be absorbed by the atom or scattered through the material. We know that the process in which the X-ray gets absorbed and transfers all of its energy to an innermost electron ejecting it out of the atom, is called the *photoelectric effect*. This process thus creates holes (or vacancies) in the innermost shell producing an unstable condition for the atom. As

the atom returns to its stable condition, transitions of electrons from the outer shells to the inner shells take place. The consequent release of energy is radiated in the form of X-ray whose energy is the difference between the binding energies of the corresponding shells. The processes involved are diagrammatically shown for a titanium atom in Fig. 14.112.

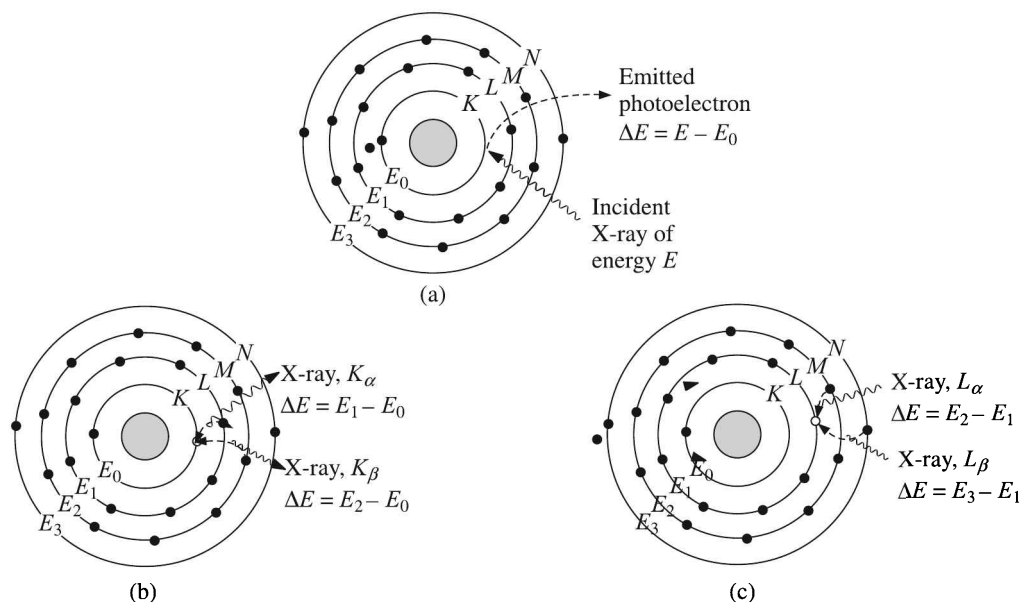


Fig. 14.112 (a) Photoelectron emission, (b) subsequent XRF emission from the K -shell (K_α and K_β), and (c) XRF emission from the L -shell.

Because each element has a unique set of energy levels, each element produces X-rays at a unique set of energies, allowing one to non-destructively measure the elemental composition of a sample. The process of emissions of characteristic X-rays is called 'X-ray fluorescence' (XRF). The spectrum analysis using XRF is called *X-ray fluorescence spectroscopy* (XFS).

In most cases the innermost K - and L -shells are involved in XRF detection. The characteristic X-rays are labelled K , L , M or N to denote the shells they originated from. Subscript α , β or γ is used to indicate transitions of electrons from higher shells. Hence, a K_α X-ray indicates that it is produced from a $L \rightarrow K$ transition of an electron, and a K_β , from a $M \rightarrow K$ transition, and so on. Since within the shells there are multiple orbits of higher and lower binding energy electrons, a further designation is made as α_1 , α_2 or β_1 , β_2 , and so on, to denote transitions of electrons from these orbits into the same lower shell.

The XFS is widely used to measure the elemental composition of materials. Since this method is fast and non-destructive, it is the method of choice for field applications and industrial production for control of materials. Depending on the application, XRF can be produced by using not only X-rays but also other primary excitation sources like α -particles, protons or high energy electron beams.

The XFS suffers from the following difficulties:

1. The process is inefficient.
2. The secondary radiation is much weaker than the incident radiation.

- The secondary radiation from lighter elements has low penetrating power because it is of low energy (long wavelength) and is severely attenuated if the beam passes through air for any distance.

Because of this, for a high-performance analysis, the tube→sample→detector path is maintained under high vacuum (around 0.075 Torr or 10 Pa residual pressure). This means, most of the working parts of the instrument has to be kept in a large vacuum chamber. The problem of maintaining vacuum where samples have to be introduced and withdrawn rapidly, poses a major challenge for the design of the instrument. For less accurate measurements, or when the sample is damaged by a vacuum (e.g. a volatile sample), a helium-swept X-ray chamber can be substituted.

In principle, beryllium ($Z = 4$) is the lightest element that can be analysed by XFS. But due to instrumental limitations and low XRF yield from the light elements, it is rather difficult to analyse elements lighter than sodium ($Z = 11$).

XFS instrumentation

The XFS measurement is normally done in two modes:

- Energy dispersive spectrometry (EDX or EDS)
- Wavelength dispersive spectrometry (WDX or WDS)

Diagrams of both arrangements are shown in Figs. 14.113 (a) and (b) respectively.

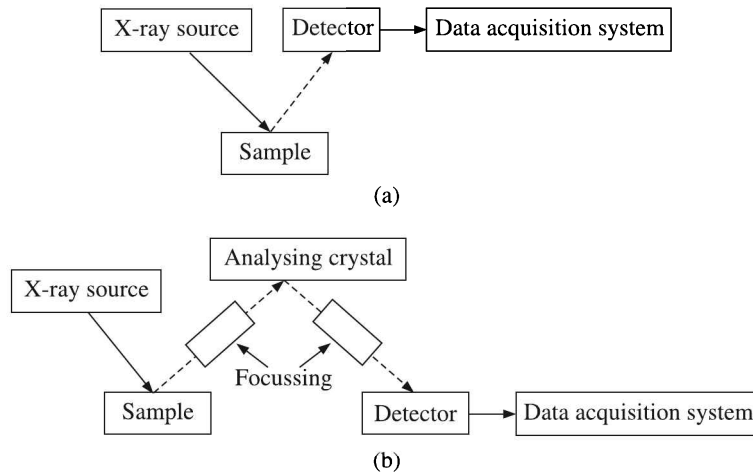


Fig. 14.113 (a) Energy dispersive spectrometry, and (b) wavelength dispersive spectrometry.

In energy dispersive spectrometers, the detector allows the determination of the energy of the photon when it is detected while in wavelength dispersive spectrometers, the photons are separated through diffraction on an analysing crystal before being detected. Although occasionally used to scan a wide range of wavelengths, producing a spectrum plot as in EDS, wavelength dispersive spectrometers are usually set up to measure only at the wavelength of the emission lines of the elements of interest.

Electron Spectroscopy for Chemical Analysis

We have already discussed that when an energetic X-ray beam is incident on a sample, photoelectrons are emitted from it. A typical X-ray photoelectric spectrum is a plot of the number of electrons detected (y -axis, ordinate) vs. the binding energy of the electrons detected (x -axis, abscissa). The method is now better known as ESCA though originally it was termed 'X-ray photoelectric spectrometry' (XPS).

Each element, that exists in or on the surface of the material being analysed, produces a characteristic set of XPS peaks at characteristic binding energy values that directly identify elements. These characteristic peaks correspond to the configuration of electrons $1s$, $2s$, $2p$, $3s$, etc. within the atoms. The number of detected electrons in each of the characteristic peaks is directly related to the amount of element present within the irradiated volume. To obtain atomic percentage values, each raw XPS signal must be corrected by dividing its signal intensity (number of electrons detected) by a *relative sensitivity factor* (RSF) and normalised over all of the elements detected.

To avoid generation of spurious secondary electrons, XPS measurement must be made under UHV conditions because electron counting detectors are typically 1 m away from the material irradiated with X-rays. The schematic diagram of an ESCA spectrometer is shown in Fig. 14.114.

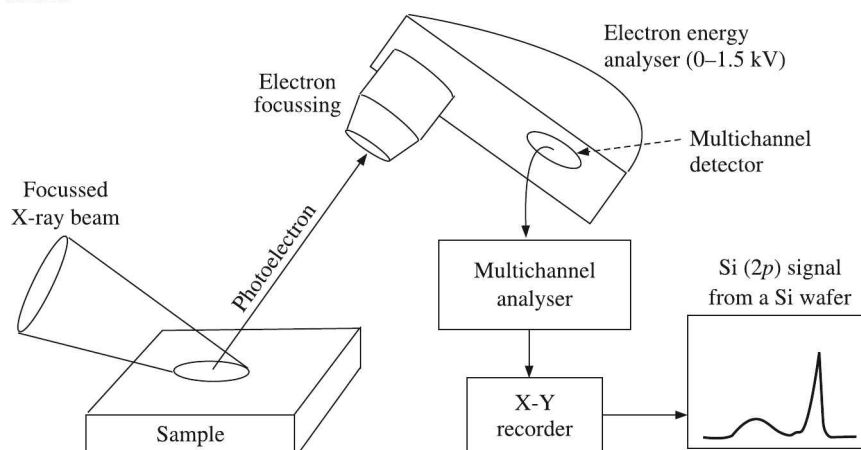


Fig. 14.114 Schematic presentation of an ESCA spectrometer.

We may note that XPS detects the photo-emitted electrons originated from within the top 10 to 12 nm of the material. All of the deeper photo-emitted electrons, which were generated as the X-rays penetrated 1 to 5 μm of the material, are either recaptured or trapped in various excited states within the material. Therefore, for most applications, it is a non-destructive technique that measures the *surface chemistry* of any material.

Monochromatic soft X-rays, such as Mg K_{α} or Al K_{α} lines having FWHM of 0.75 eV and 0.95 eV respectively are usually used for photoelectron excitation. To distinguish photoelectron peaks from Auger peaks, it is necessary to use at least two X-ray sources of different energies because photoelectron peaks shift with the energy of the incident X-ray⁴² while Auger peaks remain constant.

⁴² $\Delta E = E - E_0$ where E = energy of the incident X-ray.

ESCA can be used to detect all elements from lithium ($Z = 3$) to lawrencium ($Z = 103$). This limitation means that it cannot detect hydrogen ($Z = 1$) or helium ($Z = 2$). Detection limits for most of the elements are in the parts-per-thousand (ppt) range. The ppm level is possible, but it requires special conditions. ESCA is routinely used to analyse inorganic compounds, metal alloys, semiconductors, polymers, pure elements, catalysts, glasses, ceramics, paints, papers, inks, woods, plant parts, make-up, teeth, bones, human implants, bio-materials, viscous oils, glues, ion modified materials and many others.

14.10 Radiation Detectors

Geiger-Müller Counter

GM counter consists of a wire in the centre of a cylindrical chamber. The chamber is filled with a gas of 2 cm to 10 cm of Hg pressure. The wire and chamber act as electrodes and are insulated from each other, a bias voltage of about 700 V having been applied between them (Fig. 14.115). As an ionising radiation like α -ray passes through the GM tube, it ionises the gas causing an electrical discharge between the electrodes. The more the radiation, the more frequent is the discharge. So, by counting the rate of electric discharges, the intensity of α -radiation can be ascertained.

GM counters can be used to detect α -, β - and under some conditions X-rays and γ -radiation. The density of the gas in a GM tube being rather low, it is unlikely that a γ -radiation will interact with the gas there. Hence the GM tube is normally insensitive to γ -rays.

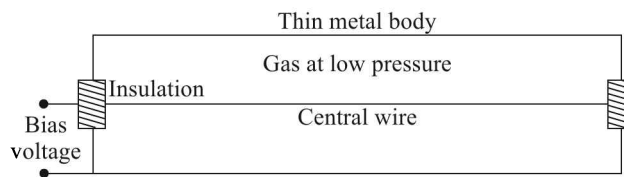


Fig. 14.115 Geiger-Müller tube.

Developed in 1908 by Hans Geiger together with Ernest Rutherford, this counter was originally capable of detecting only α -particles. In 1928 Geiger and Walther Müller, a PhD student of Geiger, improved it upon so that the counter could detect all kinds of ionising radiation. The current version of the GM counter is called the halogen counter. It was invented in 1947 by Sidney H Liebson. It has superseded the earlier GM counter because of its much longer life and lower operating voltage.

Ionisation Chamber

The ionisation chamber is used either to detect particles in air (as in a smoke detector), or to detect or measure ionising radiation.

It consists of a gas filled enclosure incorporating two conducting metal plates (or two electrodes of some shape) separated by a gap. A rather low bias voltage (6 V to 100 V) is applied between the two electrodes which may be in the form of parallel plates or coaxial

cylinders. One of the electrodes may be the wall of the vessel itself. When gas between the electrodes is ionised by any means, say by X-rays or radioactive emission, the ions move to the electrodes of the opposite sign, thus creating an ionisation current of μA -range, which may be measured by a galvanometer or an electrometer. This current is a measure of the intensity of the radiation.

Smoke detector. In its smoke detector variation, the gap between the plates is exposed to open air and a few hundred volts are applied between the electrodes. The chamber contains a small amount of americium-241, which emits α -particles. These α -particles collide with air (mostly nitrogen and oxygen) in the ionisation chamber to ionise them and produce a measurable electric current. If smoke enters the chamber, ions strike smoke particles and are neutralised. As a result there is a drop in current which triggers the alarm.

Scintillation Counter

Scintillation counters contain a transparent phosphor, usually sodium iodide or bismuth germanate crystal, or a specially treated plastic (containing anthracene) material, or an organic liquid which emits photons (scintillates) when struck by an ionising radiation. These photons are sensed by a device called *photomultiplier tube* (PMT) which contains suitably biased multiple photoelectric plates to generate an electrical current that is proportional to the number of photons that strike the PMT. The PMT is attached to an electronic amplifier and other electronic equipment to count and possibly quantify the amplitude of the signals produced by it.

Scintillation counters are popular because though inexpensive they have good quantum efficiency. The quantum efficiency of a γ -ray detector depends upon the density of electrons (per unit volume) in it. Sodium iodide and bismuth germanate have high electron densities because they are composed of elements of high atomic numbers. However, semiconductor—especially hyperpure germanium-based detectors have better intrinsic energy resolution than scintillators, and are preferred for γ -ray spectrometry. For detecting neutrons, high efficiency is gained by using scintillating materials rich in hydrogen that scatter them efficiently. β -radiation is better quantified by liquid scintillation counters.

14.11 Sample Handling Systems

The sample handling process consists essentially of the following steps:

- Step 1. Collecting a representative sample
- Step 2. Transporting the sample to the desired destination
- Step 3. Conditioning the sample
- Step 4. Switching between the sample stream and calibration standards
- Step 5. Sample rejection or venting systems
- Step 6. Handling corrosion or other reactions during the analysis

The collection of representative samples for proper analysis is as important as the analysis itself. Indeed, there is no laid down general rule or procedure for collection of all kinds of samples. For example, if we want to collect a representative sample of river water to study

its turbidity or content of pollutants, we will get different data if the samples are collected from near the banks or from the midstream. Similarly results will vary for samples collected from the surface or from the depth. So, to get a representative set of data, we have to collect several samples and make some kind of statistical analysis.

In a similar way, there are no laid down procedures for steps 2, 4, 5 and 6. So, we leave them out of our discussion here.

But there are standard procedures for conditioning the collected samples especially the gaseous samples. Several portable sample conditioners are available in the market. Sample conditioning can be divided into three rough categories:

1. Removing unwanted particles from the sample
2. Removing aerosol and water vapour from the sample
3. Keeping the sample above its dew point

Removing Particulate Matter

Many gas sample conditioning systems rely on a filter associated with the probe to remove all particulates. The more advanced gas conditioning systems include a separate filter for additional protection. Since the filter element in the sample conditioning system is sufficiently inert and trapped particulates absorb a negligible quantity of gases, filtration has virtually no observable effect on gaseous sample components.

Removing Aerosol and Water Vapour

Coalescing liquids present in the collected gas sample always present a problem. In addition to water, other compounds that are liquids at room temperature (e.g., sulphuric acid) may be present in the sample stream. The system can remove these higher-boiling-point compounds by condensation and subsequent filtration in a coalescing filter, while water remains in the sample stream until removed in the vapour state by a gas dryer. Alternatively, a suitable cooler removes all the condensing vapours at one go by condensation to a low temperature.

Sample dryers

Sample dryers are generally of two types:

1. Various forms of catch-pot, silica gel filters and the like
2. Permeation dryers

While silica gel filters are pretty common, we need to discuss about the permeation dryers in a little more detail.

Permeation dryers. The permeation dryer is a very powerful way of removing moisture from a gas stream. It is one of the compact means available to remove water from a system, particularly if feedback of the dried gas is used to improve drying effect. Many air pollutants, such as SO_2 , are water-soluble and potentially susceptible to removal in a condensation dryer from gas interaction with liquid water. In such situations, the sample gas can be conditioned by the permeation dryer, where the flue gas passes over a selectively permeable membrane

for moisture removal. In this case, water is transferred through the membrane while other pollutants are excluded, and the gas does not come into contact with condensate.

The permeation of gases through a membrane can be analysed as follows. Each gaseous component transporting through the membrane has a characteristic permeation rate that is a function of the ability to diffuse through the membrane material. The mechanism for transport is based on solubility and diffusion. The two relevant relationships are Fick's law for diffusion and Henry's law for solubility.

The diffusive flux through the membrane, according to Fick's Law, is given by

$$J_i = \frac{D_i}{t}(c_{fi} - c_{pi}) \quad (14.73)$$

where J_i is the flux (mole/(m²-s)) of the i th component

D_i is the diffusivity (m²/s) of the i th component

t is the membrane thickness (m)

c_{fi} is the i th component concentration (mole/m³) on the membrane feed side

c_{pi} is the same concentration (mole/m³) on the membrane permeate side.

From Henry's law we get

$$c_i = S_i p_i$$

where S_i is the i th component solubility (mole/m³-Pa) constant in the membrane

p_i is the partial pressure (Pa) of i th component in the gas phase.

The permeation of the i th component P_i through the membrane is the product of solubility and diffusivity

$$P_i = S_i D_i \quad (14.74)$$

The flux rate of Eq. (14.73) can be experimentally determined from the relation

$$J_i = \frac{Q_{ip} \rho_i}{A} \quad (14.75)$$

where Q_{ip} is the volumetric flow rate (m³/s) of the i th component in the permeate

ρ_i is the density (mole/m³) of i th component

A is the surface area (m²) of the membrane.

The permeation dryer can be designed from the knowledge of S_i and D_i and the consequent knowledge of P_i through Eq. (14.74).

Sample coolers

Sample coolers can be divided into two categories:

1. Refrigerant type
2. Peltier elements

Refrigerant type. Refrigerant types use compressors and evaporators like a household refrigerator to cool a sample. Their advantages are:

1. They are more efficient. Although no form of refrigerant system can be seen as particularly efficient in normal engineering terms, this type of cooler requires the least power input for a given cooling effect.

2. They can be built to produce very high levels of cooling, making them ideal for stationary installations and similar use.
3. The construction principles are well-known, making maintenance less of a problem.

However, they have a good number of disadvantages too:

1. The refrigerant used can become a problem for ecological reasons. The traditional chlorofluorocarbon (CFC) coolant that had been used in refrigerators for years is now banned⁴³ due to its contribution to the greenhouse effect. Following the phase out of CFCs and the current preparations for the planned phase-out of HCFCs, HFCs have become the most widely used refrigerants. With their stability, non-flammability, energy efficiency and zero ozone depletion potential, HFCs are suitable for a wide range of applications. But they are costlier.
2. They are not easily portable because
 - (a) A motor, compressor and evaporator have a certain weight.
 - (b) The heat exchanger of a refrigerator is a fragile unit. It consists of thin pipes or similar containing the hot fluid which tries to radiate its heat energy. These can be easily damaged, leading to a loss of the refrigerant resulting to dysfunction.
 - (c) Shaking the refrigerant in transport will lead to vapour locks which must be left to settle before the unit is used. Using the unit immediately after transport will lead to premature failure of the pump and other parts.
3. The evaporator requires a certain area and the heat exchanger must be physically separated from the cooling unit to increase efficiency.
4. Motors and compressors having moving parts, wear out with time and require replacement, making routine maintenance essential.

Peltier elements. The solid state Peltier coolers utilise the well-known Peltier effect to cool the sample. They are compact, low cost and have no moving parts. They can be instantly used after transport without any fear of failure and will cool down more rapidly than the refrigerant type.

But their chief disadvantage is that a Peltier element requires an immense amount of current to produce the cooling effect needed and must generally be cascaded to produce any useful cooling at all.

Keeping the Sample Above its Dew Point

Where ammonia is to be analysed, a temperature control must be used to keep the sample at a higher temperature. The simplest form of heating is electrical power, which is generally used. Roughly speaking, 100 to 150 watts will be needed per metre of hose length to keep the sample above the dew point at all times. Good thermal insulation will reduce the power consumption and prevent the operating personnel from incurring burns during sampling.

⁴³For details of the Montreal Protocol of the United Nations Environment Programme, see http://ozone.unep.org/Publications/MP_Handbook/Section_1.1_The_Montreal_Protocol/

Review Questions

- 14.1 (a) Explain, in brief, the working principle of a thermal conductivity-type gas analyser.
(b) Discuss the principle of operation of NDIR analyser. Name two gases which can be analysed by NDIR analyser.
- 14.2 (a) What is the paramagnetic method exclusively used for the determination of percentage of oxygen in a mixture?
(b) Explain the working of any type of paramagnetic oxygen analyser.
- 14.3 Indicate the correct choice:
- (a) IR spectroscopy
(i) has a useful range of radiation from 2.5 to 15 microns
(ii) is unsuitable for analysis of mixture of metals
(iii) is unsuitable for analysis of organic gases
(iv) uses bolometer as one of the detectors
- (b) The Duane-Hunt law represents the minimum wavelength λ (Å) of X-rays generated for an applied voltage V (volts) and is given by
(i) $V = \frac{12400}{\lambda}$
(ii) 12400λ
(iii) $V = \frac{12400}{\lambda^2}$
(iv) $12400\lambda^2$
- (c) The mathematical basis for nuclear magnetic resonance (NMR) states that, if μ is the nuclear magnetic moment of the molecule, the gyromagnetic ratio is proportional to
(i) μ
(ii) $\frac{1}{\mu}$
(iii) $\sqrt{\mu}$
(iv) μ^2
- (d) In a spectrometer, the monochromator must be able to resolve two wavelengths 599.9 nm and 600.1 nm. The required resolution is
(i) 100
(ii) 1000
(iii) 3000
(iv) 5000
- (e) Attenuation of a narrow monochromatic X-ray beam in a metal plate of thickness d is given by the equation

- (i) $I = I_0 \exp(-\mu d)$
 - (ii) $I = I_0 \exp(-\mu d^2)$
 - (iii) $I = I_0 \exp(-\mu/d)$
 - (iv) $I = I_0 \exp(-\mu/d^2)$
- (f) The transmittance of a coloured solution is 0.5. The absorbance of the solution is
- (i) 0.3
 - (ii) 0.69
 - (iii) 3.16
 - (iv) -1.5
- (g) A gas chromatograph is used for
- (i) measuring the flow rate of a gas
 - (ii) measuring the temperature of a gas
 - (iii) measuring the pressure of a gas
 - (iv) analysing the composition of a gas
- (h) The following method is widely accepted to determine oxides of nitrogen in an automobile emission
- (i) Orsat analysis
 - (ii) Gas chromatography
 - (iii) Chemiluminescence
 - (iv) Flame ionisation detection
- (i) Water vapour absorbs electromagnetic radiation primarily in the range of
- (i) X-ray
 - (ii) IR
 - (iii) UV
 - (iv) γ -ray
- (j) The transmittance of a particular solution measured is T . The concentration of the solution is now doubled. Assuming that Beer-Lambert's law holds good for both the cases, the transmittance for the second would be
- (i) $\frac{T}{2}$
 - (ii) $2T$
 - (iii) T^2
 - (iv) \sqrt{T}
- (k) Given, Planck's constant $h = 6.626 \times 10^{-34}$ Js, charge of an electron $e = 1.6 \times 10^{-19}$ coulomb, velocity of light $c = 3 \times 10^8$ m/s. In an ideal X-ray tube operated at 40 kV, the short wavelength limit of X-ray emission in \AA is ()

-
- (i) 0.155
(ii) 0.31
(iii) 0.62
(iv) 0.93
- (l) The time taken by an ionised atom, of mass m kg and charge e coulomb, pulsed into a field-free region with V volts, to reach a detector L metres away is
- (i) $\frac{1}{L} \sqrt{\frac{m}{2eV}}$
(ii) $L \sqrt{\frac{m}{2eV}}$
(iii) $m \sqrt{\frac{L}{2eV}}$
(iv) $\frac{2}{L} \sqrt{\frac{m}{2eV}}$
- (m) An X-ray source radiates two characteristic wavelengths λ_1 and λ_2 where $\lambda_1 < \lambda_2$. The shorter wavelength can be filtered by passing the radiation through
- (i) a narrow slit
(ii) an interference filter
(iii) a crystalline material whose absorption edge lies between λ_1 and λ_2
(iv) a polariser that rotates the polarisation state of radiation at λ_2
- (n) The dispersion in an X-ray diffractometer, $d\theta/d\lambda$, is given by the expression
- (i) $m/2d \cos \theta$
(ii) $m/2d \sin \theta$
(iii) $2d \sin \theta$
(iv) $2d \cos \theta$
- (o) In a particular sample, the absorbance is 0.6. For a molar concentration of solute 1.0×10^{-4} M and a 2.0 cm path length, the molar absorptivity (litre per mol per cm) is
- (i) 1.2
(ii) 1200
(iii) 3000
(iv) 3×10^{-4}
- (p) Paramagnetism can be used for the analysis of
- (i) O_2
(ii) SO_2
(iii) CO
(iv) CO_2

- (q) X-rays emitted from a molybdenum target at 35 kV have wavelength edges $K_\alpha = 63.2$ pm and $K_\beta = 63.2$ pm. These X-rays are passed through a zirconia filter of K -edge 68.9 pm. The output of the filter contains primarily X-rays of wavelengths
- 63.2 pm
 - 70.9 pm
 - Both 63.2 pm and 70.9 pm
 - Neither 63.2 pm nor 70.9 pm
- (r) Two ionic species A and B of masses in the ratio 1:2 and of equal charge are simultaneously accelerated in a time-of-flight mass spectrometer. When ion A reaches the end of a 1 m drift tube, ion B would be at a distance of
- $\sqrt{2}$ m
 - $(\sqrt{2} - 1)$ m
 - $(1/\sqrt{2})$ m
 - $\frac{1}{2}$ m
- (s) A spectrophotometer is used to measure concentration of a molecule in a solution based on Beer's law using a cuvette of path length 10 mm and detectable absorbance of 0.01. The molar absorptivity of the solution is 10^4 litres/mol/cm. The minimum concentration of the solution that can be detected is
- 1 $\mu\text{mol/litre}$
 - 2 $\mu\text{mol/litre}$
 - 5 $\mu\text{mol/litre}$
 - 10 $\mu\text{mol/litre}$
- (t) Light of intensity I_0 is equally divided and passed through 2 cuvettes P_1 and P_2 containing an analyte at concentrations c and $0.5c$ respectively. The corresponding path lengths in P_1 and P_2 are 4 cm and 1 cm. The cross-sectional areas are 1 cm^2 and 3 cm^2 respectively. The ratio of the absorbances in P_1 and P_2 is
- 1.5
 - $8/3$
 - 3
 - 8
- (u) An X-ray tube is operated at 80 kV anode voltage. In order to filter the low intensity X-rays, a 2.5 mm thick aluminum filter is used. It is given that at 80 kV anode voltage, the mass attenuation coefficient and density of aluminium are $0.02 \text{ m}^2\text{kg}^{-1}$ and $2699 \text{ kg}\cdot\text{m}^{-3}$ respectively and for copper these are $0.075 \text{ m}^2\text{kg}^{-1}$ and $8960 \text{ kg}\cdot\text{m}^{-3}$ respectively. If a copper filter is to replace the aluminium filter with the same filtering effect, the thickness of the copper filter should be
- 0.2 mm
 - 0.66 mm
 - 1.5 mm
 - 5 mm

14.4 Match the instrument to the characteristic:

- | | |
|--------------------|---------------------------|
| (a) Interferometer | (c) Wavelength dispersion |
| (b) Monochromator | (d) Magnifying power |
| | (e) Flatness measurement |

14.5 Match the electromagnetic radiation with the wavelength range:

- | | |
|------------------------|--|
| (a) X-rays | (e) 2.5 μm to 25 μm |
| (b) Vacuum ultraviolet | (f) 400 μm to 700 μm |
| (c) Visible | (g) 0.1 μm to 1 μm |
| (d) Infrared | (h) 10 nm to 100 nm |

14.6 Match the sources and the nature of radiations:

- | | |
|--------------------|-------------------|
| (a) Tungsten lamp | (e) Laser |
| (b) Cobalt-60 | (f) IR |
| (c) He-Ne | (g) UV |
| (d) Discharge lamp | (h) γ -ray |

14.7 Match the detector with the spectroscopy:

- | | |
|------------------------|-------------------|
| (a) Bolometer | (e) UV |
| (b) Photomultiplier | (f) IR |
| (c) GM Counter | (g) γ -ray |
| (d) Photographic plate | (h) X-ray |

14.8 Identify the correct set of matches from the following:

- | | |
|---------------------|--------------------------|
| (a) Beer's law | (e) X-ray diffraction |
| (b) Bragg's law | (f) Thermo-emf |
| (c) Seebeck effect | (g) Electrochemical cell |
| (d) Nernst equation | (h) Absorption of light |
- (i) a \rightarrow h, b \rightarrow e, c \rightarrow g, d \rightarrow f
(ii) a \rightarrow g, b \rightarrow f, c \rightarrow e, d \rightarrow h
(iii) a \rightarrow h, b \rightarrow e, c \rightarrow f, d \rightarrow g
(iv) None of the above

14.9 Identify the correct set of matches from the following:

- | | |
|-------------------------------|------------------------------------|
| (a) Electron capture detector | (e) IR spectroscopy |
| (b) Scintillation counter | (f) Oxygen analyser |
| (c) Thermopile | (g) Gas chromatography |
| (d) Zirconia probe | (h) Nuclear radiation spectroscopy |
- (i) a \rightarrow h, b \rightarrow e, c \rightarrow f, d \rightarrow g
(ii) a \rightarrow f, b \rightarrow h, c \rightarrow e, d \rightarrow g
(iii) a \rightarrow g, b \rightarrow e, c \rightarrow h, d \rightarrow g
(iv) None of the above

- 14.10 (a) Derive the equation $\frac{m}{Ze} = \frac{H^2 r^2}{2v}$ which represents the ion species collected by a detector in a magnetic deflection mass spectrometer.
- (b) Calculate the short wavelength cut-off of X-rays when the supplied voltage is 100000 volts, using Duane-Hunt equation.
- 14.11 In a gas chromatograph, 0.012 g of a pure compound A was injected and the area of the relevant peak was 15 sq. cm. For 0.12 g of an actual sample containing an unknown concentration of component A, the area of the peak was found to be 30 sq. cm. What percentage of compound A was present in the sample?
- 14.12 A voltage applied to an X-ray tube is increased 1.5 times. The short wave limit of an X-ray continuous spectrum shifts $\Delta\lambda = 26$ picometres. Find the initial voltage applied to the tube.
- 14.13 (a) State Beer-Lambert's law for absorption of radiation.
- (b) A particular sample of solution of coloured substance known to follow Beer's law shows 80% transmittance when measured in a cuvette of 1.0 cm optical path length.
- (i) Calculate the per cent transmittance for solution of twice the concentration in the same cuvette.
- (ii) What must be the path length in the cuvette to give the same transmittance (80%) for a solution of twice the original concentration?
- 14.14 For a narrow beam of X-ray radiation of wavelength 62 pm, the mass absorption coefficients for aluminium and lead are $3.48 \text{ cm}^2/\text{g}$ and $72 \text{ cm}^2/\text{g}$ respectively. The densities of aluminium and lead are 2.7 g/cm^3 and 11.3 g/cm^3 respectively.
- (a) By how much is the beam attenuated in an aluminium screen 2.6 cm thick?
- (b) How thick must a lead screen be to attenuate the beam as much?

Hazardous Areas and Instrumentation

In many industrial processes where flammable materials are handled or stored, any kind of ignition or spark caused by measurement systems may give rise to an explosion. To protect both plant and personnel and at the same time carry out measurements, precautions must be taken to ensure that the instrumentation is intrinsically safe.

The potentially flammable areas are known as *hazardous areas* and the materials involved include crude oil and its derivatives, alcohols, natural and synthetic process gases, carbon dust, flour, starch, grain, fibres and flyings among others.

15.1 Classification

Depending upon the overall possibility of an explosion, three separate factors have been identified leading to three different types of hazard classification as detailed in Table 15.1.

Table 15.1 Classification of hazardous areas

| <i>Classification type</i> | <i>Criterion</i> |
|----------------------------|---|
| Area classification | Probability of a hazardous mixture being present |
| Gas classification | The grouping is done considering the required spark ignition energy or flame propagation aspect |
| Temperature classification | An ignition may also be caused by a high surface temperature |

Details about the classes have been worked out mainly by the International Electrotechnical Commission (IEC), National Electrical Code, USA (NEC) and the European Committee for Electrotechnical Standardization (CENELEC¹). Other countries have accepted them with minor additions or alterations.

Area Classification

When dealing with instrumentation for explosion-risk areas, the first important task is to define the various hazardous zones. The purpose of sub-dividing the hazardous area into *zones* is to attempt to indicate the probability of a hazardous mixture being present. This probability can

¹French: Comité Européen de Normalisation Électrotechnique.

then be matched to the probability of the danger posed by the instrumentation. The present IEC standard defines the zones as given in Table 15.2. The USA and Canada classify them as *divisions*. They are also shown in the same table.

There are three zones for gases and vapours and three for dusts.

Table 15.2 Classification of hazardous areas

| <i>Division</i> NEC | <i>Zone</i> IEC and CENELEC | <i>Flammable condition</i> |
|------------------------|--------------------------------|--|
| 1 | 0 (Gases and vapours) | Continuously present or present for long periods—typically > 1000 hours/year |
| | 20 (Dusts) | |
| | 1 (Gases and vapours) | Possible but unlikely to be present for long periods—typically between 10 and 1000 hours/year |
| | 21 (Dusts) | |
| 2 | 2 (Gases and vapours) | Not likely to occur and if it occurs it will only exist for a short term—typically < 10 hours/year |
| | 22 (Dusts) | |

Apparatus or Gas Classification

Except the USA and Canada who have opted to maintain their present gas and dust classification, at present almost all countries have accepted the IEC system of apparatus grouping in a way which indicates that it can safely be used with certain gases and vapours. Actually, it is the apparatus that is grouped, but the consideration for grouping is gases.

Minimum ignition energy. The grouping is based only on the spark ignition or flame propagation aspect of the explosion-proof technique. The relevant metric for this is the Minimum Ignition Energy (MIE) which is a measure of required minimum energy for a localised ignition source, like a spark, to successfully ignite a fuel-oxidiser mixture. As shown in Figure 15.1, the ignition energy depends on the fuel concentration in air and that concentration is considered for which the required ignition energy is minimum.

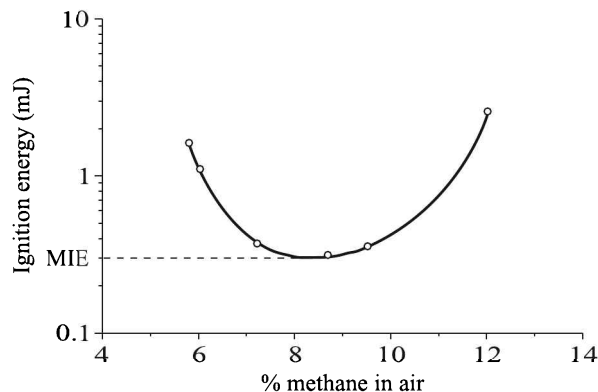


Fig. 15.1 Ignition energy vs. % of methane in air at 1 atm. and 25°C. MIE indicates the minimum ignition energy.

For most combustible fuels the minimum ignition energy is between 0.1 and 0.5 mJ in normal ambient air. However, hydrogen, acetylene and carbon disulphide have one order of magnitude lower minimum ignition energy.

Table 15.3 lists the classification based on the MIE.

Table 15.3 Apparatus or gas classification

| <i>Apparatus</i> | <i>Representative material</i> | <i>IEC</i> | <i>USA and Canada</i> | <i>MIE</i> mJ (approx) |
|-------------------|--------------------------------|-------------------|-----------------------|---------------------------|
| Gases and vapours | Acetylene | Group IIC | Class I, Group A | 0.02 |
| | Hydrogen | Group IIC | Class I, Group B | 0.02 |
| | Ethylene | Group IIB | Class I, Group C | 0.06 |
| | Propane | Group IIA | Class I, Group D | 0.3 |
| | Methane | Group I | No classification | 0.47 |
| Dusts | Aluminium | No classification | Class II, Group E | 15.0 |
| | Coal | Ditto | Class II, Group F | 60.0 |
| | Wheat | Ditto | Class II, Group G | 240.0 |
| Fibres | All | No classification | Class III | |

Temperature Classification

Flammable materials can be ignited not only by a spark but also by contact with hot surfaces. Consequently, all electrical equipment used in hazardous atmospheres are classified according to their maximum attainable surface temperature.

For intrinsically safe circuits, the maximum surface temperature is calculated or measured, including the possibility of faults occurring, in just the same way as the electrical spark energy requirements are derived.

Temperature classifications, accepted by IEC, CENELEC and NEC, are presented in Table 15.4.

Table 15.4 Temperature classification

| <i>Maximum surface temperature, (°C)</i> | <i>Class</i> |
|--|--------------|
| 450 | T1 |
| 300 | T2 |
| 200 | T3 |
| 135 | T4 |
| 100 | T5 |
| 85 | T6 |

Of course, USA and Canada divide T2, T3 and T4 further into 4, 3 and 1 subgroups respectively.

The temperature classification will be marked on items of equipment. If the hazardous area in which the equipment is being installed has gases or vapours with a low autoignition temperature then an equipment with a bigger T-number is needed so as to ensure that any hot surfaces on the equipment will not ignite the hazard.

For example, if a hazardous material has an autoignition temperature of 180°C, then it would be safe to use equipment which is marked T6 or T5 or T4. It would not be safe to

use equipment marked T3 or T2 or T1 as this equipment could exhibit surface temperatures which are hot enough to ignite the hazardous atmosphere.

Also, unless the certification documents state otherwise, the equipment is only certified in ambient temperatures up to 40°C. If exposed to higher temperatures there are two possible dangers:

1. The stated T-class temperature may be exceeded.
2. Safety components within the equipment could degenerate to an unsafe condition.

If it is expected that the equipment will be subjected to temperatures above 40°C — such as in direct sunshine or in a roof space — it is necessary to install equipment which is certified for a higher ambient temperature.

15.2 Explosion Protection of Electrical Apparatus

Electrical apparatus for use in hazardous areas needs to be designed and constructed in such a way that it will not provide a source of ignition. There are eight recognised types of protection for hazardous area electrical apparatus. Each type of protection achieves its safety from ignition in different ways and not all are equally safe. In addition to the equipment being suitable for the Gas Group and the Temperature Class required, the type of protection must be suitable for the zone in which it is to be installed. The different types of protection and the zones for which they are suitable are detailed in Table 15.5.

Table 15.5 Types of protection for gas/vapour hazards

| <i>Technique</i> | <i>Protection type</i> | <i>IEC code</i> | <i>USA, Canada Class (Division)</i> | <i>Suitable for zones</i> | <i>Typical applications</i> |
|-------------------|---------------------------------|-----------------|-------------------------------------|---------------------------|---|
| Energy limiting | Intrinsic safety | Ex ia | I(1, 2) | 0, 1, 2 | Instrumentation, control gear |
| | | Ex ib | Not recognised | 1, 2 | |
| Segregation | Pressurisation | Ex p | I(1, 2) | 1, 2 | Control room, analysers |
| | Oil immersion | Ex o | I(1, 2) | 1, 2 | Transformers, switchgear |
| | Powder filling Encapsulation | Ex q Ex m | Not recognised Not recognised | 1, 2 1, 2 | Instrumentation Instrumentation, control |
| Refined | Increased safety | Ex e | Not recognised | 1, 2 | Motors, lighting fittings |
| Mechanical design | Non-incendiary | Ex m(n) | Not recognised | 2 | Motors, lighting |
| Containment | Flameproof | Ex d | I(1, 2) | 1, 2 | Switchgear, motors, pumps |
| Special | Special | Ex s | Not recognised | 1, 2 | |

Equipment complying with European (CENELEC) standards will frequently bear the code EEx (as opposed to Ex). But the use of EEx is being phased out for equipment designed and certified to the latest editions of the European Standards.

Similar types of protection have been evolved for dust hazard locations. But we are not mentioning them here.

If an equipment is marked

Ex ib IIC T6

it indicates the following:

- Hazardous area equipment built in accordance with a European standard (Ex)
- Intrinsic safety protection type ib
- Apparatus or gas class IIC
- Temperature class T6

Barriers or isolators are placed between the 'safe' and 'hazardous' circuits. The safe devices will not bear the temperature class marking as follows:

Ex ib IIB

Explosion/Flame-proof Enclosures

The term enclosure means, a case that is added to an electrical equipment in such an order that it offers a degree of protection to personnel against specific environmental conditions as well as incidental contact with the equipment.

The IEC developed *Ingress Protection* (IP) codes while the National Electrical Manufacturers' Association of USA (NEMA) developed another set of codes.

The IP Code (or Rating) consists of the letters IP followed by two digits. The first digit indicates the degrees of protection provided against the intrusion of solid objects (including body parts like hands and fingers), while the second digit indicates the ingress of water in mechanical casings having electrical enclosures. The standard aims at providing the users with more detailed information than vague marketing terms such as *waterproof*.

Table 15.6 gives a brief description of different IP code numbers.

For example, an electrical socket rated IP22 is protected against insertion of fingers and will not be damaged or become unsafe during a specified test in which it is exposed to vertically or nearly vertically dripping water. IP22 is the typical minimum requirement for the design of electrical accessories for indoor use.

The NEMA of USA also has evolved protection ratings for enclosures similar to the IP codes of the IEC. However, NEMA also dictates other product features, such as

1. Corrosion resistance
2. Gasket ageing
3. Construction practices

which are not addressed by IP codes. Thus, while it is possible to map IP Codes to NEMA ratings that satisfy or exceed the IP Code criteria, it is not possible to do the reverse since the IP Code does not mandate the additional requirements.

Table 15.6 Meaning of IP code numbers

| <i>1st digit</i> | <i>Solid particle protection against</i> | <i>2nd digit</i> | <i>Liquid ingress protection against</i> |
|------------------|--|------------------|--|
| 0 | Nothing. | 0 | Nothing. |
| 1 | Foreign objects of size > 50 mm. | 1 | Vertically falling water drops. |
| 2 | Foreign objects of size > 12.5 mm. | 2 | Vertically dripping water when the enclosure is tilted at an angle up to 15°. |
| 3 | Foreign objects of size > 2.5 mm. | 3 | Water falling as a spray at any angle up to 60° from the vertical. |
| 4 | Foreign objects of size > 1 mm. | 4 | Water splashing against the enclosure from any direction. |
| 5 | Dust. | 5 | Water projected by a nozzle (6.3 mm diameter) against the enclosure from any direction. |
| 6 | Dust-tight. | 6 | Water projected in powerful jets (12.5 mm diameter nozzle) against the enclosure from any direction. |
| | | 7 | Ingress of water when the enclosure is immersed in water up to 1 m of submersion. |
| | | 8 | Continuous immersion of the equipment in water under conditions to be specified by the manufacturer. |

Table 15.7 indicates the minimum NEMA rating that satisfies a given IP code. But, as discussed, it is given only to have an idea about NEMA ratings and not to map NEMA to IP.

With this background on classification and protection measures for hazardous areas, we now move on to the instrumentation part.

Table 15.7 IP to NEMA mapping

| <i>IP code</i> | <i>Corresponding min. NEMA rating</i> |
|----------------|---------------------------------------|
| IP 20 | 1 |
| IP 54 | 3 |
| IP 65 | 4, 4X |
| IP 67 | 6 |
| IP 68 | 6P |

15.3 Intrinsically Safe Instrumentation

The concept of intrinsic safety is different from other recognised methods of protection, and as such it is worth considering why it has its name.

If something is *intrinsically safe*, this implies that it is safe by itself without any help from outside. So, to maintain the intrinsic safety of an apparatus, it is necessary to use it properly and not to do anything which will interfere with its inherent safety.

To understand the underlying principle we may quote some examples. An electrical switch on its own is intrinsically safe. It may be carried to a hazardous area and operated as many times as we want, provided it is not connected to anything. However, if it is connected to a circuit in which, say 40 A is flowing at 200 V dc, its intrinsic safety is ruined. Similarly, an RTD, say Pt-100, is intrinsically safe by itself. But once it is connected directly across a 24 V supply, it might get dangerously hot and might ignite a hazardous gas or vapour.

In this context, the following definitions are typical:

Intrinsic safety. A protection technique based on the restriction of electrical or thermal energy which can cause ignition by either sparking or by heating effects.

Intrinsically safe circuit. Equipment and wiring which is incapable of releasing sufficient electrical or thermal energy under the normal or abnormal conditions to cause ignition of a specific hazardous atmospheric mixture in its most easily ignited concentration.

Intrinsically safe instrumentation in hazardous areas can be of two types:

1. Pneumatic instrumentation
2. Intrinsically safe electronic instrumentation

Pneumatic Instrumentation

Compressed air is used to generate and transmit signals in a pneumatic instrumentation system. The signal is usually generated with the help of nozzle-flapper transducers², the standard signal range being 3 to 15 psig (0.2 to 1.0 bar).

Pneumatic measurement systems are free from electrical disturbances and hence safe in hazardous areas. But they have many disadvantages as well. The comparison between pneumatic and electrical measurement systems is given in Table 15.8.

Table 15.8 Comparison between pneumatic and electrical measurement systems

| <i>Pneumatic system</i> | <i>Electrical system</i> |
|---|--|
| 1. Simple, robust, reliable and not affected by electrical interference. | 1. Somewhat complex and affected by electrical interference. |
| 2. Bigger in size, contains moving parts and therefore, requires maintenance. | 2. Can be very small, no moving parts and almost maintenance-free. |
| 3. Time delay in transmitting signal considerable. Delay increases with distance of transmission. | 3. Almost instantaneous transmission. |
| 4. Perfectly safe in hazardous areas. | 4. Needs to deploy proper barriers in hazardous areas. |

Suppose, we need to measure temperature between 0°C and 100°C and need to transmit it over a distance pneumatically. One way of doing it is to introduce a Bourdon tube in the enclosure as shown in Fig. 10.3 at page 378 and allow the end of the tube to act as the flapper of a nozzle-flapper transducer to produce a pneumatic signal.

But the problem associated with the flapper-nozzle transducer is that its time constant is rather long, typically a few minutes, because the nozzle and orifice diameters being very

²See Section 6.1 at page 170.

small, they offer a high resistance to flow. As a result, the nozzle-flapper transducer cannot be directly connected to a pneumatic transmission line. The signal has to be sent via a device which allows an increased air flow and therefore a lower time constant. This is achieved through a *relay amplifier*.

Relay amplifier

The heart of a pneumatic relay is a double valve. Air from the supply can flow to the transmission line via the double valve when the diaphragm is depressed (Fig. 15.2). This path is free from restrictions and therefore, allows a high flow rate of air into the transmission line.

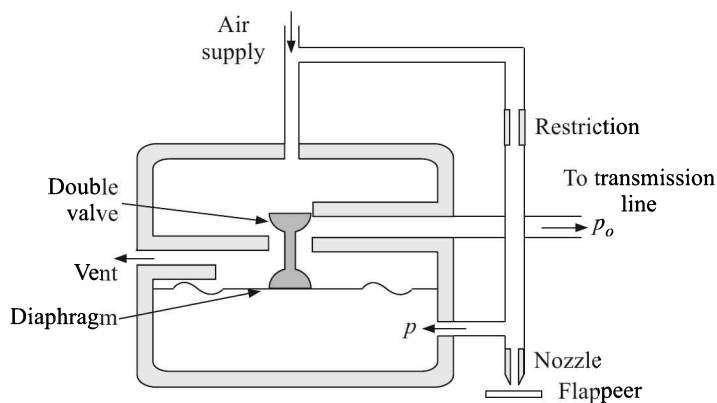


Fig. 15.2 Diagram of a pneumatic relay.

The diaphragm moves up and down according to the feedback it receives from the nozzle-flapper transducer. There is another no-restriction-path via the double valve. It connects the transmission line to a vent port. This is necessary to allow a high flow rate of air out of the line when it is necessary to depressurise the line.

If p be the back pressure from the nozzle
 p_o be the output pressure
 α be the area of cross-section of the diaphragm
 k be the stiffness of the diaphragm
 y be the the deflection of the diaphragm at any instant

then the force acting on the diaphragm is $p\alpha$, because the pressure above the diaphragm is the atmospheric pressure which is equal to zero gauge pressure. This force is balanced by the deflection of the diaphragm. Therefore,

$$p\alpha = ky$$

or

$$y = \frac{\alpha}{k}p$$

Now, the relation between p and p_o will depend on the factors like resistances offered by the supply and vent paths, which, in turn, depend on the geometry of the ball valve. However,

through a careful design, this relation can be made very nearly linear. Therefore, the steady-state sensitivity of the relay, defined by

$$K = \frac{\Delta p_o}{\Delta p}$$

is nearly constant and its value varies between 1 and 20, depending on the construction of the relay. The typical air flow rate through a relay is 1 kg/h. Owing to such high flow rate, the time constant comes down to a few seconds.

Pressure regulators

As the name suggests, the primary function of a pressure regulator is to restrict the flow of gas through it to the demand for gas placed upon the system. If the load flow decreases, then the regulator flow must decrease also and vice versa, such that the controlled pressure does not decrease owing to a shortage of gas in the pressure system.

A pressure regulator generally consists of the following elements:

1. Restricting element
2. Loading element
3. Measuring element

Their functions are as follows:

Restricting element. The restricting element is a type of valve that can offer a variable restriction to the flow.

Loading element. Its function is to apply the needed force to the restricting element. It can be a weight, a spring, a piston actuator, or more commonly the diaphragm actuator in combination with a spring.

Measuring element. Its function is to determine when the inlet flow is equal to the outlet flow. The diaphragm is often used as a measuring element because it can also serve as a combined element.

A single-stage regulator is shown in Fig. 15.3. Here, a diaphragm is used with a valve to regulate pressure. As pressure in the upper chamber increases, the diaphragm is pushed upward, causing the valve to reduce flow. This brings the pressure back down. The top screw

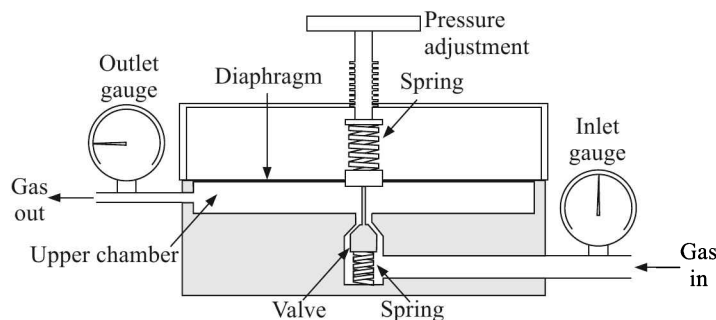


Fig. 15.3 Schematic diagram of a single-stage pressure regulator.

can be adjusted to increase the downward pressure on the diaphragm. Then the upper chamber will require more pressure to maintain the equilibrium. In this way, the outlet pressure of the regulator is controlled.

A pressure regulator may be used to maintain a steady flow of supply gas to the system.

Pneumatic transmitters

As we have seen in Section 6.1 at page 170 that at the industrial supply pressure of 20 psig, the nozzle-flapper transducer has nearly linear characteristic between 3 and 15 psig which is the industrial standard for pneumatic signal transmission. The nozzle-flapper transducer, owing to the small diameters of the orifice and the nozzle, has a large time-constant. The relay is used to lower the time constant. Now, if the pressure variation generated by measurement system, be it temperature or pressure or anything else, is much higher or too lower, it has got to be either attenuated or amplified to the standard level. Pressure transmitters precisely do that. Pressure transmitters are usually of two types, namely

1. Force-balance type
2. Torque-balance type

Force-balance transmitter. In any balancing system, a self-balance system continuously balances an adjustable quantity against a sensed quantity. Once the balance is achieved, the adjustable quantity indicates the sensed quantity. A common balance for weighing materials is an example.

Such a system is perfectly linear, which is why these balances are very useful in measurement systems. Figure 15.4 illustrates a force-balance pneumatic transmitter which balances a sensed differential pressure with an adjustable air pressure which becomes a pneumatic output signal.

The differential pressure, sensed by the liquid-filled capsule, transmits a force to the force bar. The moment acting on the force bar at the fulcrum, deflects its top and, therefore the nozzle, away from the flapper. This displacement is sensed by the pneumatic relay which sends a different amount of air pressure to the bellows unit. The bellows presses against the range bar. The range bar, in turn, pivots on the range-nut fulcrum to counteract the initial deflection of the force bar. When the system returns to the equilibrium, the air pressure within the bellows represents linearly the process differential pressure applied to the capsule.

The adjustable gains achieved at two stages, once at the fulcrum and then at the range nut, allow to trim the output signal to the desired limit of 3 to 15 psig.

Torque-balance transmitter. A schematic diagram of the torque-balance pneumatic transmitter is presented in Figure 15.5. The downward force F generated by the process pressure tends to rotate the torque bar in the counter-clockwise (CCW) direction because of its resting on a fulcrum. As a result, the distance between the flapper and nozzle decreases causing an increased back pressure p as well as the relay output pressure p_o . This increased pressure is fed to the transmission line and also back to the bellows which presses on the torque bar to generate two forces, namely

1. A downward force of magnitude αp_o , where α is the area of cross-section of the bellows
2. An upward force F_o produced by the spring

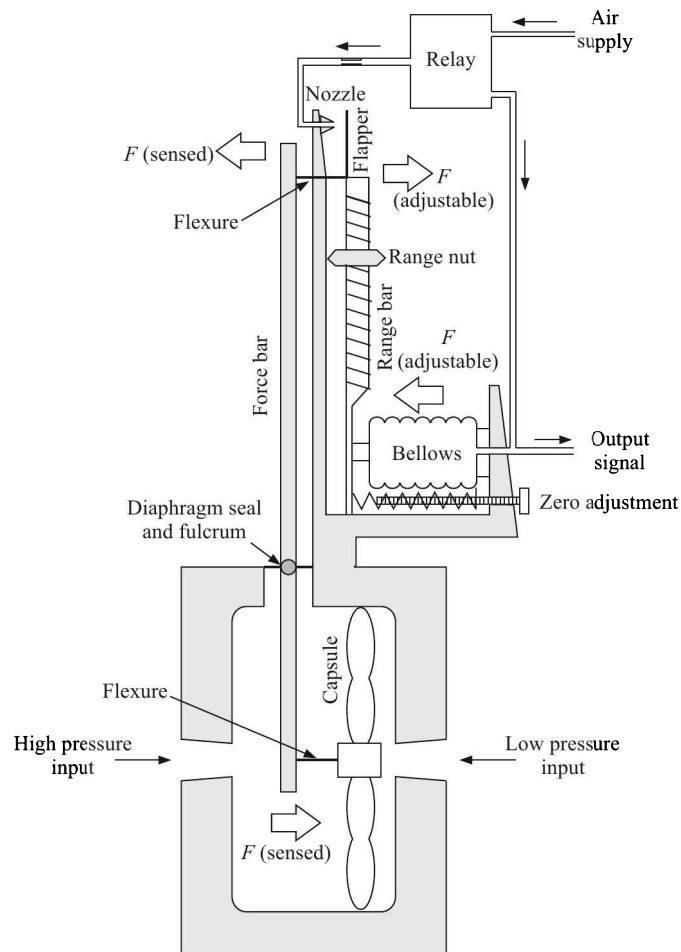


Fig. 15.4 Force-balance pneumatic transmitter.

Now the torque bar will try to rotate in the clockwise (CW) direction and an equilibrium will reach when the CCW and CW moments balance. Thus, for equilibrium,

$$\underbrace{Fx + F_0y}_{\text{CCW moments}} = \underbrace{\alpha p_o y}_{\text{CW moment}}$$

or

$$p_o = \frac{x}{\alpha y} F + \frac{F_o}{\alpha}$$

The last factor being a constant of the system, the output pressure is linearly proportional to the input force, or for that matter, the input pressure. As in the force-balance, the output pressure can be adjusted to the required pressure range by adjusting the position of the fulcrum.

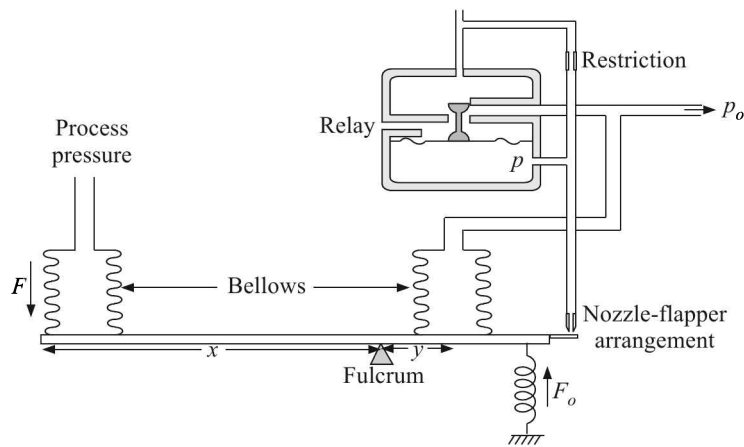


Fig. 15.5 Torque-balance pneumatic transmitter.

Transportation lag. The pneumatic transmission, it can be shown, is a first order system with a transport lag or input-delay. The corresponding transfer function can be written as

$$\frac{\Delta p_o}{\Delta p_i} = \frac{K e^{-\tau_D s}}{1 + \tau s}$$

where τ_D is the input-delay. It has been observed that the ratio $\tau_D/\tau \cong 0.24$ irrespective of the length of the transmission line and the pipe diameter. We present below data for different lengths of line for a pipe of 1/4 inch outside diameter:

| Length(ft) | τ_D (s) | τ (s) |
|------------|--------------|------------|
| 500 | 0.77 | 3.2 |
| 1000 | 2.3 | 9.7 |
| 2000 | 7.4 | 31 |

It is amply clear from the data that pneumatic transmission of signal over 100 ft, where total time lag is >0.4 s, is not suitable. In fine, pneumatic signals are best transmitted for safety in hazardous areas where the line length does not exceed 100 ft. Outside of the hazardous area, they may be converted to electric signals by using any suitable pressure transducer as discussed in Section 8.5 at page 296.

Intrinsically Safe Electronic Transmission

We know hazardous area may contain flammable gasses or vapours, combustible dusts, or ignitable fibres or flyings. There are different systems used in Europe or the United States to classify the type of hazard and whether the hazard is always present or only present in an emergency condition. An intrinsically safe circuit is designed for the worst case, which would be to assume the explosive atmosphere is always present and the electrical or thermal energy is the lowest required to cause an ignition.

The general scheme of an intrinsically safe design is illustrated in Fig. 15.6.

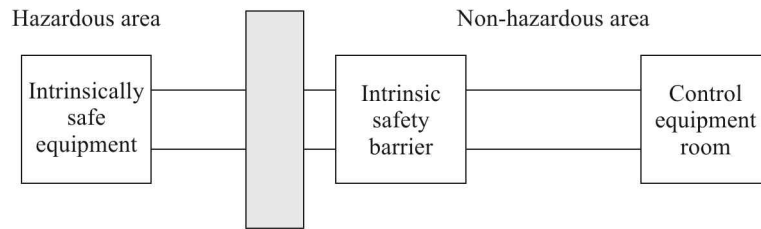


Fig. 15.6 Diagram of an intrinsically safe system.

Most applications require a signal to be sent out of or into the hazardous area. The equipment mounted in the hazardous area must first be approved for use in an intrinsically safe system. The barriers designed to protect the system must be mounted outside of the hazardous area and in an area designated as non-hazardous or safe in which there is no hazard or will not be present at any time.

The approval agencies recognise devices known as *simple apparatus*. A simple apparatus does not generate or store more than the following:

1. 1.2 V
2. 100 mA
3. 20 mJ
4. 25 mW

This type of device does not need certification from a third party. Even though a device is considered a simple apparatus, it must be connected to an intrinsic safety barrier. Examples of simple apparatus are:

| | |
|------------------|---|
| Discrete Inputs | 1. Limit, Pressure, Float, Flow and Temperature switches 2. Push buttons |
| Analogue Inputs | 1. Thermocouples 2. RTDs |
| Discrete Outputs | LEDs |

Other equipment which has been designed for and is available for use in hazardous areas with intrinsically safe barriers includes:

1. Strain gauges
2. Potentiometers
3. Proximity switches
4. Infrared temperature sensors
5. Electromagnetic flowmeters
6. Solenoid valves
7. 4-20 mA dc two-wire transmitters

But, their installation needs to be certified by an accredited third party.

Intrinsically safe barriers

The barrier can be either of the following:

1. Zener diode barrier
2. Transformer isolation barrier

Zener diode barrier. Zener diode barriers limit the energy supply to the hazardous area to a level below the minimum ignition energy of the specific air/gas mixture. This is accomplished by protecting against the following faults:

1. Shorting of wires connected to the hazardous area side of the barrier.
2. Proper grounding of wires connected to the hazardous area side of the barrier.
3. Allowing an unsafe voltage to be applied to the safe area side of the barrier through misconnection or failure of the power supply.

As a result of these protections, the barrier will prevent unsafe levels of voltage/current from passing into the hazardous area. A schematic diagram of a simple Zener diode barrier is shown in Fig. 15.7.

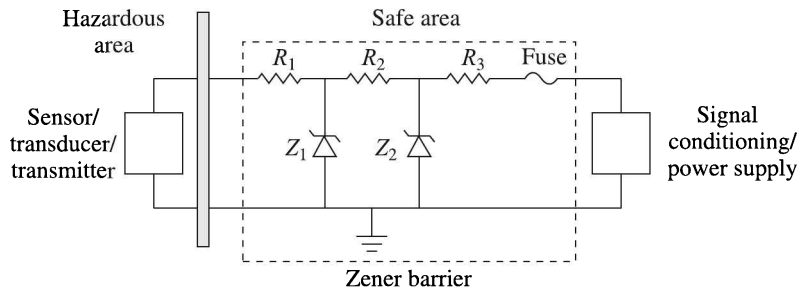


Fig. 15.7 A simple Zener diode barrier circuit.

The circuit comprises three resistors for current limitation, a fuse which opens in the case of excessive current and redundant Zener diodes which limit the voltage to a safe level. The fuse resistance prevents the flow of a high current so that Zener diodes Z_1 and Z_2 are protected. The current rating of the fuse is substantially lower than the surge current rating of the diodes. So, if a fault current is increasing fast, the fuse should blow and Z_2 should provide the fault current a safe path to the ground. If the fuse fails, the second and third lines of protection are provided by the two Zener diodes which should have an appropriate voltage rating.

The following example will give an idea as to how to find out whether a Zener barrier is safe or not.

Example 15.1

Figure 15.8 shows an equivalent circuit of a Zener barrier used for the measurement of temperature by a thermocouple in a hydrogen-air atmosphere.

Symbols used for the sensor, cable network and the Zener barrier are self-explanatory. If $V_Z = 10 \text{ V}$, $R_1 = 50 \Omega$, $C = 1.85 \text{ nF}$, $L = 60 \mu\text{H}$ and the minimum ignition energy for hydrogen-air is $19 \mu\text{J}$, examine from the calculation of maximum total energy stored in the circuit if the instrumentation is safe or not.

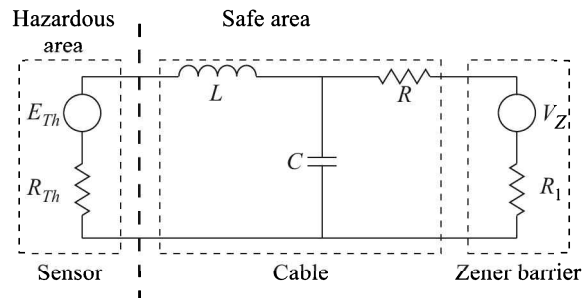


Fig. 15.8 Example 15.1.

Solution

Energy will remain stored in the capacitor and the inductor. So,

$$\begin{aligned}
 E_{\max} &= \frac{1}{2}CV_Z^2 + \frac{1}{2}Li^2 \\
 &= \frac{1}{2}\left[CV_Z^2 + L\left(\frac{V_Z}{R_1}\right)^2\right] \\
 &= \frac{1}{2}\left[(1.85 \times 10^{-9})(10)^2 + (60 \times 10^{-6})\left(\frac{10}{50}\right)^2\right] \text{ J} \\
 &= 1.2925 \mu\text{J}
 \end{aligned}$$

Since, $E_{\max} \ll 19 \mu\text{J}$ (MIE for hydrogen-air), the barrier is perfectly safe.

Disadvantages of Zener diode barriers. Though Zener diode barriers provide suitable protection, operate properly and are cost effective, they suffer from the following disadvantages:

1. They require a high integrity intrinsically safe grounding. The resistance of the ground line must not exceed 1Ω .
2. A regulated power supply has to be used to prevent the possibility of the voltage
 - (a) either rising too high that may cause the Zener diodes to conduct and the fuse to blow
 - (b) or falling too low which may prevent the minimum operating voltage from reaching the field device
3. Process control loops have to totally float above ground. Therefore, it is possible to acquire electrical noise on the control signal.
4. Since the barrier is encapsulated in epoxy, in case the protective fuse blows, it cannot be replaced and the barrier has to be discarded.

Transformer isolation barrier. A transformer isolated barrier (TIB) contains a Zener barrier for the voltage and current limitation. However, it does not require an intrinsically safe grounding. The transformer in a TIB provides a high degree of isolation between the primary and secondary windings. Therefore, the ground connection is unnecessary. The diagram of a simple TIB is shown in Fig. 15.9.

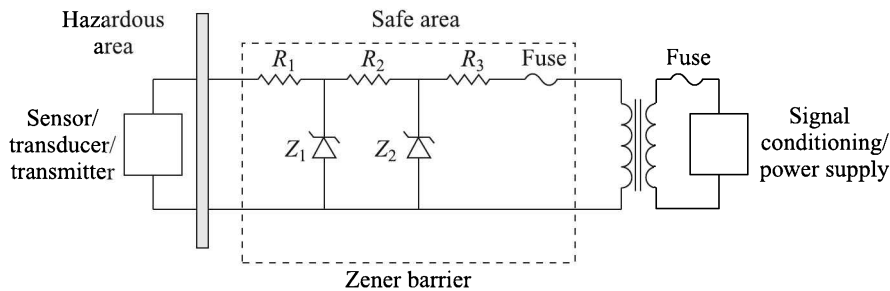


Fig. 15.9 A simple transformer isolated barrier.

Advantages of TIBs. The advantages of TIBs are as follows:

1. Transformer isolated barriers do not require a high integrity intrinsically safe ground since the transformer isolates the hazardous area connections from the safe area connections.
2. Any process control loop (mA or mV signal) connected to a TIB will remain floating, unlike Zener barriers that ground one side of the signal or at best provide a quasi-floating system. This prevents the dreaded ground looping (see *Multiple earths* at page 14).
3. TIBs can be repaired, unlike Zener barriers that are encapsulated in epoxy.
4. TIBs contain separate current limiting circuitry (shown as 'Fuse') that will spare the protective barrier fuse to blow in case of a short circuit.

Review Questions

15.1 Indicate the correct choice:

- (a) Petroleum refineries fall in which of the following class of hazardous area classification?

| | |
|-----------------|--------------------|
| (i) Class I | (ii) Class II |
| (iii) Class III | (iv) None of these |
- (b) Pneumatic relays are used to perform

| | |
|--------------------------------|----------------------------|
| (i) arithmetic operation | (ii) logarithmic operation |
| (iii) boosting of input signal | (iv) integration |
- (c) Grain elevators belong to which class of hazardous area?

| | |
|-----------------|--------------------|
| (i) Class I | (ii) Class II |
| (iii) Class III | (iv) None of these |

15.2 What is the purpose of installing a pneumatic regulator? Explain with a neat sketch the function of a pneumatic regulator.

15.3 What is the basis of classification of hazardous areas? give examples of Class I, Class II and Class III type of hazardous areas. Explain different types of Ingress Protection (IP) code for the instrument enclosure.

- 15.4
- (i) What do you mean by intrinsic safety?
 - (ii) What are the different areas categorised as far as intrinsic safety is concerned?
 - (iii) How is the Zener barrier used for intrinsic safety?

-
- 15.5 (i) With the help of a diagram explain the force-balance mechanism of pneumatic transmission.
- (ii) Why is a pneumatic relay commonly used at the output of a flapper-nozzle assembly?
- (iii) What are the possible advantages and disadvantages of a pneumatic transmitter over an electronic transmitter?
- 15.6 (i) What is the difference between IP codes and NEMA codes for enclosures?
- (ii) What is meant by intrinsic safety? What is the advantage of using an intrinsically safe circuit?
- (iii) What does IP65 signify? What is its NEMA equivalent?
- 15.7 What is a pneumatic relay? Describe with a neat sketch the bleed type pneumatic relay.

Signal Conditioning

The total area of signal conditioning can be divided into the following sections:

1. Bridge circuits
2. Processes
3. Recovery of signals
4. Signal conversion

We will treat the subject in that order.

16.1 Bridge Circuits

Signals in an instrument are produced by transducers. We have already pointed out that transducers are of two types—active and passive. While active transducers themselves generate signals, passive ones—like strain gauges, resistance thermometers, resistive-, inductive-, capacitance-displacement transducers—need be excited by external voltage sources to generate signals. One of the necessary conditions of measuring these signals is that the measuring system must not load the measured medium. Bridge circuits are very suitable for this purpose. Therefore, before we consider other aspects of signal conditioning, we will discuss a few basic bridge circuits.

Various types of bridge circuits are used to measure change in resistance, capacitance and inductance of passive transducers. In static measurements a null method is used allowing no flow of current through the measuring instrument. This is an ideal situation because here the measuring instrument does not load the measured medium.

Bridges are of two types: (i) dc bridge and (ii) ac bridge. While either of the two types may be used for resistance measurement, for capacitance or inductance measurement we have to have recourse to ac bridges.

DC Bridge: Wheatstone Bridge

The basic circuit of the bridge is given in Fig. 16.1 along with its Thevenin-equivalent. It is clear that the internal resistance and open-circuit voltage of the Thevenin generator are given by

$$R_0 = \frac{R_1 R_2}{R_1 + R_2} + \frac{R_3 R_4}{R_3 + R_4} \quad (16.1)$$

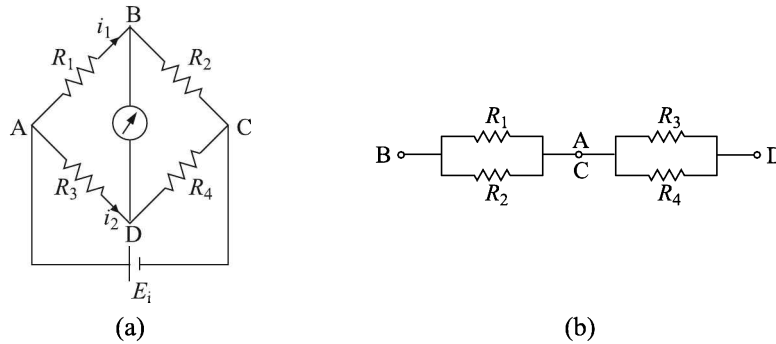


Fig. 16.1 (a) Basic dc bridge circuit, and (b) its Thevenin equivalent.

$$\begin{aligned}
 E_o &= i_1 R_1 - i_2 R_3 \\
 &= \left[\frac{R_1}{R_1 + R_2} - \frac{R_3}{R_3 + R_4} \right] E_i \\
 &= \left[\frac{R_1 R_4 - R_2 R_3}{(R_1 + R_2)(R_3 + R_4)} \right] E_i
 \end{aligned} \tag{16.2}$$

In the null method $E_o = 0$, which yields

$$R_1 R_4 - R_2 R_3 = 0$$

or

$$\frac{R_1}{R_2} = \frac{R_3}{R_4}$$

If R_m is the resistance of the measuring device, then with the help of Eqs. (16.1) and (16.2) the current i_m flowing through it can be obtained as

$$\begin{aligned}
 i_m &= \frac{E_o}{R_0 + R_m} \\
 &= \frac{R_1 R_4 - R_2 R_3}{R_1 R_2 (R_3 + R_4) + R_3 R_4 (R_1 + R_2) + R_m (R_1 + R_2)(R_3 + R_4)} E_i
 \end{aligned}$$

On invoking the condition of initial equality of resistances in all arms, if R_1 changes to $R + \Delta R$, we get

$$\begin{aligned}
 i_m &= \frac{(R + \Delta R)R - R^2}{(R + \Delta R)R(2R) + R^2(2R + \Delta R) + R_m(2R + \Delta R)(2R)} E_i \\
 &= \frac{\Delta R/R^2}{2(1 + \Delta R/R) + (2 + \Delta R/R)(1 + 2R_m/R)} E_i \\
 &= \frac{\Delta R/R^2}{4(1 + R_m/R)} E_i \\
 &= \frac{\Delta R/R}{4(R + R_m)} E_i
 \end{aligned} \tag{16.3}$$

Therefore, the change in voltage across the measuring instrument is given by

$$\Delta E_o = i_m R_m = \frac{\Delta R/R}{4(1 + R/R_m)} E_i \quad (16.4)$$

Now, if it is a voltage-sensitive bridge which means that if the bridge output is fed to a CRO or a digital voltmeter having a very high input resistance (i.e. $R_m \rightarrow \infty$), Eq. (16.4) gives

$$\Delta E_o = \frac{\Delta R}{4R} E_i \quad (16.5)$$

while the output of a current-sensitive bridge which measures the output with the help of a finite resistance device, such as a galvanometer, is given by Eq. (16.3).

Calibration of the bridge

This is done by attaching a known resistance R_c along with a switch to R_2 in parallel (Fig. 16.2).

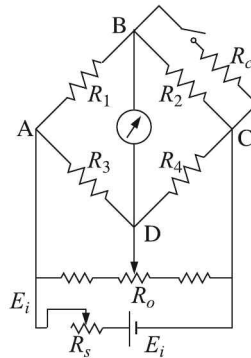


Fig. 16.2 Calibration of a Wheatstone bridge.

Initially, the switch is open and the bridge is balanced. Next the switch is closed when the voltmeter registers an output of, say, E_o volts. This output is caused by a change in resistance of

$$\Delta R = R_2 - \frac{R_2 R_c}{R_2 + R_c} = \frac{R_2^2}{R_2 + R_c}$$

Therefore, the sensitivity of the bridge S is given by

$$S = \frac{E_o}{\Delta R} = \frac{R_2 + R_c}{R_2^2} E_o \text{ V}/\Omega$$

The bridge sensitivity can be adjusted with the help of R_s . The zero adjustment, i.e. making the output zero with a zero input resistance, is made with the help of R_o which is a part of a parallel balancing arrangement.

The maximum supply voltage E_i is determined from the condition of maximum allowable self-heating of the transducers. Example 16.1 will clarify the requirement.

Example 16.1

A thermistor of resistance 1 k Ω , temperature coefficient of resistance 4.0 per cent/ $^{\circ}\text{C}$ and

having an internal temperature rise of $0.2\text{ }^\circ\text{C/mW}$ around $27\text{ }^\circ\text{C}$ is included in a dc bridge with three fixed resistors each of value $1\text{ k}\Omega$. A voltmeter of high input impedance is connected to the bridge output.

- Calculate the maximum bridge voltage and hence the maximum voltage sensitivity for the bridge if the internal temperature is not to exceed $0.1\text{ }^\circ\text{C}$.
- Calculate the maximum voltmeter drift, referred to the input, that can be tolerated if the overall system precision is to be within $\pm 2\text{ }^\circ\text{C}$.

Solution

- (a) Let E_i be the bridge supply voltage
 R_t be the thermistor resistance
 R be the resistance in other arms of the bridge ($= 1\text{ k}\Omega$)
 ΔR_t be the change in the thermistor resistance ($= R_t\alpha\Delta T$)
 E_o be the corresponding bridge output

Then assuming $\Delta R_t \ll R_t$, it can be seen that the current through the arms containing the thermistor is

$$i = \frac{E_i}{R_t + R} = \frac{E_i}{2R}$$

because $R_t = R$. Therefore, power dissipation through the thermistor is

$$i^2 R = \frac{E_i^2}{4R^2} R = \frac{E_i^2}{4R}$$

which raises the temperature of the thermistor to

$$\frac{E_i^2}{4R} \times 0.2 \times 10^3\text{ }^\circ\text{C}$$

Equating this temperature rise to the maximum allowable limit of $0.1\text{ }^\circ\text{C}$, we get after substituting the value of R

$$E_i = \sqrt{\frac{0.1 \times 4}{0.2}} V \simeq 1.4\text{ V}$$

According to Eq. (16.5) the bridge output corresponding to a temperature change of ΔT is

$$\Delta E_o = \frac{\Delta R_0}{4R} E_i = \frac{R_t \alpha \Delta T}{4R} E_i$$

Therefore, the maximum bridge sensitivity is

$$\frac{\Delta E_o}{\Delta T} = \frac{\alpha E_i}{4} \text{ (since } R_t = R) = \frac{0.045 \times 1.4}{4} \text{ V}/^\circ\text{C} = 15.75\text{ mV}/^\circ\text{C}$$

- (b) Thus, the maximum voltmeter drift for a system precision of $\pm 2\text{ }^\circ\text{C}$ is $4 \times 15.75\text{ mV} = 63\text{ mV}$.

AC Bridges

AC bridges comprise ac devices, namely inductances and capacitances, over and above the typical dc device—resistance. Many ac bridges are used for various purposes, such as measurement of inductance, capacitance, etc. Here we will consider only those applications which incorporate push-pull transducers.

Push-pull transducers work on a differential principle. For example, in inductance types if the input signal changes the inductance of one part from L to $L + \Delta L$, the other part is so coupled that its inductance becomes $L - \Delta L$. Such a transducer has the following advantages:

1. It produces a more linear relationship between the bridge output and the measurand than that produced by a single element transducer.
2. When connected to adjacent arms of an ac bridge, an automatic temperature compensation is achieved, because identical opposite changes in the two parts produce no effect on the bridge output.

A push-pull inductance transducer can be connected to an ac bridge in two ways. But before we analyse those possibilities, let us formulate the guiding conditions of an ac bridge.

In the bridge shown in Fig. 16.3, the following conditions will satisfy in the balanced condition:

$$i_1 = \frac{e_i}{Z_1 + Z_2} \quad (16.6)$$

$$i_2 = \frac{e_i}{Z_3 + Z_4} \quad (16.7)$$

$$i_1 Z_1 = i_2 Z_3 \quad (16.8)$$

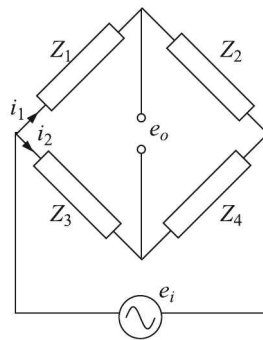


Fig. 16.3 A generalised ac bridge.

On substituting Eqs. (16.6) and (16.7) in Eq. (16.8), we get after simplification

$$Z_2 Z_3 = Z_1 Z_4 \quad (16.9)$$

Thus, Eq. (16.9) gives us the relation between the bridge components when it is in a balanced condition, i.e. $e_o = 0$.

As said before, the ac bridge with a push-pull inductance transducer can be constructed in two ways as shown in Fig. 16.4.

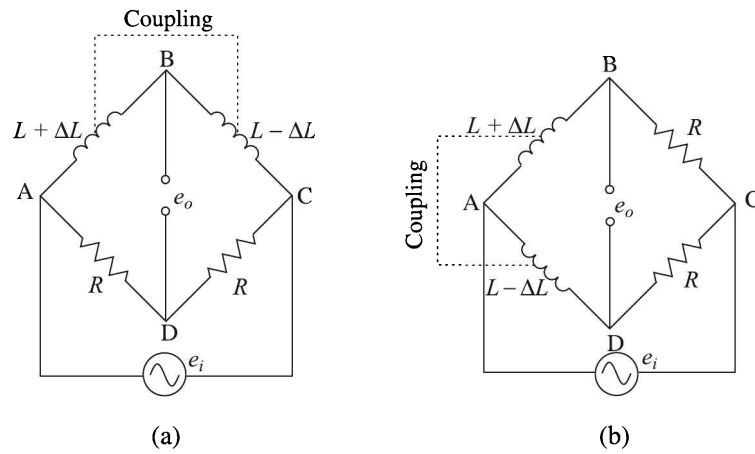


Fig. 16.4 AC bridge with a push-pull inductance transducer—two configurations.

In the first case,

$$e_B = \frac{j\omega(L + \Delta L)}{2j\omega L} e_i$$

$$e_D = \frac{R}{2R} e_i$$

Hence,

$$e_o = e_B - e_D = \frac{e_i}{2} \left[\frac{L + \Delta L}{L} - 1 \right]$$

$$= \frac{\Delta L}{2L} e_i$$

In the second case,

$$e_B = \frac{R}{R + j\omega(L + \Delta L)} e_i$$

$$e_D = \frac{R}{R + j\omega(L - \Delta L)} e_i$$

Therefore,

$$e_o = e_B - e_D = R \left[\frac{1}{R + j\omega(L + \Delta L)} - \frac{1}{R + j\omega(L - \Delta L)} \right] e_i$$

$$= R \left[\frac{2j\omega\Delta L}{R^2 + 2j\omega RL - \omega^2 L^2 + \omega^2 (\Delta L)^2} \right] e_i$$

$$\approx \frac{\Delta L}{L} e_i$$

assuming $R = \omega L$ and neglecting the term in $(\Delta L)^2$. Therefore, under similar conditions the output of the latter bridge circuit is double that of the former.

Blumlein bridge

Inductance type. This bridge which can be used for a push-pull inductive transducer uses two closely coupled inductive ratio arms, as shown in Fig. 16.5(a).

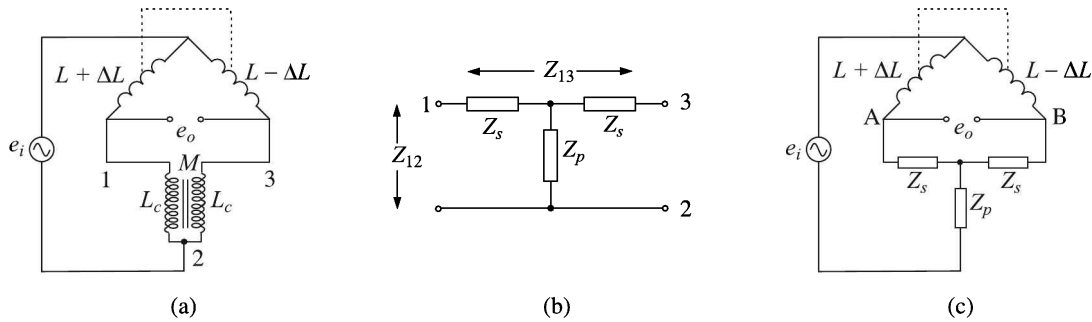


Fig. 16.5 (a) Blumlein bridge, (b) T-type network, equivalent to the closely-coupled ratio arms of the Blumlein bridge, and (c) Blumlein bridge with equivalent T-type network.

If the T-type network, shown in Fig. 16.5(b) is equivalent to the closely coupled ratio arms represented by 1–2–3, we have

$$Z_{12} = Z_s + Z_p \equiv j\omega L_c \quad (16.10)$$

$$Z_{13} = 2Z_s \equiv 2(j\omega L_c - j\omega M) \quad (16.11)$$

for a particular winding sense of the ratio arms. From Eqs. (16.10) and (16.11), we get

$$Z_s = j\omega(L_c - M)$$

$$Z_p = j\omega M$$

If k is the coupling factor between the two coils, $M = \pm kL_c$ and for tight coupling $k = \pm 1$.

When the bridge is balanced, the current through the ratio arms is of the same magnitude and direction, and for the aforesaid particular winding sense $k = +1$. Then, $Z_s = 0$ which means there is no voltage drop across the ratio arms.

Note: This result is important because it implies that any stray or cable capacitance which would be in parallel to the ratio arm will not affect measurements in any way, thus simplifying shielding and earthing requirements.

When the bridge is off balance, the instantaneous potentials at 1 and 3 are not equal and hence a current circulates around the bridge. Under such condition the coupling factor changes its sign to become -1 . As a result $Z_s = 2j\omega L_c$. To work out the sensitivity of the bridge, we note from Fig. 16.5(c) that if Z_B is the equivalent impedance of the bridge, the bridge supply voltage, e'_i , is

$$e'_i = \frac{Z_B}{Z_B + Z_p} e_i \quad (16.12)$$

Now, Z_B consists of two parallel arms—the inductance of one arm is $j\omega(L + \Delta L) + Z_s$ while that of the other is $j\omega(L - \Delta L) + Z_s$. Thus,

$$\begin{aligned} Z_B &= \frac{[j\omega(L + \Delta L) + Z_s][j\omega(L - \Delta L) + Z_s]}{[j\omega(L + \Delta L) + Z_s] + [j\omega(L - \Delta L) + Z_s]} \\ &= \frac{j\omega(L + \Delta L + 2L_c)(L - \Delta L + 2L_c)}{2(L + L_c)} \end{aligned} \quad (16.13)$$

Equation (16.13) coupled with Eq. (16.12) and $Z_p = -j\omega L_c$ gives

$$e'_i = \frac{(L + \Delta L + 2L_c)(L - \Delta L + 2L_c)}{(L + \Delta L + 2L_c)(L - \Delta L + 2L_c) - 2L_c(L + 2L_c)} e_i \quad (16.14)$$

The bridge output is given by

$$\begin{aligned} e_o = e_A - e_B &= \left[\frac{L + \Delta L}{L + \Delta L + 2L_c} - \frac{L - \Delta L}{L - \Delta L + 2L_c} \right] e'_i \\ &= \frac{4L_c \Delta L}{(L + \Delta L + 2L_c)(L - \Delta L + 2L_c)} e'_i \\ &= \frac{4L_c \Delta L}{(L + \Delta L + 2L_c)(L - \Delta L + 2L_c) - 2L_c(L + 2L_c)} e_i \quad [\text{using Eq. (16.14)}] \\ &= \frac{\Delta L}{L} \cdot \frac{4L_c/L}{\left(1 + \frac{\Delta L}{L} + \frac{2L_c}{L}\right) \left(1 - \frac{\Delta L}{L} + \frac{2L_c}{L}\right) - \frac{2L_c}{L} \cdot \left(1 + \frac{2L_c}{L}\right)} e_i \\ &\simeq \frac{\Delta L}{L} \cdot \frac{4L_c/L}{[1 + (2L_c/L)]} e_i \quad [\text{neglecting } (\Delta L/L)^2] \end{aligned} \quad (16.15)$$

Therefore, the bridge-sensitivity factor S_B is given by

$$S_B = \frac{e_o}{e_i(\Delta L/L)} = \frac{4L_c/L}{1 + 2(L_c/L)} \quad (16.16)$$

In both the coupled and uncoupled cases, S_B is independent of the frequency of the supply. From the plot of S_B vs. L_c/L (Fig. 16.6) it is evident that

1. the bridge with coupled ratio arms has a higher sensitivity for all L_c/L ;
2. for $L_c/L \gg 1$, $S_B = 2$ and, therefore, it is independent of the variations of ratio arm inductance.

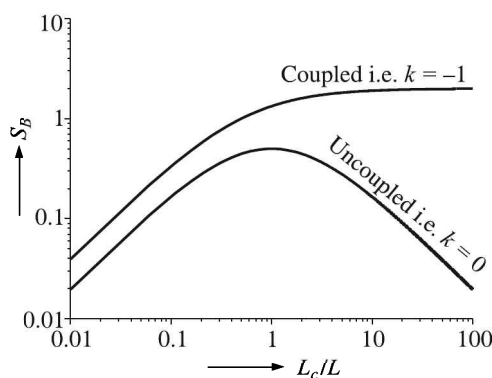


Fig. 16.6 Sensitivity plot of Blumlein bridge.

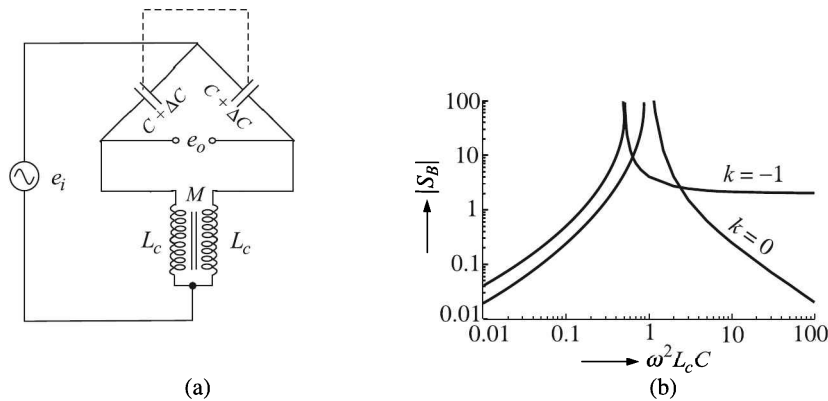


Fig. 16.7 (a) Blumlein bridge with push-pull capacitance transducers, and (b) its sensitivity plot.

Capacitance type. A Blumlein bridge having push-pull capacitance transducers, as shown in Fig. 16.7(a), is very useful. A similar analysis to that for the inductance-type shows that the output voltage and bridge-sensitivity factor for the coupled case are given by

$$e_o = \frac{\Delta C}{C} \cdot \frac{4\omega^2 L_c C}{(2\omega^2 L_c C - 1)} e_i$$

$$S_B = \frac{e_o}{e_i(\Delta C/C)} = \frac{4\omega^2 L_c C}{2\omega^2 L_c C - 1}$$

while that for the uncoupled (i.e. $k = 0$) case is given by

$$S_B = \frac{2\omega^2 L_c C}{(1 - \omega^2 L_c C)^2} \quad (16.17)$$

The S_B vs. $\omega^2 L_c C$ plot for both the coupled and uncoupled cases is shown in Fig. 16.7(b).

From this analysis it is evident that for

| <i>Coupled case</i> | <i>Uncoupled case</i> |
|---|---|
| 1. Resonance occurs at $\omega^2 L_c C = 1/2$ | 1. Resonance occurs at $\omega^2 L_c C = 1$ |
| 2. Beyond resonance, when $\omega^2 L_c C \gg 1$, $S_B = 2$ and therefore, it is independent of the variations of the bridge frequency and the ratio-arm inductances | 2. There is no region in the characteristics curve where the sensitivity does not vary with frequency or inductance |

To sum up, the Blumlein bridge with tightly coupled ratio arms

1. Picks up less noise, and
2. Offers greater constancy of sensitivity and higher sensitivity than a bridge with uncoupled inductive or resistive ratio arms.

Note: This analysis is true only if Q factors of L and L_c are high.

Linearisation by Bridge Circuits

Transducers, having nonlinear characteristics, can be compensated for by a nonlinear element, such as a deflection bridge, to produce more or less a linear output. For example, the temperature-resistance characteristic of a thermistor is nonlinear as shown in Fig. 16.8 (a). If the thermistor forms one arm of a dc Wheatstone bridge which has nonlinear resistance vs. output characteristic as shown in Fig. 16.8 (b), the combination can be made to generate a nearly linear temperature vs. voltage characteristic as shown in Fig. 16.8 (c).

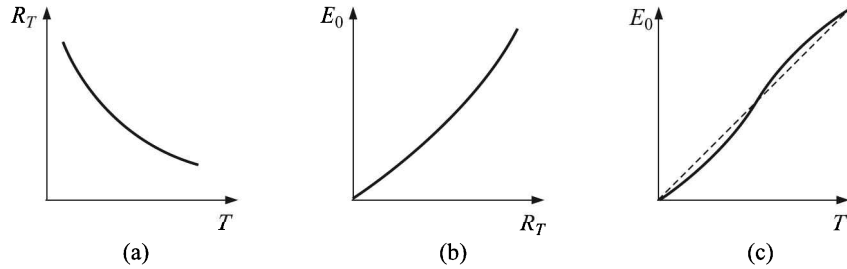


Fig. 16.8 (a) Thermistor characteristic, (b) Wheatstone bridge characteristic, and (c) output voltage vs. temperature curve of the bridge with thermistor in one arm.

Let us consider a practical example and see how the goal can be achieved. We have seen in Eq. (16.2) that the ratio of the output and input voltages of the Wheatstone bridge is given by

$$\begin{aligned} \frac{E_o}{E_i} &= \frac{R_1}{R_1 + R_2} - \frac{R_3}{R_1 + R_4} \\ &= \frac{1}{1 + (R_2/R_1)} - \frac{1}{1 + (R_4/R_3)} \end{aligned} \quad (16.18)$$

We consider the case when R_1 of the bridge is the temperature sensing resistance of the thermistor. We denote it as R_T , while other resistors are of fixed values. So, we rewrite Eq. (16.18) as

$$\frac{E_o}{E_i} = \frac{1}{1 + (R_2/R_T)} - \frac{1}{1 + (R_4/R_3)} \quad (16.19)$$

The bridge is initially balanced at a fixed reference temperature, say 0°C , when the output voltage is zero and the corresponding value of R_T is R_0 . Then, we have from the balancing condition

$$\frac{R_2}{R_0} = \frac{R_4}{R_3}$$

or

$$R_2 = \frac{R_4}{R_3} R_0 \quad (16.20)$$

Substituting this value of R_2 in Eq. (16.19) we get

$$\frac{E_o}{E_i} = \frac{1}{1 + \frac{R_0}{R_T} \cdot \frac{R_4}{R_3}} - \frac{1}{1 + \frac{R_4}{R_3}} \quad (16.21)$$

Now, let us put $x = R_T/R_0$ and $r = R_4/R_3$. Then, Eq. (16.21) becomes

$$\frac{E_o}{E_i} = \frac{1}{1 + (r/x)} - \frac{1}{1 + r} = \frac{x}{x + r} - \frac{1}{1 + r} \quad (16.22)$$

The E_o/E_i vs. x curves for different values of r are shown in Fig. 16.9. It can be seen that an appropriate nonlinear curve can be found that will compensate for the otherwise nonlinear characteristics curve of the thermistor. We better consider an actual situation.

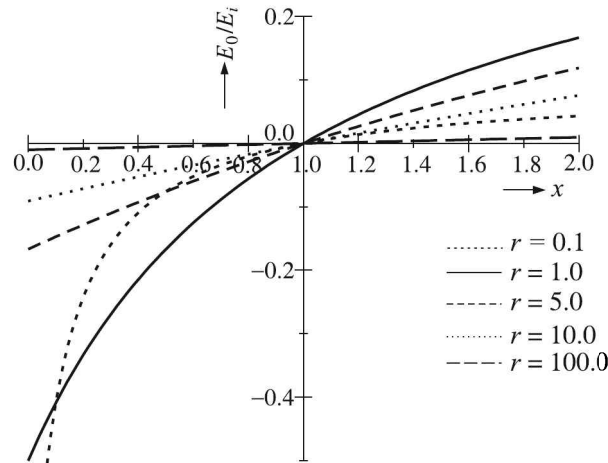


Fig. 16.9 E_o/E_i vs. x curves for different values of r .

We have seen in Example 10.6 that the thermistor resistance is $4 \text{ k}\Omega$ at the ice-point and 800Ω at 50°C . Corresponding x values are $1 (= 4000/4000)$ and $0.2 (= 800/4000)$. Suppose, we want an output of 0 V and 1.0 V from the bridge corresponding to temperatures 273 K and 323 K respectively, and that the bridge output should be nearly linear for all temperatures in this range, then at the mid-point, i.e. at $T = 298 \text{ K}$, we should have $E_o = 0.5 \text{ V}$. Thus, we get the data of Table 16.1 for the three temperatures, the mid-point thermistor resistance having been calculated from the relation obtained in Example 10.6.

Table 16.1 Data for three temperatures

| $T(\text{K})$ | $E_o(\text{V})$ | $R_T(\Omega)$ | x |
|---------------|-----------------|---------------|--------|
| 273 | 0.0 | 4000.0 | 1.0 |
| 298 | 0.5 | 1672.12 | 0.4784 |
| 323 | 1.0 | 800.0 | 0.2 |

Substituting appropriate values from Table 16.1 in Eq. (16.22) we get the following three equations

$$0 = \frac{1}{1 + r} - \frac{1}{1 + r} \quad (16.23)$$

$$\frac{0.5}{E_i} = \frac{0.4784}{0.4784 + r} - \frac{1}{1 + r}$$

$$\frac{1}{E_i} = \frac{0.2}{0.2 + r} - \frac{1}{1 + r}$$

Of the three, Eq. (16.23) is trivial. Solving the other two, we get $r = 0.7158$ and $E_i = -2.744$ V. So, from Eq. (16.20), we get

$$R_2 = rR_0 = 2863.2 \Omega$$

So, we have designed the compensating Wheatstone bridge, output of which can be obtained from Eq. (16.19) as

$$E_o = -2.744 \left[\frac{1}{1 + (2863.2/R_T)} - \frac{1}{1.7158} \right] = 1.5993 - \frac{2.744}{1 + (2863.2/R_T)} \quad (16.24)$$

Substituting the relation for R_T obtained from Example 10.6 in Eq. (16.24), we get

$$E_o = 1.5993 - \frac{2.744}{1 + \frac{2863.2}{\exp\{[(1/T) - 1] \div B\}}} \quad (16.25)$$

where $A = 7.4084 \times 10^{-4}$ and $B = 3.5232 \times 10^{-4}$. The plot of Eq. (16.25), given in Fig. 16.10, shows that E_o vs. T curve is nearly linear in the given temperature range.

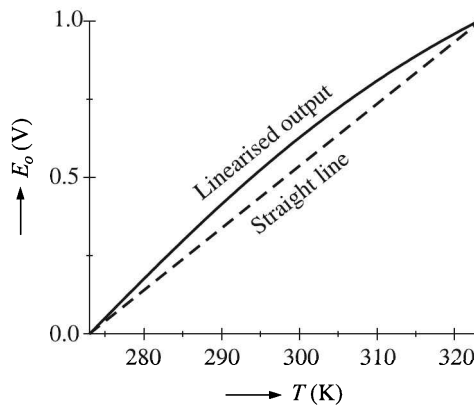


Fig. 16.10 E_o vs. T plot of Eq. (16.25).

It may be mentioned here that by choosing a higher value of r , rather than limits of E_o , might have yielded a better linearity. But then, a higher E_o and consequently a higher current flow through the thermistor might have resulted. That would be detrimental to the temperature measurement process because the hot thermistor would upset the temperature of the measurand.

Cold Junction Compensation of Thermocouple

As discussed in Section 10.4 (page 409), we will now see how a bridge circuit can be utilised to generate compensating voltage for the cold junction of thermocouples.

Let the cold junction temperature be T for which the required compensating voltage is E_o and let T have a small variation that will require E_o to vary as

$$E_o = kT \quad (16.26)$$

where k is a constant. This variation in E_o can be achieved by incorporating an RTD in one of the arms of a Wheatstone bridge as shown in Fig. 16.11. For small temperature changes, the resistance-temperature relation of the RTD can be written as

$$R_T = R_0(1 + C_1T)$$

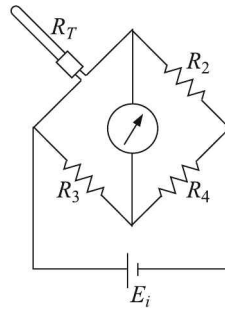


Fig. 16.11 Wheatstone bridge with RTD in one arm.

where R_0 is the resistance of the RTD at 0°C and C_1 is a constant. If, as before, we define $x = R_T/R_0$, its value is nearly 1 because $C_1 = 3.92 \times 10^{-3}/^\circ\text{C}$. Therefore, if a large value of r , say 100, is chosen, Eq. (16.22) reduces to

$$\frac{E_o}{E_i} = \frac{1}{1+r}(x-1) \cong \frac{1}{r}(x-1) \quad [\because r \gg 1]$$

which means,

$$\begin{aligned} E_o &= E_i \cdot \frac{R_3}{R_4} \left(\frac{R_T}{R_0} - 1 \right) \\ &= E_i \cdot \frac{R_3}{R_4} [(1 + C_1T) - 1] \\ &= E_i \cdot \frac{R_3}{R_4} \cdot C_1T \equiv kT \end{aligned}$$

Thus, the bridge that produces the desired output as postulated in Eq. (16.26), is the required cold junction compensation.

16.2 Conditioning Processes

Signal conditioning processes indeed constitute a vast discipline encompassing almost the entire gamut of analogue and digital electronics. Once the analogue signal is suitably conditioned, it is best converted to a digital signal so that it can be processed by a digital computer and analysed. We will not, however, consider signal conditioning at length but will touch different topics to gain an idea about the conditioning processes and circuitry involved.

A transducer detects a measurand and converts it into an electrical signal. If the signal is not strong enough to be displayed by a display device, such as a galvanometer, it has to be

amplified. Thus, amplification constitutes signal conditioning here. Signal conditioning may involve many more processes which can be broadly divided into two categories as shown in Table 16.2.

Table 16.2 Two categories of signal conditioning processes.

| <i>Linear processes</i> | <i>Nonlinear processes</i> |
|---|--|
| Amplification, attenuation, integration, differentiation, addition and subtraction. | Modulation, demodulation, sampling, filtering, clipping and clamping, squaring, linearising or multiplication by another function. |

Amplification

A display device draws energy from the measuring circuit itself, thus often loading the circuit and lowering the value of the measurand. Also in many applications the transducers do not generate enough power to drive display devices. Hence is the necessity of instrumentation amplifiers which generate enough power to drive display devices on one hand and protect the measurand from being loaded on the other hand. The desirable properties of an instrumentation amplifier are enumerated below:

Input impedance. The amplifier should have a high input impedance so that the transducer stage is not loaded. The rule of thumb is to choose the input impedance of the amplifier as 10 times the transducer impedance.

Output impedance. A low output impedance is desirable in order to minimise loading of the amplifier by the subsequent display/recording device. However, in this context one has to consider the output drive capability of the amplifier. An example will make the situation clear.

Example 16.2

A galvanometer has a sensitivity of 0.2 cm/mA and a span of ± 10 cm. If the output of an amplifier delivering ± 5 V has to be displayed by this galvanometer with an error of less than $\pm 0.2\%$, find the output resistance of the amplifier.

Solution

The span and sensitivity of the galvanometer suggest that for maximum readability, the variation of current should be $(\pm 10 \div 0.2)$ mA = ± 50 mA. Thus the galvanometer resistance R_L should be

$$R_L = \frac{5}{50 \times 10^{-3}} \Omega = 100 \Omega$$

With a 0.2% loading, the source resistance, i.e. the amplifier output resistance should be

$$0.998 = \frac{1}{1 + (R_o/R_L)} \cong 1 - \frac{R_o}{R_L}$$

or
$$\frac{R_o}{R_L} = 0.002$$

or
$$R_o = 100 \times 0.002 = 0.2 \Omega$$

Gain and frequency response. There is always an attenuation at the input of an amplifier depending on the output impedance of the source and the input impedance of the amplifier. The effective amplification obtained is thus the product of the gain (or gain factor) of the amplifier and the attenuation at its input. For a faithful measurement of a time-varying quantity, this gain of the amplifier should be stable.

The term frequency response actually refers to the variation of the gain of an amplifier with frequency of the input signal. In ac amplifiers the frequency response curve looks like Fig. 16.12(a) where the frequency range over which the flat response occurs depends upon the value of the load impedance Z_L .

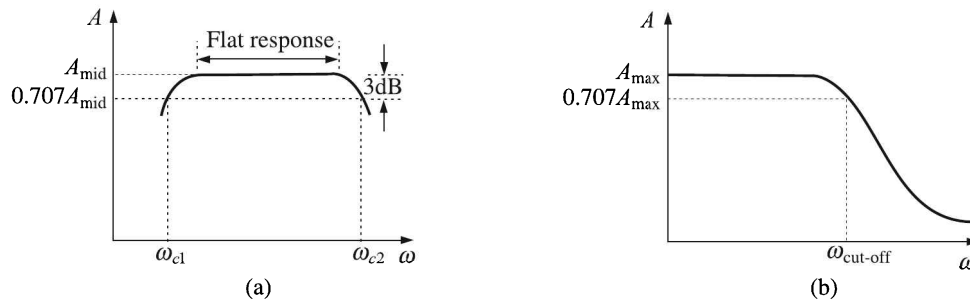


Fig. 16.12 Frequency response: (a) ac amplifier, and (b) dc amplifier.

The gain corresponding to the flat response is called midband gain, A_{mid} . Thus an ac amplifier in effect acts as a band-pass filter¹ having a band-width of $(\omega_{c2} - \omega_{c1})$. As is well-known, the 3 dB points correspond to a voltage change of 70.7% and a power change of 50%. The ac amplifier is evidently free from drift because a slowly varying input does not produce any output.

The frequency response of a dc amplifier on the other hand, corresponds to that of low-pass filter [Fig. 16.12(b)] by which low frequency signals up to a cut-off frequency are amplified faithfully. This property of a dc amplifier makes it susceptible to a major source of error called *drift*.

Noise. Noise is unwanted signal which contaminates the true signal. All active devices generate some noise due to various reasons depending upon the frequency f of the input signal. And since amplifiers are made of active devices, their performance is limited by such noises.

Operational Amplifier

The type of amplifier which finds applications in all kinds of measurement and instrumentation is the operational amplifier (or op-amp). It was originally designed to perform mathematical operations such as addition, sign-changing, integration, differentiation, etc. in analogue computers and hence the name op-amp. Basically, it is a very high-gain direct-coupled amplifier with a high input impedance and a low output impedance. Our purpose here is to provide an introduction to the basics of op-amp applications without delving deep into the subject.

¹See Section 16.2 at page 773.

An op-amp consists of a number of direct-coupled transistors, diodes, etc.², but for all practical purposes the details of the circuitry need not be considered and the device can be represented by a triangular symbol, as shown in Fig. 16.13, with two inputs V_1 (inverting) and V_2 (non-inverting), and an output V_o .

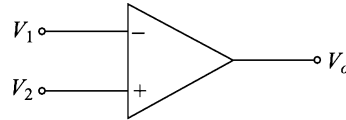


Fig. 16.13 Symbolic representation of an op-amp.

It is understood that all voltages are with respect to the ground and therefore the ground line is not usually shown. By definition,

$$V_o = A(V_2 - V_1)$$

where A is the voltage gain of the op-amp. A is also referred to as the *open-loop dc gain* and its value may lie anywhere between 80 dB (10,000) and 140 dB (10^7). That is why op-amps are used in circuits with considerable negative feedback.

Closed-loop voltage gain

The ratio V_o/V_i of an op-amp, with connections as shown in Fig. 16.14(a) is called its *closed-loop voltage gain*.

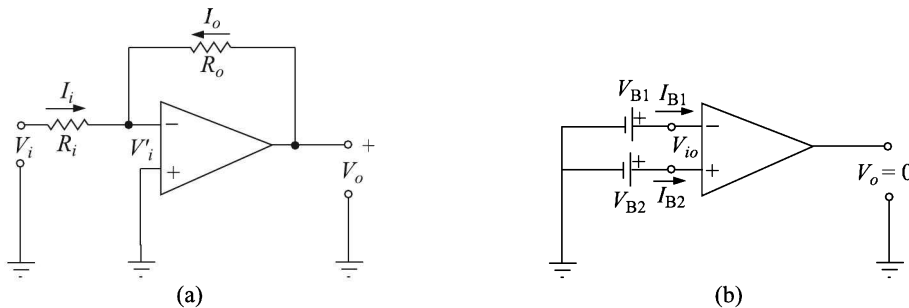


Fig. 16.14 (a) Closed-loop connection of an op-amp, and (b) offset voltage and bias currents of op-amp.

Here V_i may be a dc voltage or an ac signal within the bandwidth of the amplifier which may be as high as 1 MHz. Since the input impedance Z_i of the op-amp is very high, $I_i \cong I_o$. This prompts us to write

$$\frac{V_i - V_i'}{R_i} = \frac{V_i' - V_o}{R_o}$$

²See, for example, *Electronic Devices and Circuits* A Mottershead, Prentice Hall of India (1991) p 413, Fig. 23-13 for the circuit of an op-amp.

or

$$\begin{aligned} \frac{V_o}{R_o} &= V_i' \left(\frac{1}{R_o} + \frac{1}{R_i} \right) - \frac{V_i}{R_i} \\ &= -\frac{V_o}{A} \left(\frac{1}{R_o} + \frac{1}{R_i} \right) - \frac{V_i}{R_i} \quad [\text{substituting } V_o = -AV_i'] \end{aligned}$$

An ideal op-amp is perfectly balanced, that is $V_o = 0$ when $V_1 = V_2 = 0$. A real op-amp, however, exhibits an unbalance caused by a mismatch of input transistors within the op-amp circuitry. Often, an input offset voltage applied between the two input terminals is required to balance the amplifier in order to drive the output voltage V_o to zero.

The idealised mode of the op-amp must be modified to include the offset voltage and bias currents as depicted in Fig. 16.14(b). The main specifications used to describe op-amp performance are given in Table 16.3.

Table 16.3 Specifications used to describe op-amp performance

| Specification | Definition | Typical data | |
|--|---|--------------|--------------|
| | | BJT 741 | BIFET AD 611 |
| Input offset voltage, V_{io} (mV) | Voltage which must be applied between the input terminals to balance the amplifier. | ≤ 5 | ≤ 0.5 |
| Input offset current, I_{io} (nA) | The difference between the separate currents entering the input terminals of a balanced op-amp. According to Fig. 16.14, $I_{io} \equiv I_{B1} - I_{B2}$ when $V_o = 0$. | ≤ 200 | ≤ 0.010 |
| Input bias current, I_B (nA) | The mean of the separate currents entering the input terminals of a balanced op-amp. Put mathematically, $I_B \equiv (I_{B1} + I_{B2})/2$ when $V_o = 0$. | ≤ 500 | ≤ 0.025 |
| Full-power bandwidth (kHz) | If ΔV is the maximum output swing that can be obtained without significant distortion (at a given load resistance), the maximum frequency at which a sinusoid of ΔV is obtained is called the full-power bandwidth. | 10 | 200 |
| Slew rate (V/ μ s) | The rate of change of the closed-loop op-amp output voltage under large signal conditions. | 0.5 | 13 |
| Unity gain frequency (MHz) | The frequency at which the open-loop gain of the amplifier becomes unity (i.e. 0 dB). | 1 | 2 |
| Common mode rejection ratio, CMRR (dB) | To be defined later. | 80 | 80 |

Typical op-amp applications

A few typical op-amp applications in instrumentation are considered here.

Inversion. Consider the arrangement as shown in Fig. 16.15. Here $R_i = R_o = 10\text{ k}\Omega$, say. Then $V_o = -V_i$.

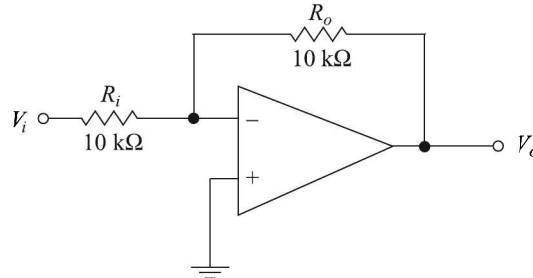


Fig. 16.15 Signal inversion.

Addition. The arrangement of Fig. 16.16 may be used to obtain an output which is a linear combination of a number of input signals.

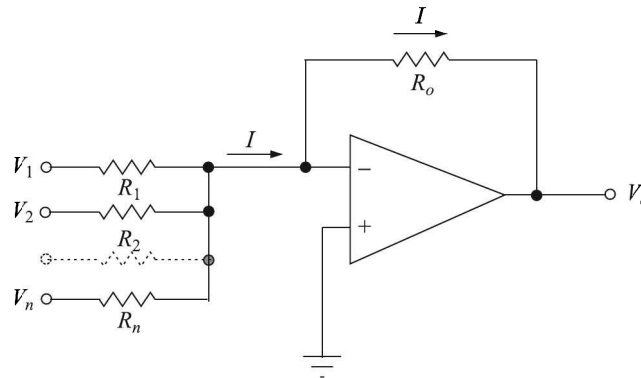


Fig. 16.16 Signal addition.

Since a virtual ground exists at the op-amp input, we have

$$I = \frac{V_1}{R_1} + \frac{V_2}{R_2} + \dots + \frac{V_n}{R_n}$$

or

$$V_o = R_o I = - \left(\frac{R_o}{R_1} V_1 + \dots + \frac{R_o}{R_n} V_n \right)$$

If $R_1 = R_2 = \dots = R_n = R_o$, then

$$V_o = - \sum V_i$$

In case non-inverting addition is desired, the resistance ladder may be connected to the non-inverting input with the inverting input grounded.

Subtraction. In the circuit of Fig. 16.17 we have the input voltage to the second stage V'_1 as

$$V'_1 = \frac{R'_o}{R_1} V_1$$

Hence the output after the second stage is given by,

$$V_o = - \left(\frac{R_o}{R_3} V_2 + \frac{R_o}{R_2} V_1' \right) = - \left(\frac{R_o}{R_3} V_2 - \frac{R_o}{R_2} \frac{R_1'}{R_1} V_1 \right)$$

If $R_1 = R_2 = R_3 = R_1' = R_o$, we get

$$V_o = V_1 - V_2$$

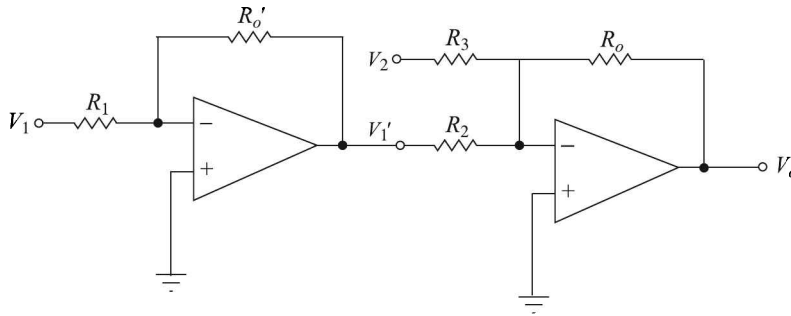


Fig. 16.17 Signal subtraction.

Multiplication and division. These operations can be performed by choosing suitable values of R_o and R_i in the basic op-amp configuration.

Integration. If R_o in the basic configuration is replaced by a capacitor C as shown in Fig. 16.18(a), the circuit behaves as an integrator, because here, $i_R = v_s/R$, and $i_C = Cdv_o/dt$ and since the current flow through the op-amp is negligibly small, $i_R = -i_C$. Hence,

$$\frac{v_s}{R} = -C \frac{dv_o}{dt}$$

or

$$v_o = -\frac{1}{RC} \int v_s dt$$

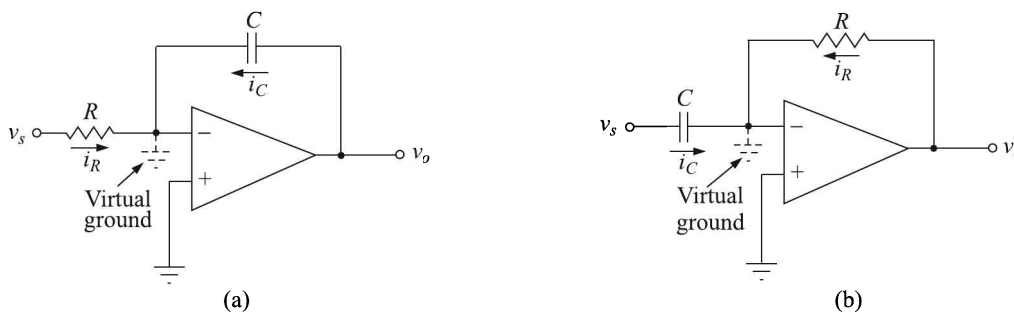


Fig. 16.18 Op-amp: (a) integrator, and (b) differentiator.

The amplifier, therefore, provides an output voltage which is proportional to the integral of the input voltage.

In case the input voltage is constant, i.e. $v_s = V$, the output will be a ramp $v_o = -Vt/RC$. Such an integrator makes an excellent sweep circuit for an oscilloscope and is called a *Miller Integrator* or *Miller Sweep*.

Another typical application of the integrator is found in capacitive displacement transducers to obtain linear relation between the input and the output³. The charge produced by piezoelectric transducers is also converted to a voltage with the help of the op-amp integrator.

Differentiation. If the positions of R and C in the previous circuit are interchanged, as shown in Fig. 16.18(b), the resultant circuit is a differentiator.

The inverting input of the op-amp being a virtual ground, we observe

$$i_C = C \frac{dv_s}{dt}$$

$$i_R = -\frac{v_o}{R}$$

Since, $i_C = i_R$, solving for v_o yields

$$v_o = -Ri_C = -RC \frac{dv_s}{dt}$$

It is interesting to note that if the input to the circuit is $v_s = A \sin \omega t$, then the output is $v_o = -ARC\omega \cos \omega t$. That means, the output amplitude increases linearly with frequency, i.e. the circuit has high gain at high frequencies. This property makes the circuit useful as *frequency to voltage converter*.

Voltage-to-current conversion. Consider the question of driving a deflection coil in a cathode ray tube. Here it is necessary to convert a voltage signal to a proportional current output. We consider two cases—(i) the load impedance Z_L has neither side grounded (i.e. it is floating) and (ii) Z_L is grounded. Two corresponding simple circuits are presented in Fig. 16.19.

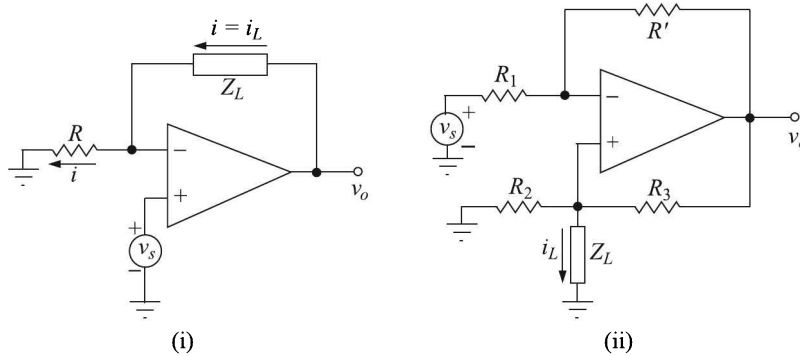


Fig. 16.19 Voltage-to-current conversion: (a) floating load, and (b) grounded load configurations.

In the first case, the inverting input being a virtual ground,

$$i_L(t) = i(t) = \frac{v_s(t)}{R}$$

³See Section 6.2 at page 187.

Since the same current flows through the signal source and the load, the signal source should be capable of providing the load current. It is also interesting to note that i is independent of Z_L .

In the second case it can be shown that if $R_3/R_2 = -R'/R_1$, then

$$i_L(t) = -\frac{v_s(t)}{R_2}$$

Current-to-voltage conversion. Figure 16.20 shows how an op-amp can be used as current-to-voltage converter.

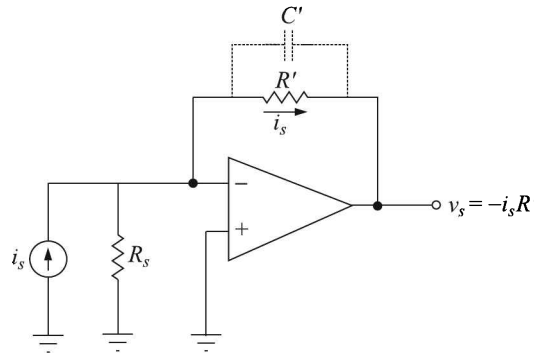


Fig. 16.20 Current-to-voltage conversion.

Owing to the existence of a virtual ground at the op-amp input, the current in R_s is zero and i_s flows through the feedback resistor R' making the output voltage $v_o = -i_s R'$. R' is generally shunted by a capacitor to reduce high-frequency noise and any possible oscillation. This circuit makes an excellent current-measuring instrument since it is an ammeter with zero voltage across the meter.

It finds application in measuring voltages generated by photocells and photomultiplier tubes.

Voltage follower or unity gain buffer. Suppose an op-amp is used as a non-inverting amplifier as shown in Fig. 16.21(a).

Here, making $V_i = 0$ requires that

$$V_i = \frac{R_1}{R_1 + R_2} V_o - V_s = 0$$

Substituting $A_v = V_o/V_s$, we get

$$A_v = \frac{R_1 + R_2}{R_1} = 1 + \frac{R_2}{R_1}$$

If $R_2 = 0$, R_1 is unnecessary and $A_v = 1$. The circuit, then having the appearance as in Fig. 16.21(b), is called a voltage follower or unity gain buffer. Such a circuit has a high input resistance, low output resistance and unity gain resembling an emitter- or source-follower.

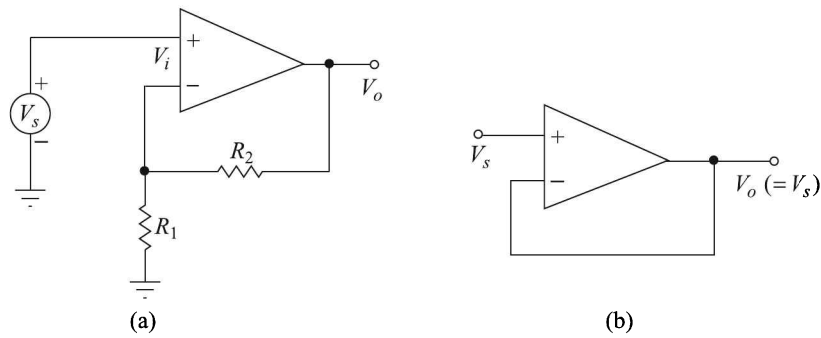


Fig. 16.21 (a) General non-inverting amplifier, and (b) voltage follower.

Differential amplifier

The use of an op-amp as a differential amplifier is very useful in instrumentation. Transducers, such as strain gauges, thermocouples and hot-wire anemometers generate a small difference signal which usually must be amplified. Instrumentation amplifiers provide an output that is a precise multiple of the difference between two input signals. A simple instrumentation amplifier can be constructed by using one op-amp as a differential amplifier as shown in Fig. 16.22(a). Here the output voltage V_o is given by

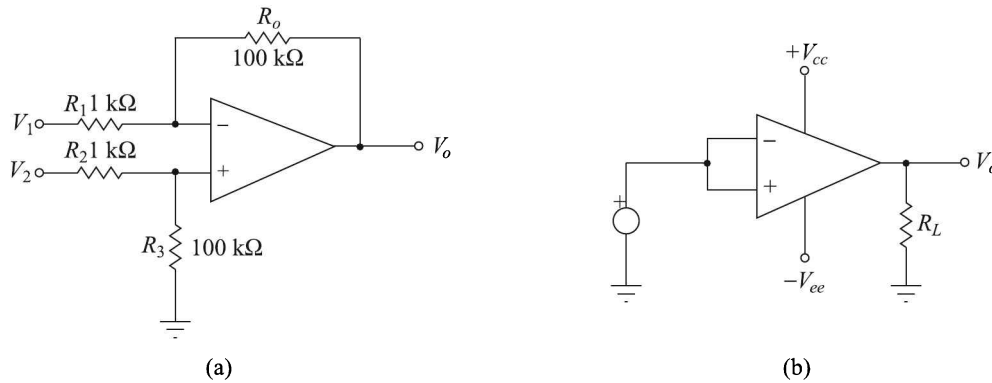


Fig. 16.22 (a) Differential amplifier, and (b) common-mode configuration.

$$V_o = A(V_+ - V_-)$$

where, A is the internal gain of the op-amp. An op-amp has a very high gain; $A \cong 10^5$ or more, which implies a small $V_+ - V_-$ for a finite V_o . Therefore, for all practical purposes, $V_+ \cong V_-$. Now,

$$V_1 = \frac{R_1}{R_1 + R_o} V_o$$

$$V_o = \frac{R_o}{R_1 + R_o} V_1$$

Hence we get by superposition

$$V_- = \frac{R_1}{R_1 + R_o} V_o + \frac{R_o}{R_1 + R_o} V_1$$

Also,

$$V_+ = \frac{R_3}{R_2 + R_3} V_2$$

Since $V_+ \cong V_-$, we have

$$\frac{R_1}{R_1 + R_o} V_o + \frac{R_o}{R_1 + R_o} V_1 = \frac{R_3}{R_2 + R_3} V_2$$

or

$$V_o = \frac{(R_o + R_1)R_3}{(R_2 + R_3)R_1} V_2 - \frac{R_o}{R_1} V_1 \quad (16.27)$$

Now

$$\begin{aligned} A_D &= \frac{R_o}{R_1} = \frac{R_3}{R_2} \quad (16.28) \\ &= \frac{R_o + R_1}{R_1} = \frac{R_2 + R_3}{R_2} \end{aligned}$$

or

$$\frac{R_o + R_1}{R_2 + R_3} = \frac{R_1}{R_2} \quad (16.29)$$

and therefore from Eq. (16.27) with the help of Eqs. (16.28) and (16.29), we get

$$V_o = A_D(V_2 - V_1) \quad (16.30)$$

A_D is called the *differential mode gain* of the op-amp. Note that it is equal to the internal gain A of the op-amp. Thus any difference signal ($V_2 - V_1$) is multiplied by the differential mode gain to produce an output.

Common-mode configuration. When the same input voltage is applied to both the input terminals of an op-amp, it is said to be operating in a common-mode configuration [Fig. 16.22(b)]. A common-mode voltage can be ac, dc or a mixture. Ideally, in such a case there should not be any output. But owing to imperfections in a practical op-amp, some common-mode output voltage V_{com} , however small, will appear.

Thus, there exists a common-mode voltage gain, A_{com} defined by

$$A_{com} = \frac{V_{o,com}}{V_{com}}$$

The common mode rejection ratio (CMRR) is defined as,

$$\text{CMRR} = \frac{\text{Differential mode gain}}{\text{Common-mode gain}} = \frac{A_D}{A_{com}}$$

Generally, the CMRR value is very large and is therefore conveniently specified in dB as

$$\text{CMRR(dB)} = 20 \log \frac{A_D}{A_{com}}$$

Instrumentation amplifier

The circuit in Fig. 16.22 represents a simple instrumentation amplifier, which we have already analysed as a differential amplifier. Here the output voltage is given by Eq. (16.30) as

$$V_o = \frac{R_o}{R_1}(V_2 - V_1)$$

But the problem with this simple instrumentation amplifier is that R_1 has to be kept low in order to have a high gain. A low R_1 , in turn, implies a low input impedance and low common mode rejection. These are in conflict with instrumentation amplifier requirements which are

1. High impedance
2. High CMRR
3. Low offset voltage
4. Low temperature coefficient of offset voltage

If the input to the differential amplifier of Fig. 16.22 is modified with the inclusion of two inverting op-amps as shown in Fig. 16.23, and if we make $R_1 = R_2 = R_3 = R_o$, the overall differential output becomes

$$V_o = \left(1 + 2\frac{R_i}{R_c}\right)(V_1 - V_2)$$

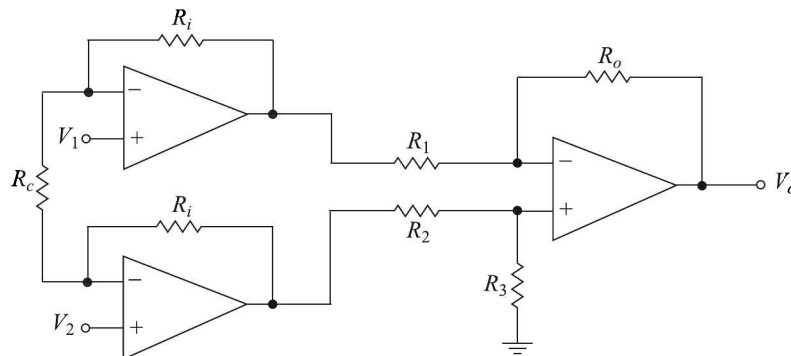


Fig. 16.23 Simple instrumentation amplifier.

The corresponding common mode gain = 1. The output op-amp then acts as an inverting unity gain differential amplifier and the combination will have all the desired attributes of an instrumentation amplifier.

Example 16.3

For the instrumentation amplifier shown in Fig. 16.24, verify that

$$V_o = \left[1 + \frac{R_2}{R_1} + \frac{2R_2}{R}\right](V_2 - V_1)$$

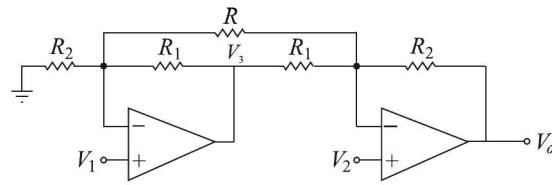


Fig. 16.24 Example 16.3.

Solution

We assume that the op-amps are ideal and so, $V_+ = V_- = V_1$. Now, applying KCL to the left op-amp, we get

$$\frac{V_1}{R_2} + \frac{V_1 - V_2}{R} + \frac{V_1 - V_3}{R_1} = 0$$

$$\Rightarrow V_1 \left[\frac{1}{R_1} + \frac{1}{R} + \frac{1}{R_2} \right] - \frac{V_2}{R} = \frac{V_3}{R_1}$$

This yields

$$V_3 = V_1 \left[1 + \frac{R_1}{R} + \frac{R_1}{R_2} \right] - \frac{V_2 R_1}{R} \quad (i)$$

Applying KCL to the right op-amp, we get

$$\frac{V_1 - V_2}{R} + \frac{V_3 - V_2}{R_1} + \frac{V_o - V_2}{R_2} = 0$$

$$\Rightarrow \frac{V_1}{R} - V_2 \left[\frac{1}{R} + \frac{1}{R_1} + \frac{1}{R_2} \right] + \frac{V_3}{R_1} = -\frac{V_o}{R_2}$$

Substituting the value of V_3 from Eq. (i) and on rearranging, we get

$$\begin{aligned} -\frac{V_o}{R_2} &= V_1 \left[\frac{1}{R_2} + \frac{1}{R_1} + \frac{2}{R} \right] - V_2 \left[\frac{1}{R_2} + \frac{1}{R_1} + \frac{2}{R} \right] \\ &= \left[\frac{1}{R_2} + \frac{1}{R_1} + \frac{2}{R} \right] (V_1 - V_2) \end{aligned}$$

$$\Rightarrow V_o = \left[1 + \frac{R_2}{R_1} + \frac{2R_2}{R} \right] (V_2 - V_1)$$

Modulation

The variation of a high-frequency carrier characteristic proportional to a lower frequency signal is called *modulation*. We know that a wave is characterised by three parameters, namely,

1. Amplitude
2. Frequency
3. Phase

All these parameters can be modulated to transmit information and hence we have amplitude modulation (AM), frequency modulation (FM) and phase modulation (PM) techniques. We explain them in that order.

Amplitude modulation (AM)

As the name suggests, here the signal modulates the amplitude of the carrier. The process is shown schematically in Fig. 16.25.

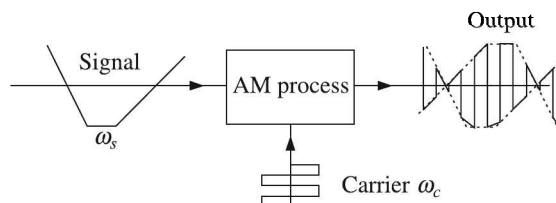


Fig. 16.25 Amplitude modulation. Note that the output amplitude is an envelope which consists of the signal and its mirror image.

To understand the process, we make a simple analysis with a sine-wave signal of angular frequency ω_s and a similar carrier of angular frequency ω_c so that

$$e_c = A_c \sin \omega_c t \quad (16.31)$$

where the terms have their usual significance, the subscript c denoting the carrier. In AM the value of A_c swings about its unmodulated value A_o and for a sinusoidal signal of angular frequency ω_s , A_c can be written as

$$A_c = A_o(1 + m \sin \omega_s t) \quad (16.32)$$

where, m is the modulation index, or, the degree of modulation varying between 0 and 1. If $m = 1$, it means that the modulation is 100%.

If the modulated output is designated as e_m , it can be written by combining Eqs. (16.31) and (16.32) as

$$\begin{aligned} e_m &= A_o(1 + m \sin \omega_s t) \sin \omega_c t \\ &= A_o(\sin \omega_c t + m \sin \omega_c t \sin \omega_s t) \\ &= A_o \left[\sin \omega_c t + \frac{m}{2} \{ \cos(\omega_c - \omega_s)t - \cos(\omega_c + \omega_s)t \} \right] \\ &= A_o \left[\sin \omega_c t + \frac{m}{2} \sin \{ (\omega_c - \omega_s)t + 90^\circ \} + \frac{m}{2} \sin \{ (\omega_c + \omega_s)t - 90^\circ \} \right] \end{aligned} \quad (16.33)$$

Equation (16.33) shows that

1. the output signal comprises three frequencies, namely $(\omega_c - \omega_s)$, ω_c , and $(\omega_c + \omega_s)$ [Fig. 16.26(a)];
2. the frequencies other than ω_c , called *side frequencies*, have the same amplitude of $(1/2) mA_o$
3. with respect to the input, the lower side frequency has a phase-shift of 90° while for the upper side frequency, it is -90° [Fig. 16.26(b)]
4. the bandwidth is $[(\omega_c + \omega_s) - (\omega_c - \omega_s)] = 2\omega_s$.

Note: A multiplier circuit can be utilised to produce AM signals.

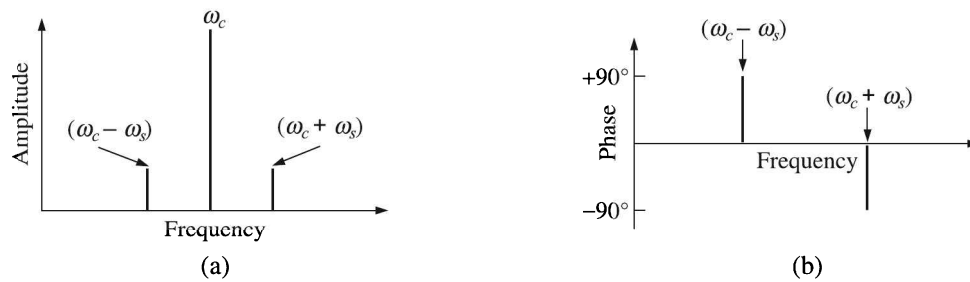


Fig. 16.26 AM signal: (a) frequency spectrum, and (b) phase shifts of different frequencies.

An application of AM. Amplitude modulation is typically used in amplifying small output signals from strain gauges. We have already seen that dc amplifiers suffer from drift while ac amplifiers are not suitable for amplifying slowly varying inputs because they cut off low frequency inputs.

Amplitude modulation comes in handy for such a situation where the slowly varying input modulates the amplitude of a 5 V, 5000 Hz supply voltage which is then conveniently amplified by a high-gain ac amplifier to produce a faithful output without any drift. The arrangement is shown schematically in Fig. 16.27.

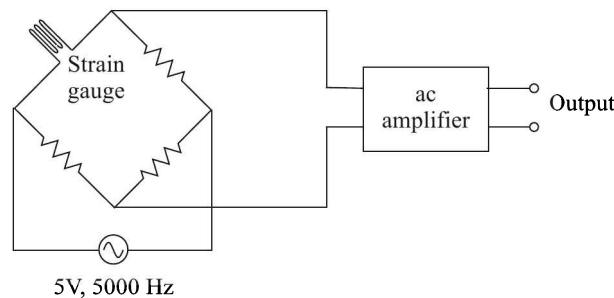


Fig. 16.27 Utilisation of AM to amplify strain-gauge signals.

If the frequency varies at the rate of 20 Hz owing to the variation of the strain-gauge resistance, the resulting modulated signal will have frequencies 4980 Hz and 5020 Hz which lie in the flat response region of an ac amplifier. Therefore, the output will have a faithful dynamic response.

The additional advantage of AM and ac amplifier is the rejection of low frequency noise which is generally picked up by connecting cables from the 50 Hz household supply voltage.

While utilising AM, one point that has to be kept in mind is that the carrier frequency has to be at least 10 times the signal frequency.

Chopper-stabilised dc amplifier. The concept of AM is utilised in constructing chopper-stabilised amplifiers to amplify dc inputs. A schematic diagram is shown in Fig. 16.28.

The dc input is alternately connected to A and B terminals of the switch S_1 which, in turn, is connected to the centre-tapped primary of a transformer. As a result the secondary of the transformer produces a square-wave ac signal proportional to the input dc. The resulting ac is amplified by an amplifier where the slowly varying dc component, or the drift of the input, gets

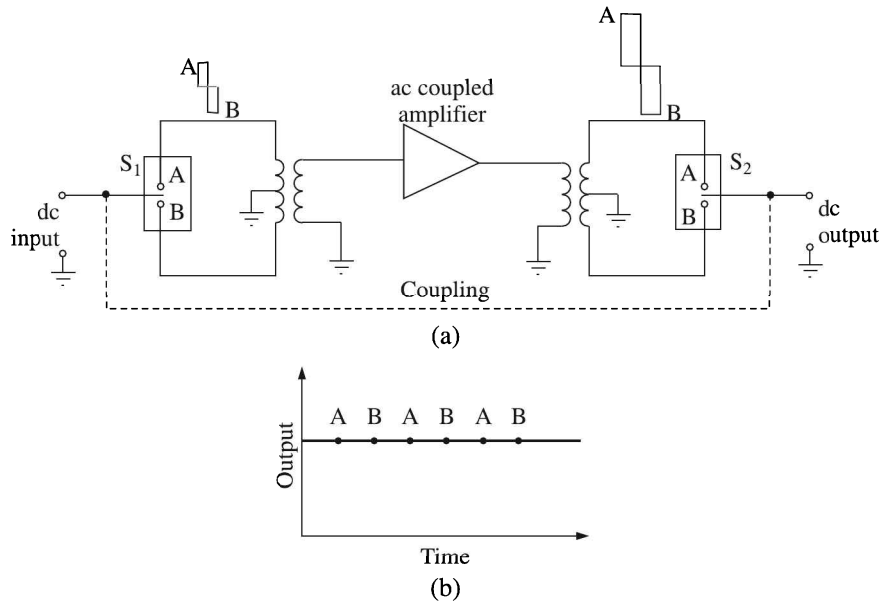


Fig. 16.28 Chopper-stabilised dc amplifier: (a) schematic circuit, and (b) output characteristic. A and B denote switch positions.

lost because of the frequency response of an ac amplifier which resembles that of a band-pass filter.

The amplified voltage is chopped by a mechanically coupled chopper so that it operates in synchronisation with the input chopping. As a result, the dc value of the input is restored but amplified by the ac gain.

Electromechanical choppers have rather short life span in comparison to electronic ones. BJTs, light-activated devices and FETs have been used to construct such devices. But considering the vital requirement of chopper that it must not inject any current into the circuit being chopped, MOSFETs are the most successful because they have no junction as a source of leakage current.

Amplitude modulation for digital signals. So far, we have talked about AM for analogue signals. AM can also be employed for transmitting digital signals. The AM outputs corresponding to two-level and four-level digital signals are shown in Fig. 16.29.

In two-level, each bit is transmitted separately, while in four-level, two consecutive bits are taken together and a pulse of corresponding level is transmitted. For example, 10 in binary equals 2 in decimal. Therefore, to transmit 10, a third level (0 corresponds to level 1) pulse is generated.

The speed of transmission can be increased by transmitting 4-levels instead of 2-levels as can be visualised from the diagram.

We have seen that AM produces two identical symmetrical side-bands around the carrier frequency. Since both side-bands contain the same information, it is good enough if only one of them is transmitted.

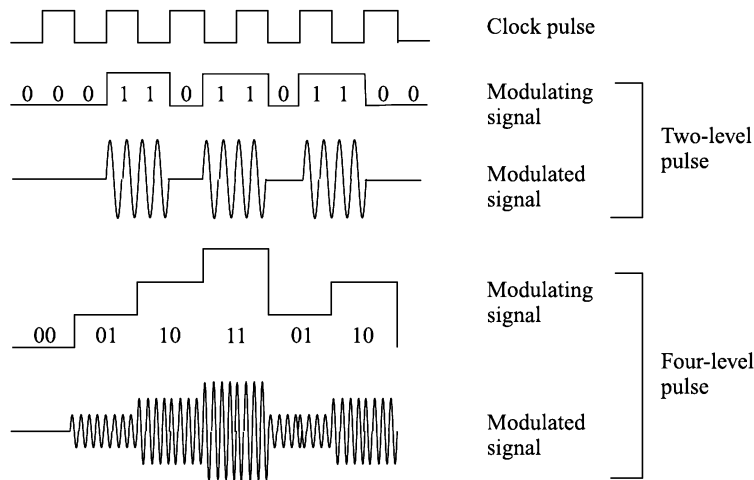


Fig. 16.29 AM digital signals—two-level and four-level pulses.

Frequency modulation (FM)

In this technique the instantaneous frequency of the carrier is made to vary in accordance with the amplitude of the signal, while the carrier amplitude and phase are held constant. A comparison of AM, FM and PM waveforms, shown schematically in Fig. 16.30, will help understand the difference between the three.

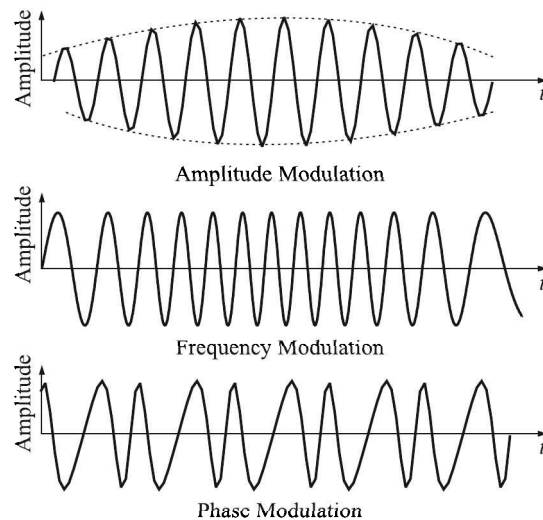


Fig. 16.30 Comparison between AM, FM and PM waveforms.

The simplest way to achieve FM for voice signals is to use a capacitor microphone in the tank circuit of an oscillator, as shown in Fig. 16.31(a). The vibrating diaphragm of the capacitor microphone changes its capacitance in accordance with the voice signal. This, in turn, changes the frequency of the oscillator.

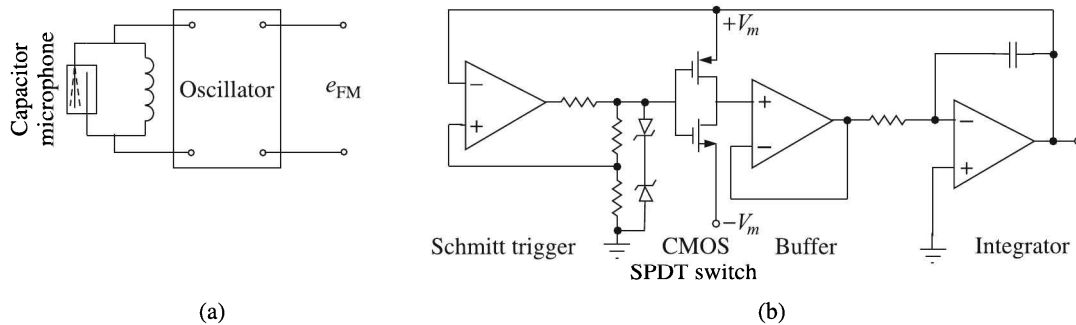


Fig. 16.31 A simple method to generate FM waveform: (a) from voice signals, and, (b) from voltage signals.

For voltage signals, a voltage-to-frequency converter circuit which is basically a voltage controlled oscillator (VCO), may be utilised to produce FM. Typically, a VCO comprises Schmitt trigger, a CMOS SPDT switch, a buffer and an integrator as shown in Fig. 16.31(b).

FM is widely used in digital data transmission and recording. We will take it up while discussing recording techniques.

Phase modulation

While in principle phase modulation can be used for analogue signals, in practice it is used for modulating digital signals. Usually it is known as 'phase-shift keying' (PSK). As the name suggests, here the phase of the carrier frequency is varied according to the data transmitted. Since both amplitude and frequency are held constant here, PSK is often employed in high-speed data transmission.

Modulating and modulated signals are schematically shown in Fig. 16.32 where 0s and 1s have 0° and 180° phase-shift respectively. In order that the receiver can identify the phase angles properly, it is necessary to provide some form of reference signal to the receiver.

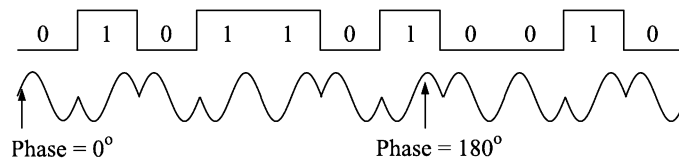


Fig. 16.32 A digital signal and its PSK equivalent. Note that each time the bit flips, the phase of the carrier flips.

Demodulation or Detection

A modulated signal cannot be used to drive the display system of an instrument. Hence is the necessity of converting it back to its original form, i.e. a variation of voltage, as produced by the transducer. The relevant process is called 'demodulation or detection'. Obviously, the process involves filtering the carrier frequency.

In AM, detection can be *half-wave* or *full-wave* and *phase-sensitive* or *non-phase-sensitive*. The meaning of these terms will be clear from Fig. 16.33.

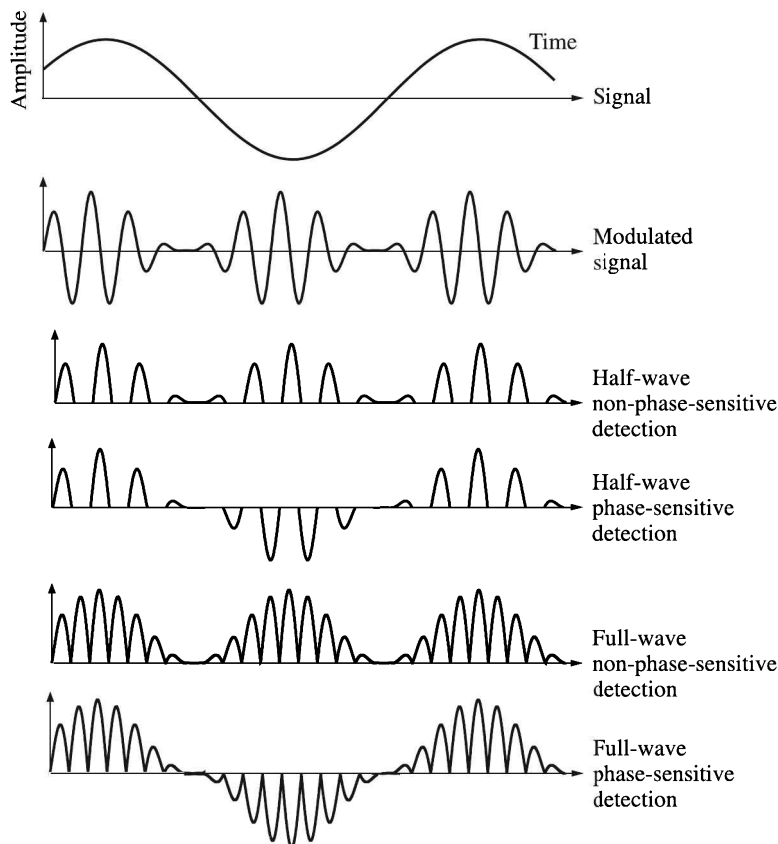


Fig. 16.33 An AM signal and the corresponding different types of detection.

In most of the applications, a phase-sensitive detection is necessary in order to recover fully the original magnitude as well as the sign of the voltage produced by the transducer. After detection the carrier frequency has got to be filtered from the signal. This requires a low-pass filter which is a device that allows low frequency signals to pass faithfully but attenuates high frequencies drastically. For successful operation of such a filter it is better that the frequency to be attenuated, i.e. the carrier, is widely separated from the signal. This is why the carrier frequency is chosen to be at least 5 to 10 times that of the signal.

A basic demodulator circuit can be constructed by extracting voltage from two oppositely charged matched capacitors which produce a null voltage if the input is zero (Fig. 16.34).

For a positive-going input, if the induced voltage across the secondary matches the phase of the oscillator signal, the upper capacitor will produce a higher voltage than that of the lower capacitor, and vice versa. In order that the sign of the original dc signal be recovered, the same reference oscillation which drives the modulator has to be used in the demodulation.

We discussed only the AM demodulation here. FM and PM demodulations will be considered in a subsequent chapter while discussing magnetic recording of signals.

Now, we will consider an interesting application of phase-sensitive detection, which we mentioned in Section 6.2 although without elaborating on what it meant.

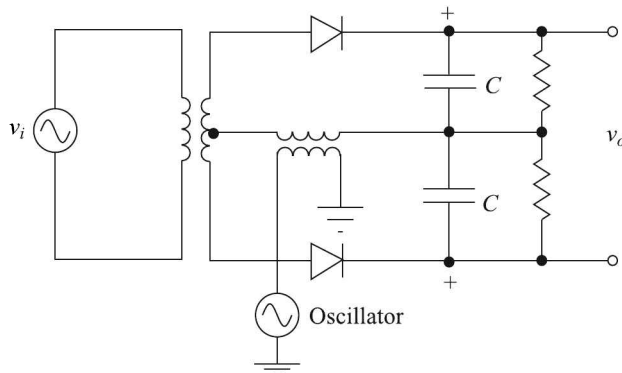


Fig. 16.34 A basic demodulator circuit.

An application of phase-sensitive detection

Consider the measurement of displacement by an LVDT (see Section 6.2). The output of the LVDT is a sine wave whose amplitude is proportional to the displacement of the core. We can measure the displacement by measuring the output with the help of a calibrated ac voltmeter. But such a measurement will not give any indication as regards the direction of displacement of the core from the null position. Of course, an oscilloscope may be used to detect that—but measurement with a CRO is an involved procedure and it is rather difficult to track rapid core movements in such measurements. Phase-sensitive detection is the solution for this kind of measurement as will be evident from Fig. 6.14.

Filters

In a measurement system, the signal is more often than not corrupted by unwanted noise generated by many factors. Therefore, to isolate the signal generated by the transducer from the noise, or in other words to improve the signal-to-noise ratio, it is necessary to use filters.

Consider a system with an input signal containing many components at different frequencies. A filter is necessary to separate a chosen band of frequencies from those present. As we have discussed in Section 4.1 (see page 80), such a situation is best described by considering the frequency transfer function $G(j\omega)$.

In terms of their response to frequencies, filters are ideally classified into four categories:

1. Low-pass
2. High-pass
3. Band-pass
4. Band-reject

Low-pass filter. A low-pass filter transmits all frequencies from zero (dc) to a pre-determined cut-off frequency ω_c without loss. For inputs with frequency components $\omega > \omega_c$, it gives a zero output [Fig. 16.35(a)]. That is,

$$|G(j\omega)| = \begin{cases} G_o & \text{for } \omega < \omega_c \\ 0 & \text{for } \omega > \omega_c \end{cases}$$

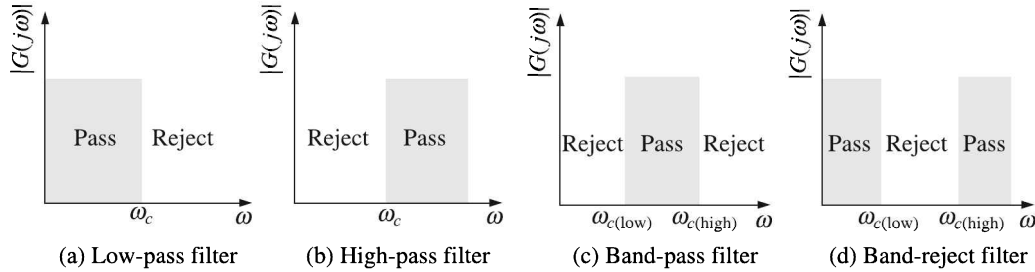


Fig. 16.35 Frequency domain transfer functions of different kinds of filters.

High-pass filter. A high-pass filter transmits all frequencies above a pre-determined cut-off frequency ω_c without loss. For inputs having $\omega < \omega_c$, it gives a zero output [Fig. 16.35(b)]. Stated symbolically,

$$|G(j\omega)| = \begin{cases} G_o & \text{for } \omega > \omega_c \\ 0 & \text{for } \omega < \omega_c \end{cases}$$

Band-pass filter. The characteristic of a band-pass filter is better stated symbolically as [Fig. 16.35(c) for a graphical presentation],

$$|G(j\omega)| = \begin{cases} G_o & \text{for } \omega_{c(\text{low})} < \omega < \omega_{c(\text{high})} \\ 0 & \text{for all other } \omega \end{cases}$$

Band-reject filter. A band-reject filter is the reverse of the band-pass filter [Fig. 16.35(d)] and its transfer function in the frequency domain can be written as,

$$|G(j\omega)| = \begin{cases} 0 & \text{for } \omega_{c(\text{low})} < \omega < \omega_{c(\text{high})} \\ G_o & \text{for all other } \omega \end{cases}$$

The characteristics as depicted in Fig. 16.35 are only ideal while practical circuits produce somewhat different responses. Figure 16.36 depicts a realistic low-pass response.

The indicated passband is the range of frequencies that is transmitted without excessive attenuation. Note that the amplitude ratio has a *ripple* of magnitude $(G_0 - G_1)$ and therefore, is not really constant, its typical value lies between 0.5 and 1 dB. The frequency ω_c for which $|G(j\omega)| = G_0 - 3$ (in dB) is often used to indicate the edge of the pass-band, i.e. the cut-off frequency. The meaning of other terms is apparent from the diagram. Note that the ripples may be present in the stopband as well.

Thus, a filter is specified by:

1. Cut-off frequency, ω_c i.e. the range of passband frequencies
2. Stopband attenuation $G_0 - G_2$
3. Stopband frequency range, which means specifying ω_s
4. Allowable passband ripple

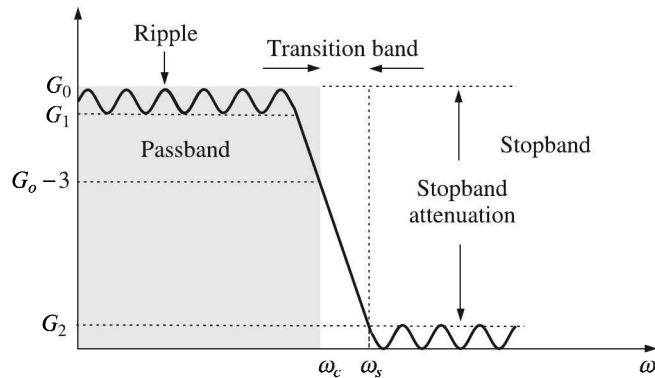


Fig. 16.36 Frequency response curve of a realistic low-pass filter (ripple is shown in an exaggerated way).

Filters are also classified as

1. Active, and
2. Passive

Active filters are a class of frequency-selective circuits employing resistors, capacitors and op-amps. As modern IC fabrication precludes the use of inductors, such filters are of advantage particularly for frequencies below about 100 kHz.

Passive filters use passive components like inductors, resistors and capacitors, though inductors are generally avoided because they are not only bulky, heavy and nonlinear, but also generate stray magnetic fields. We will now discuss the fabrication of passive filters to understand their application in signal conditioning.

Low-pass RC filter

A simple low-pass RC filter and its frequency response are shown in Fig. 16.37. We can at once

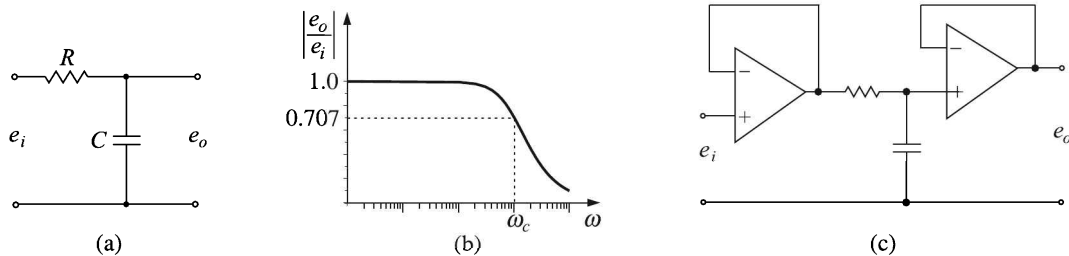


Fig. 16.37 Low-pass RC filter: (a) a simple circuit, (b) its frequency response, and (c) with unity-gain buffers at its input and output.

see that at low frequencies the capacitor would offer a high reactance and would, therefore, behave as an open circuit, allowing the entire input voltage e_i to appear at the output so that the amplitude ratio is 1. Above the cut-off frequency, where the gain falls below +3 dB (0.707), the capacitor conducts sufficiently to cut-off higher frequencies [Fig. 16.37(b)].

The sinusoidal transfer function of such a network can be written as

$$\frac{e_o}{e_i}(j\omega) = \frac{1/j\omega C}{R + 1/j\omega C} = \frac{1}{1 + j\omega\tau} \quad (16.34)$$

Therefore,

$$\text{Gain, } A = \left| \frac{e_o}{e_i}(j\omega) \right| = \frac{1}{\sqrt{1 + (\omega\tau)^2}} \quad (16.35)$$

Since the cut-off frequency corresponds to +3 dB attenuation, which means $A = 1/\sqrt{2}$, from Eq. (16.35) we get

$$\text{Cut-off frequency, } \omega_c = \frac{1}{\tau} = \frac{1}{RC} \quad (16.36)$$

The simple RC arrangement is sometimes inconvenient because the source and load impedances tend to modify the cut-off frequency and the passband gain. This difficulty can be overcome by introducing unity-gain buffers at the input and output of this filter [Fig. 16.37(c)] which enable the filter operate without being influenced by the input and load impedances.

An application of low-pass filter. Consider the dynamic measurement of strain by a strain-gauge through a bridge arrangement, as shown in Fig. 16.38(a). The strain varies the frequency at, say, 2 Hz.

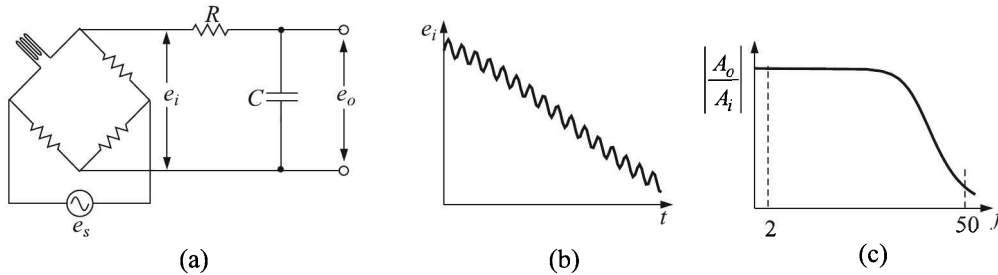


Fig. 16.38 Application of low-pass filter (LPF) in dynamic measurement of strain: (a) circuit, (b) bridge output without LPF contains 50 Hz noise, and (c) characteristic of LPF to cut off 50 Hz noise.

It is very likely that the bridge output will pick up 50 Hz noise from the power-supply line making the output appear as shown in Fig. 16.38 (b). This noise can be filtered by introducing a low-pass RC filter having a characteristic as depicted in Fig. 16.38 (c).

A low-pass filter can be designed by implementing the following steps:

- Step 1. Choose a value of high cut-off frequency ω_c .
- Step 2. Select a value of C less than or equal to 1 μF . Mylar or tantalum capacitors are preferred because of their better performance.
- Step 3. Calculate the value of R from the value of the time constant τ using Eq. (16.36).

Example 16.4

Design a simple RC low-pass filter for its application in the dynamic measurement of strain such that it has a 3 dB attenuation at 50 Hz.

Solution

From the given condition, $20 \log \frac{e_o}{e_i} = -3$, or $\frac{e_o}{e_i} = 0.708$. From Eq. (16.34), we have

$$\frac{e_o}{e_i} = \frac{1}{\sqrt{1 + \omega^2\tau^2}} = 0.708$$

Substituting $\omega = 2\pi \times 50$ in the above equation, we get

$$\tau = 3.2 \times 10^{-3} \text{ s} = RC$$

whence

$$R = \frac{3.2 \times 10^{-3}}{0.33 \times 10^{-6}} \cong 10 \text{ k}\Omega$$

if we take a capacitor of value $0.33 \mu\text{F}$.

High-pass RC filter

A high-pass RC filter can be formed by interchanging the resistor and capacitor in a low-pass filter [Fig. 16.39 (a)]. The sinusoidal transfer function of such a filter is therefore given by

$$\frac{e_o}{e_i}(j\omega) = \frac{R}{R + (1/j\omega C)} = \frac{j\omega RC}{1 + j\omega RC}$$

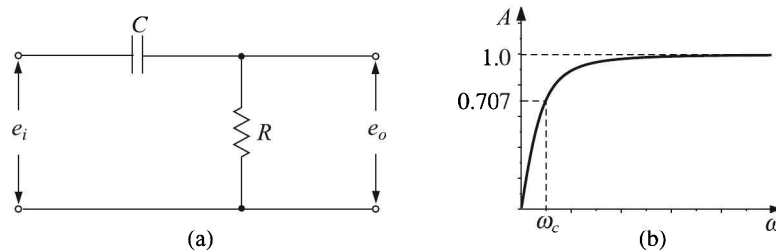


Fig. 16.39 High-pass RC filter: (a) a simple circuit, and (b) its frequency response curve.

Thus,

$$\text{Gain } A = \left| \frac{e_o}{e_i} \right| = \frac{\omega\tau}{\sqrt{1 + \omega^2\tau^2}} = \frac{1}{\sqrt{1 + (1/\omega^2\tau^2)}}. \quad (16.37)$$

Equation (16.37) shows that for low ω , gain A is small, while $A \rightarrow 1$ for high ω . In other words, this circuit allows high frequencies to pass unimpeded but offers a high impedance to low frequencies. The gain vs. frequency plot shown in [Fig. 16.39 (b)] gives a graphic description of the state of affairs.

An application of high-pass filter. Consider the measurement of the intensity of infrared radiation by a suitable detector. Since infrared radiation detection basically means sensing heat, the fluctuation of ambient temperature will interfere with the measurement. Such interference can easily be avoided by chopping the radiation incident on the detector, thus converting the output to ac [Figs. 16.40(a) and (b)].

This output can now be high-pass filtered to reject the low frequency variation of the base-line which owes its origin to the slow variation of the ambient temperature [Fig. 16.40 (c)]. In fact, an ac amplifier itself acts as a high-pass filter because as the name suggests it does not provide any dc gain and thus the effects of the dc offset voltage are eliminated.

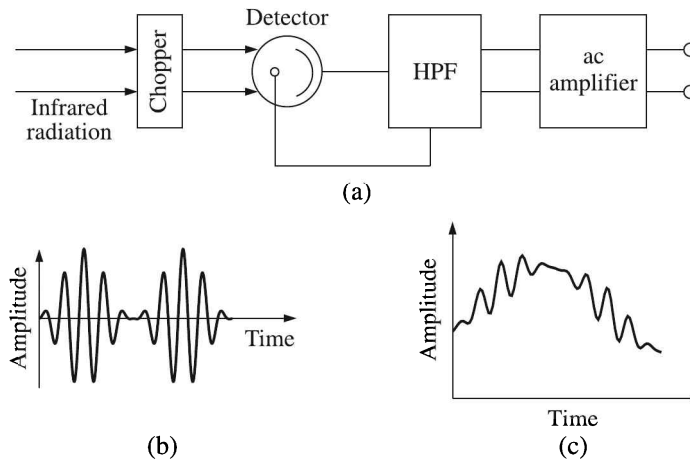


Fig. 16.40 Application of high-pass filter (HPF) in infrared detection: (a) block diagram, (b) chopper output, and (c) the HPF output.

High-pass RC filter as a differentiator. The sinusoidal transfer function for a high-pass filter is given by

$$\frac{e_o}{e_i}(j\omega) = \frac{j\omega\tau}{j\omega\tau + 1} \quad (16.38)$$

Hence, for $\omega\tau \ll 1$, we have $\frac{e_o}{e_i}(j\omega) = j\omega\tau$. The corresponding transfer function in the s -space is given by

$$\frac{E_o(s)}{E_i(s)} = s\tau$$

which on inverse Laplace transform yields

$$e_o = \tau \frac{de_i}{dt} \quad (16.39)$$

Equation (16.39) shows that for $\omega\tau \ll 1$, the high-pass filter behaves as a differentiator.

Band-pass filter

If a low-pass and a high-pass filter are connected in series (i.e. cascaded) [Fig. 16.41(a)], the resulting transfer function of the circuit is

$$\frac{E_o(s)}{E_i(s)} = \frac{1}{1 + \tau_1 s} \times \frac{\tau_2 s}{1 + \tau_2 s}$$

where $\tau_1 = R_1 C_1$ and $\tau_2 = R_2 C_2$ with $\tau_2 > \tau_1$. The characteristics of a band-pass filter are shown in Fig. 16.41(b).

In the passband, the behaviour of the circuit closely resembles that of purely resistive network. Hence, the gain in the passband is given by

$$A = \left| \frac{e_o}{e_i} \right| = \frac{R_2}{R_1 + R_2}$$

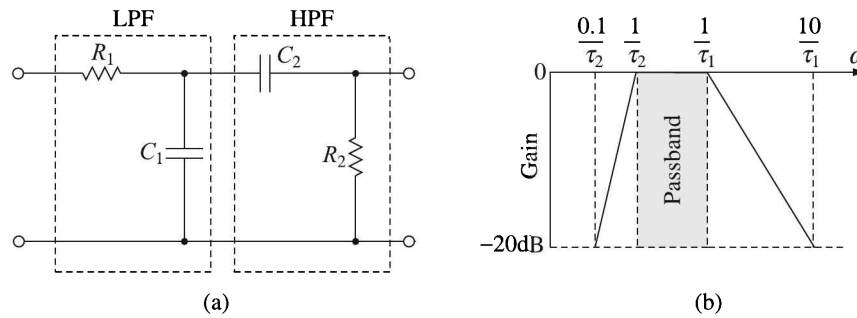


Fig. 16.41 Band-pass filter: (a) circuit, and (b) the characteristic.

Band-reject filter

Figure 16.42 (b) shows a simple band-reject filter which is basically a parallel combination of high-pass and low-pass networks as shown in Fig. 16.42 (a).

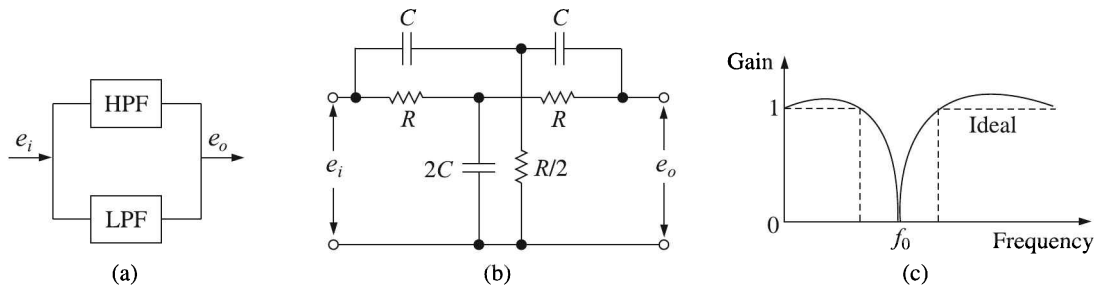


Fig. 16.42 Band-reject filter: (a) block diagram, (b) circuit, and (c) the characteristic.

Because of the peculiarity of its construction, it is called a *twin-T filter*. At low frequencies the resistive circuit (low-pass network) conducts while at high frequencies the capacitive circuit (high-pass network) conducts, the gain in both the cases being almost 1. At some intermediate frequency-band the circuit completely absorbs the entire input energy transmitting nothing to the output, thus making the gain nearly zero. The ideal and actual characteristics of such filters are shown in Fig. 16.42(b).

Narrow-band band-reject circuits are also called *notch filters*, the frequency f_0 corresponding to zero gain being called the *notch frequency*.

Some of the nonlinear processes, e.g. clipping and clamping, squaring, linearisation or multiplication by another function can very effectively be done with the help of op-amps. A detailed discussion of these circuits is beyond the scope of this book. Interested readers may consult any standard text on op-amps⁴ for these circuits.

Dynamic Compensation

Dynamic compensation becomes necessary when the dynamic characteristics of a system cannot be altered to a desired extent by altering its own parameters. For example, we have

⁴See, for example, *Op-Amps and Linear Integrated Circuits*, 4th ed., RA Gayakwad, Prentice-Hall of India (2006).

seen in Section 4.5 that the thermocouple is a first-order system and its time constant is given by

$$\tau = \frac{mc}{K_t A}$$

If it is necessary to make its response faster by lowering the value of τ , the value of m needs to be lowered. That means, we need to use a finer wire for the thermocouple. But then, finer wires may not withstand mechanical shocks of a dynamic environment. In such a situation, τ can be adjusted to our will by introducing a dynamic compensation circuit, of course, at the expense of its static sensitivity.

The general principle of dynamic compensation is as follows. Suppose, the transfer function of a system is $G(s)$ with a specified value p of a certain parameter. We can change it to p' by coupling it with a circuitry having the transfer function $G'(s)/G(s)$, where $G'(s)$ incorporates our requisite p' . Figure 16.43 explains it graphically.

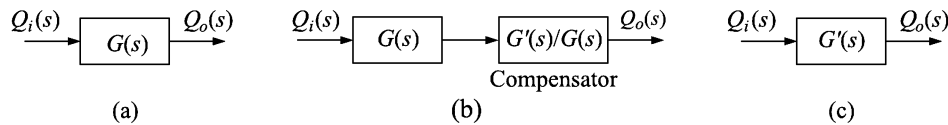


Fig. 16.43 Dynamic compensation: (a) original situation, (b) inclusion of compensator circuit, and (c) the situation after compensation.

Let us consider dynamic compensation of first-order and second-order systems to illustrate the principle. For simplicity, we will consider passive elements such as L , C , R for compensation, although active compensation with the help of op-amps is more versatile and powerful.

First-order system

The transfer function of a first-order system is given as

$$G(s) = \frac{K}{\tau s + 1}$$

We want to change τ to τ' such that $\tau' < \tau$. The compensation network is shown in Fig. 16.44 (a), where R_0 , C_0 combination is so chosen that $R_0 C_0 = \tau =$ time constant of the preceding first-order system.

In the frequency-domain, the input-output relation for the passive network can be written as

$$e_o(j\omega) = \frac{R}{R_0 \parallel C_0 + R} e'_i(j\omega) = \frac{R}{R_0} \frac{e'_i(j\omega)}{\frac{j\omega C_0}{R_0 + (1/j\omega C_0)} + R} = \frac{R(j\omega C_0 R_0 + 1)}{R_0 + R + j\omega C_0 R_0 R} e'_i(j\omega) \quad (16.40)$$

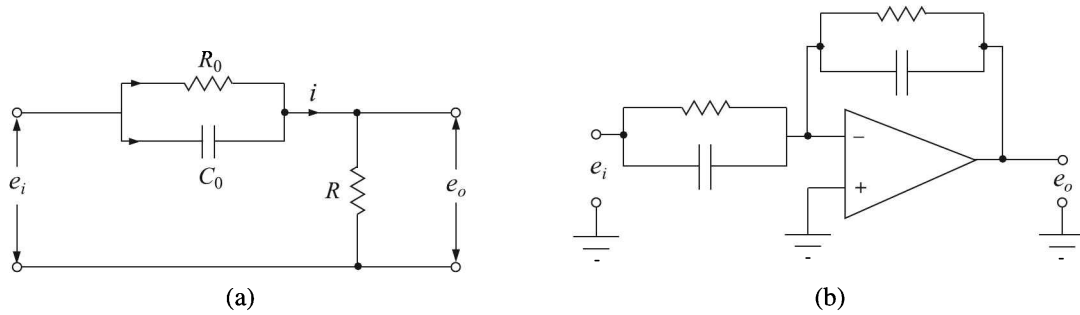


Fig. 16.44 Compensation for first-order system: (a) passive network, and (b) active compensation.

On writing Eq. (16.40) in the time-domain, we get

$$\begin{aligned}
 E_o(s) &= \frac{R(C_0 R_0 s + 1)}{R_0 + R + C_0 R_0 R s} E'_i(s) \\
 &= \frac{[R/(R_0 + R)](R_0 C_0 s + 1)}{[R/(R_0 + R)]R_0 C_0 s + 1} E'_i(s) \\
 &\equiv \frac{\alpha(\tau s + 1)}{\alpha\tau s + 1} E'_i(s) \qquad (16.41)
 \end{aligned}$$

where $\alpha = R/(R_0 + R)$ and $\tau = R_0 C_0$.

Writing $\alpha\tau = \tau'$, we get from Eq. (16.41) the transfer function of the compensating circuit as

$$\begin{aligned}
 \frac{E_o(s)}{E'_i(s)} &= \frac{\alpha(\tau s + 1)}{\tau' s + 1} \\
 &\equiv \frac{\tau s + 1}{K} \cdot \frac{K'}{\tau' s + 1} \quad \text{where } \alpha \equiv \frac{K'}{K} \\
 &\equiv \frac{G'(s)}{G(s)}
 \end{aligned}$$

as desired.

Note: Since $\alpha = R/(R_0 + R)$, it is always less than 1. Hence, $\tau' < \tau$. But, the effective static sensitivity $K' = \alpha K$. Therefore, $K' < K$. Thus a dynamic compensation always lowers the static sensitivity. Sometimes it may lower it by as much as 80% as Example 16.5 shows.

But there is nothing to worry about because the decrease in sensitivity can be compensated for by including a subsequent amplification stage. The advantage of active compensators [Fig. 16.44(b)] is that they perform both the functions at the same time.

Example 16.5

A first-order thermocouple has a time constant $\tau = 0.01$ s. Design a compensating network for the thermocouple to reduce its time constant by a factor of 5. What is the resultant loss in sensitivity of the thermocouple?

Solution

The compensating network will look like Fig. 16.44 (a). We are given

$$\tau = R_0 C_0 = 0.01 \text{ s} \quad (\text{i})$$

$$\tau' = \frac{R}{R_0 + R} \tau = \frac{\tau}{5}$$

or

$$\frac{R}{R_0 + R} = \frac{1}{5} \quad (\text{ii})$$

Thus, we are left with two equations to evaluate three parameters, namely R_0 , C_0 and R . Let us arbitrarily choose $R = 1 \text{ M}\Omega$. Then,

$$\frac{R_0 + R}{R} = 5 \quad [\text{from Eq. (ii)}]$$

or

$$R_0 = 4 \text{ M}\Omega$$

$$C_0 = \frac{0.01}{4 \times 10^6} \text{ F} = 2500 \text{ pF} \quad [\text{from Eq. (i)}]$$

If the original sensitivity was $K \text{ mV}/^\circ\text{C}$, the present sensitivity is $K' = \alpha K$. Therefore,

$$\text{Loss in sensitivity} = \frac{K - \alpha K}{K} \times 100\% = \left(1 - \frac{1}{5}\right) \times 100\% = 80\%$$

Second-order system

In second-order systems, we need to adjust the damping ratio ζ . In accordance with the principle discussed before, it may be compensated for as follows.

The transfer function of a second-order system is

$$G(s) = \frac{K}{\frac{s^2}{\omega_n^2} + 2\frac{\zeta s}{\omega_n} + 1}$$

Suppose, we want to change ζ to ζ_1 with the help of a compensation network. Such a passive network is shown in Fig. 16.45.

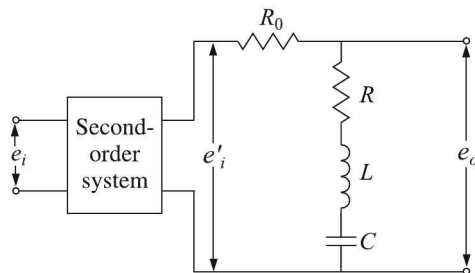


Fig. 16.45 Passive compensation network for the second-order system.

The frequency-domain transfer function for this network can be written as

$$\begin{aligned}\frac{e_o}{e'_i}(j\omega) &= \frac{R + j\omega L + \frac{1}{j\omega C}}{(R_o + R) + j\omega L + \frac{1}{j\omega C}} \\ &= \frac{(j\omega)^2 \frac{1}{\omega_n'^2} + 2\zeta' \frac{j\omega}{\omega_n'} + 1}{(j\omega)^2 \frac{1}{\omega_n'^2} + 2\zeta_1 \frac{j\omega}{\omega_n'} + 1}\end{aligned}\quad (16.42)$$

where

$$\begin{aligned}\omega_n' &= \frac{1}{\sqrt{LC}} \\ \zeta' &= \frac{R}{2\sqrt{L/C}} \\ \zeta_1 &= \frac{R + R_o}{2\sqrt{L/C}}\end{aligned}$$

In the time-domain, Eq. (16.42) becomes

$$\frac{E_o}{E'_i}(s) = \frac{\frac{s^2}{\omega_n'^2} + 2\zeta' \frac{s}{\omega_n'} + 1}{\frac{s^2}{\omega_n'^2} + 2\zeta_1 \frac{s}{\omega_n'} + 1}$$

If we now choose L and C in such a way that $\omega_n' = \omega_n$ and $\zeta' = \zeta$ of the preceding second-order stage, then the transfer function of the combined second order system and the compensation network becomes

$$\begin{aligned}G'(s) \equiv \frac{E_o}{E_i} &= \frac{K}{\frac{s^2}{\omega_n^2} + 2\zeta \frac{s}{\omega_n} + 1} \cdot \frac{\frac{s^2}{\omega_n^2} + 2\zeta \frac{s}{\omega_n} + 1}{\frac{s^2}{\omega_n^2} + 2\zeta_1 \frac{s}{\omega_n} + 1} \\ &= \frac{K}{\frac{s^2}{\omega_n^2} + 2\zeta_1 \frac{s}{\omega_n} + 1}\end{aligned}$$

which is exactly what we want.

Example 16.6

The natural frequency, the damping ratio and the static sensitivity of a second-order transducer are 500 rad/s, 0.3 and 5 mV/ μm respectively. It is necessary to change the damping ratio to 0.65 for a good frequency response. Design a suitable compensating network.

Solution

The network will look like Fig. 16.45. We are given

$$\frac{1}{\sqrt{LC}} = 500 \text{ rad/s} \quad (\text{i})$$

$$\frac{R}{2\sqrt{L/C}} = 0.3 \quad (\text{ii})$$

$$\frac{R + R_0}{2\sqrt{L/C}} = 0.65 \quad (\text{iii})$$

We have to determine four parameters, namely, R , R_0 , L and C whereas we have only three equations. Let us arbitrarily choose $L = 10 \text{ H}$. Then, from Eqs. (i), (ii) and (iii) we get

$$C = \left(\frac{1}{500\sqrt{L}} \right)^2 = \frac{1}{25 \times 10^4 \times 10} \text{ F} = 0.4 \mu\text{F}$$

$$R = 0.3 \times 2\sqrt{L/C} = 0.3 \times 2 \times \frac{10}{2 \times 10^{-3}} \Omega = 3 \text{ k}\Omega$$

$$R_0 = 0.65 \times 2\sqrt{L/C} - R = 0.65 \times 2 \times \frac{10}{2 \times 10^{-3}} - 3000 \Omega = 3.5 \text{ k}\Omega$$

Signal Conditioning System

With this background of bridges, amplifiers, detectors, modulators and filters we can now set up generalised signal conditioning systems.

DC systems

The generalised dc signal conditioning system is shown in Fig. 16.46.

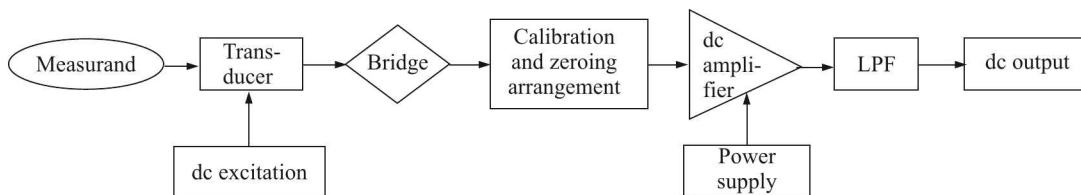


Fig. 16.46 The dc signal conditioning system.

In such a system the bridge is generally of the Wheatstone variety, the transducer constituting one arm or more. Calibration and zeroing arrangement may be a simple manually adjustable potentiometer or an elaborate automatic compensating device. The dc amplifier may need to possess a high CMRR, but it must have a good thermal stability. The LPF is necessary to lop off unwanted ac pick-ups.

The dc signal conditioner is cheaper, can be calibrated easily and its overload recovery is better, though its chief disadvantage is drift. The point has already been discussed.

AC system

A generalised ac signal conditioning system looks almost similar to dc's except that a carrier oscillator provides excitation to the ac bridge as well as reference signal to the phase sensitive detector (Fig. 16.47).

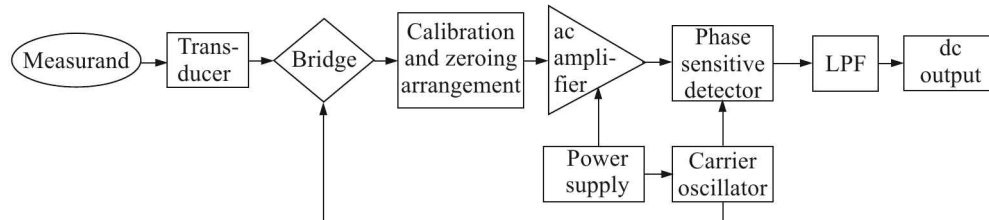


Fig. 16.47 The ac signal conditioning system.

Transducers, which constitute part of bridges, are generally excited by ac sources of frequency 50 Hz – 200 kHz. One has to choose this carrier frequency carefully so that it is at least 10 times the signal frequency. A phase-sensitive detection is necessary to ensure polarity of the dc output.

The carrier-type ac signal conditioner requires a stable carrier oscillator. This requirement is more stringent than obtaining stabilised dc source for the dc signal conditioner. However, save this difficulty ac signal conditioning system has many plus points, such as freedom from drift, very high signal-to-noise ratio by using active filters, and complete elimination of the carrier frequency through phase-sensitive detection.

Signal Transmission

The signal received by a transmitter is usually of non-standard type while that transmitted by it has to remain within a standard range. Signals that are used to transmit data can be of many kinds, the most common of them being pneumatic, electronic and optical. Of course radio and hydraulic signals are used too. But they are not so common presumably because there may be interference in radio signals and leakages in hydraulic transmitters.

However, at the present scenario, pneumatic and electronic signals are predominant. We defer a discussion on the transmission of pneumatic signals right now.

The 4–20 mA current loop

The analogue electronic signals that are transmitted are current signals that lie normally in the 4–20 mA range. Why current signals are preferred to voltage signals has been discussed in Section 2.2. under the heading *Interference* at page 12. The live zero of output is 4 mA rather than 0 mA which enables one to distinguish between a fault situation (output = 0 mA) from a zero output (output = 4 mA). Following the pneumatic standard of 3 to 15 psig, 3 to 15 mA was tried in the past. But somehow the standard gravitated to 4 to 20 mA from 1950s and is still in vogue for analogue signals.

Closed-loop systems with high-gain negative feedbacks enjoy good immunity from disturbances. For this reason, closed-loop transmission systems were initially popular despite a loss in gain in such systems. But later, with the advent of precision mechanical components

and highly reliable integrated electronics, these were replaced with open-loop systems where there is no loss of gain.

The requirement of conversion of non-standard process signals to standardised signals will be apparent from the following example. Suppose our measurement range of a temperature is 100 to 500°C which means that the span of the measurement is 400°C. So, at the lowest temperature the transmitter should output 4 mA and at the highest temperature, 20 mA. The steady-state gain K is

$$K = \frac{\text{Change in output}}{\text{Change in input}} = \frac{16}{400} = 0.04 \text{ mA}/^\circ\text{C}$$

So, it is necessary to build a transmitter that senses temperature change and generates current change with a gain of 0.04 mA/°C.

Such transmitters are commercially available which convert outputs of RTD, thermocouple, and bridge-circuit offsets to proportional output current. Open-loop transmitters for millivolt and resistance input are shown in block diagram forms in Fig. 16.48.

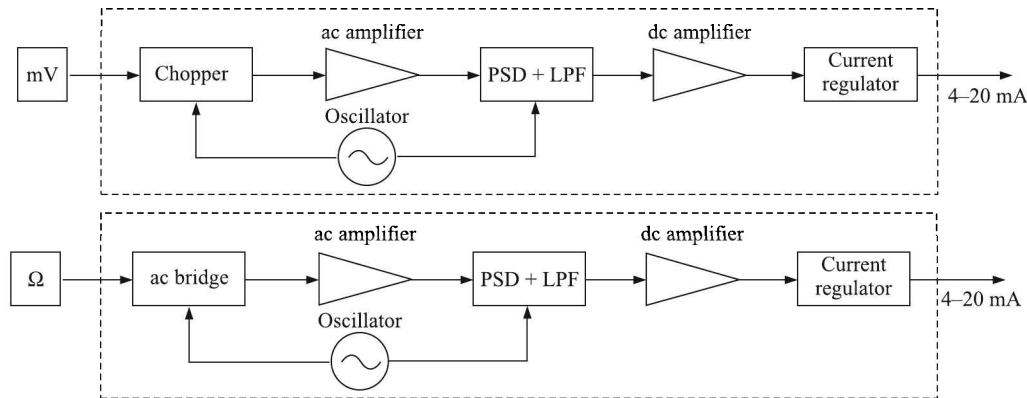


Fig. 16.48 Open-loop transmitters for thermocouple and RTD output.

Next comes the question of transmitting the signal to the control room which may be located at a considerable distance. Multiple loads can be series connected in a transmitter loop, providing considerable control and display opportunities. Loads today typically have full scale input requirements of 1 V, 5 V and 10 V to maintain the devices within their operational specifications. This voltage is often referred to as *compliance voltage*.

Transmitter standards. The American National Standards Institute (ANSI) and The Instrumentation Systems and Automation Society (ISA) have designated type numbers with suffix letters as transmitter identifiers⁵. The type number is the number of wires necessary to provide transmitter power. Shields and IO wires are excluded. Suffix letters identify the load resistance capability. The following are the type numbers:

1. ISA Type 2 or Two-wire transmitter
2. ISA Type 3 or Three-wire transmitter
3. ISA Type 4 or Four-wire transmitter

⁵ANSI/ISA-50.1-1982 (R1992) formerly ANSI/ISA-S50.1-1982 (R1992) *Compatibility of Analog Signals for Electronic Industrial Process Instruments*, see <http://www.ISA.org/> or <http://www.ANSI.org/>.

ISA Type 2. The Type 2 is a two-wire transmitter energised by the loop current where the loop source voltage (compliance) is included in the receiver. The transmitter floats and the signal ground is in the receiver (Figure 16.49). A two-wire transmitter is mostly dc powered.

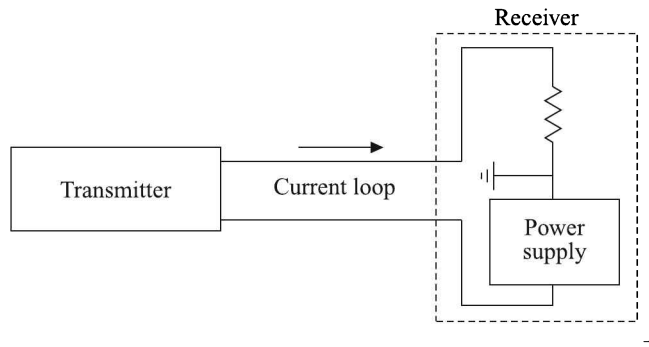


Fig. 16.49 Two-wire transmission of data.

ISA Type 4. The Type 4 is a four-wire transmitter energised by a supply voltage at the transmitter. The transmitter sources the loop current to a floating receiver load (Figure 16.50). Although that means an increase in installation cost, but there are situations which demand a four-wire transmission.

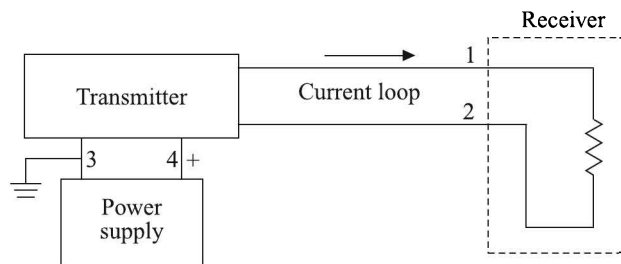


Fig. 16.50 Four-wire transmission of data. 1, 2, 3 and 4 indicate wire numbers.

For example, the electromagnetic flow meter often comprises a sensing element, an electromagnet and other electronic components. Because of distance restrictions, the electronics enclosure needs to be mounted near the sensor. This necessitates a four-wire installation where two wires provide the electrical power and the other two transmit the output signal to a receiver which may be a distributed control system (DCS), a data acquisition system (DAQ), a recorder or a display.

Frequently transmitters are placed at a considerable distance from the receivers in a plant and has field inputs. Field inputs are usually referenced to field grounds or in some cases actually connected to a field ground (for example, the grounded thermocouple). Receiver grounds are rarely identical to field grounds; therefore, isolation is required to eliminate potential ground loop problems⁶. ISA recommends that

In no event should transmitters with grounded outputs be connected to grounded receivers. They require an isolator in the loop or floating receiver system.

⁶See *Multiple earths* in Section 2.2 at page 14.

The power for a four-wire transmitter can be either ac or dc and the output signal can be either current or voltage. Four-wire transmitters are usually not used for conventional temperature, pressure or level measurements.

Opto-isolator. An opto-isolator contains a device that converts electrical input signal into light, a closed optical channel (also called dielectric barrier), and a photosensor, which detects the incoming light. It either generates electric energy directly, or modulates electric current flowing from an external power supply. Figure 16.51 shows a typical circuit incorporating an opto-isolator (aka *opto-coupler*).

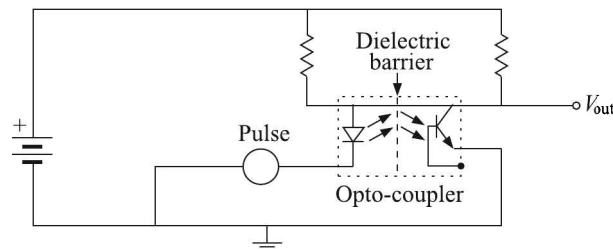


Fig. 16.51 Schematic diagram of a circuit showing an opto-coupler. The source of light (LED) is on the left, dielectric barrier is in the centre, and the sensor (phototransistor) is on the right.

The source of light is almost always a near infrared light-emitting diode (LED) and the sensor can be a photoresistor, a photodiode, a phototransistor, a silicon-controlled rectifier (SCR) or a triac. The cheapest kind has a phototransistor. Because LEDs can sense light in addition to emitting it, the construction of symmetrical, bidirectional opto-isolators is possible.

Opto-couplers allow us to send digital (and sometimes analogue) signals between circuits with separate grounds. This function can also be achieved by employing isolation transformers, which use inductive coupling between electrically isolated input and output sides. Unlike transformers, opto-isolators can pass dc or slow-moving signals and do not require matching impedances between input and output sides. Both transformers and opto-isolators are effective in breaking ground loops.

ISA Type 3. We took up the three-wire transmission after considering two-wire and four-wire transmissions because the three-wire transmitter is a blend of the two-wire and four-wire versions.

The Type 3 is a three-wire transmitter energised by a supply voltage at the transmitter. The transmitter sources the loop current. The transmitter common is connected to the receiver common. The three-wire transmission is shown in Figure 16.52.

Thus the three-wire transmitter utilises the best of both two-wire and four-wire transmitters. One less wire than a four-wire transmitter is required and the outputs are provided for both 4–20 mA current signals and 0–10 V voltage signals. Because they are mostly dc powered and they do not require an isolator, the cost of these transmitters is lower.

Sub-classification of transmitters. All 4–20 mA transmitters may not necessarily be identical in their ability to provide current into different loads. For example, a typical 4–20 mA transmitter module cannot drive a 100 k Ω load, because this requires a compliance

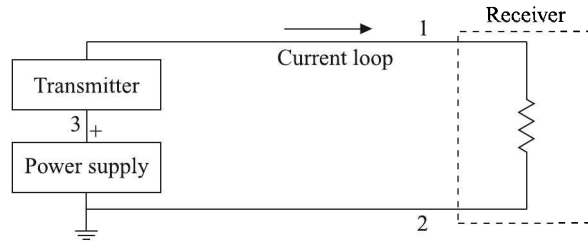


Fig. 16.52 Three-wire transmission of data. 1, 2 and 3 indicate wire numbers.

source of 2000 V ($20 \text{ mA} \times 100 \text{ k}\Omega$). To address this problem, besides the type classifications, ISA specified classes to identify the load of a transmitter vis-à-vis its power supply voltages. These are given in Table 16.4.

Table 16.4 Transmitter classes

| Class | Maximum load resistance (Ω) | Minimum supply voltage (V dc) |
|-------|--------------------------------------|-------------------------------|
| H | 300 | 21 |
| L | 800 | 32.7 |
| U | 300 to 800 | 23 to 32.7 |

The class standard ensures that modules of identical classes are interchangeable with respect to their drive capabilities. Combining the type and class categories, a Type 2L transmitter from one manufacturer can replace one from another manufacturer without replacing other devices in the circuit.

Choosing a power supply. If a 4–20 mA transmitter is not having a proper power supply, it will be unable to produce a 100% output reading. Understanding the electrical response of different transmitters is key in designing, installing and maintaining 4–20 mA loops with sufficient power to operate through the entire variable range it is purported to transmit.

It is rather easy to determine which power supply best meets our requirements. The thing that needs to be considered is that the power supply must supply voltage equal to or greater than the voltage drops to the combined system. The procedure will be clear from the following example.

Example 16.7

We need to set-up a 4–20 current loop for a pressure transducer with specified operating voltage of 12–30 V. The transducer is located at a distance of 2000 ft from the control room. The connecting 2-wire transmitter comprises a 24-gauge solid copper wire of resistance $2.62 \Omega/100 \text{ ft}$. The data acquisition system, shown in Figure 16.53, maps current to a range of 1 to 5 V. Determine the power supply necessary for this transmitter and the class thereof.

Solution

The typical shunt resistor to map a current range of 4–20 mA to 1–5 V is 249Ω , as shown below.

$$0.004 \text{ A} \times 249 \Omega = 0.996 \text{ V}$$

$$0.020 \text{ A} \times 249 \Omega = 4.98 \text{ V}$$

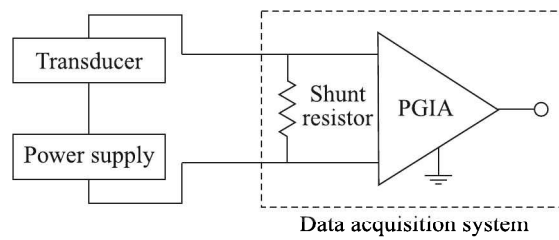


Fig. 16.53 Current loop and data acquisition system of Example 16.7. PGIA indicates the programmable gain instrumentation amplifier.

The pressure transducer requires a minimum operating voltage of 12 V. The cable resistance is

$$2000 \text{ ft} \times \frac{2.62 \Omega}{100 \text{ ft}} \times 2 = 104.8 \Omega$$

The corresponding voltage drop is

$$104.8 \Omega \times 0.020 \text{ A} = 2.096 \text{ V}$$

So, the total voltage requirement is

$$12 + 5 + 2.096 = 19.096 \text{ V}$$

The total load resistance exceeds 353.8 (249 + 104.8) Ω because we do not know the transducer resistance. Therefore a 2U transmitter with at least 24 V power supply will be suitable.

Smart Transmitters

Of late, *smart* or intelligent transmitters have come into play. These are open-loop transmitters with microcomputers incorporated into the system. This new technology has made a sea-change in the overall performance, accuracy and flexibility of instrumentation, signal conditioning, and transmission. A block diagram of such a transmitter is presented in Fig. 16.54.

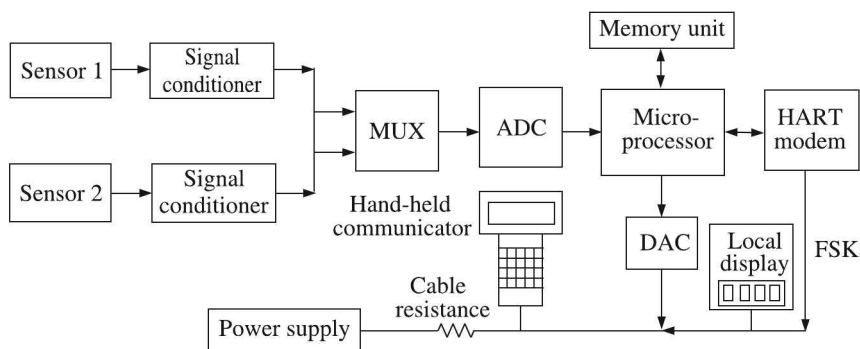


Fig. 16.54 Block diagram of smart transmitter using the HART protocol.

A smart transmitter can convert analogue signals to digital signals, making the communication fast and easy. It can even send both analogue and digital signals. One

implementation of the smart transmission is the Highway Addressable Remote Transducer⁷ (HART). It uses a bus topology that can accommodate 15 smart devices with one power source. The FSK digital signal (1200 Hz for '1' and 2200 Hz for '0', amplitude 0.5 mA) rides on the 4–20 mA analogue current loop (Figure 16.55). We experience a similar concurrent transmission of analogue telephone signals and digital broadband signals through the landline.

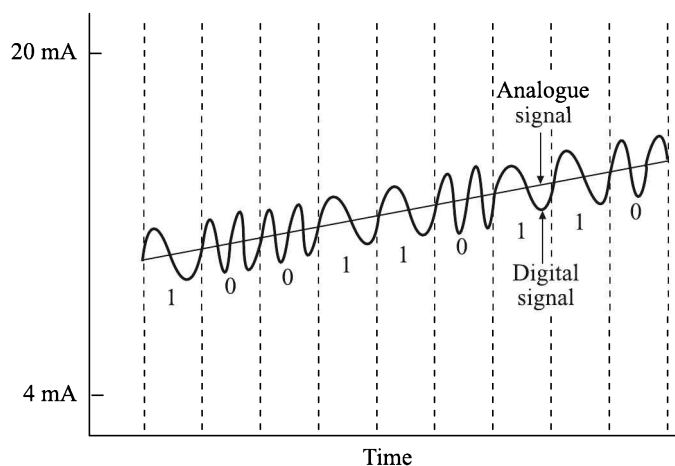


Fig. 16.55 Schematic diagram showing concurrent transmission of analogue and digital signals through the 4–20 mA current loop in HART.

With digital signalling, HART-enabled transducers allow responses to proprietary commands to perform certain actions or reply with their status which is not possible with the 4–20 mA analogue signalling. The communicator is a hand-held device that allows digital *commands* to be delivered to the HART-enabled transducers. The communicator has a display that lets the operator see the input/output information and it can be directly connected anywhere in the loop.

The protocol command set of the HART is organised into three groups as given in Table 16.5.

In fine, the HART transmitter incorporates the following features:

1. Configuration
2. Re-ranging
3. Characteristics
4. Signal conditioning
5. Self-diagnosis

Configuration. The HART can be configured to meet the requirements of the process in which it is used. For example, it can be set to read almost any range or type of thermocouple, RTD or thermistor. This feature reduces the need for a large number of specific replacement devices.

⁷Product of Rosemount Inc.

Table 16.5 HART protocol commands

| <i>Group</i> | <i>Functions</i> |
|--------------------------|---|
| Universal commands | Implemented by all HART-enabled devices and provide interoperability across products of different manufacturers. These commands include among others: <ol style="list-style-type: none"> 1. Manufacturer and device type 2. Primary variable and units 3. Current input and percentage of change 4. Four predefined dynamic variables 5. Eight character tag 6. Sixteen character descriptor and date |
| Common-practice commands | This set is used in many HART field devices but not all. It includes functions as: <ol style="list-style-type: none"> 1. Writable transfer ranges 2. Ability to set zero and span 3. Perform self-test, etc. |
| Device-specific commands | These commands are unique to a particular field device. The functions included in the command set are <ol style="list-style-type: none"> 1. Start, stop, clear totaliser 2. Select primary variable 3. PID controller set point and tuning parameter manipulation. |

Re-ranging. The range under which the transmitter functions can be altered from the comfort of the control room rather than visiting the location of the transmitter to effect the change. For example, using the communicator, the operator can change from a Pt-100 RTD to a Type K thermocouple simply by reprogramming the transmitter when the transmitter switches to measuring voltage from measuring resistance.

Alternatively, if measuring pressure, the operator can determine at any time which unit to use—mmWG or psi or Pa, etc.

Characteristics. A HART transmitter is able to act as a stand-alone transmitter when it can send output signal to a Distributed Control System (DCS) or a Programmable Logic Controller (PLC).

Signal conditioning. Functions like scanning the average, eliminating noise spikes or introducing a delay to smooth the response for a rapidly fluctuating process, can be done by HART.

Self-diagnosis. By performing self-test, the HART can locate a malfunctioning circuit-board.

Fieldbus

A fieldbus is a completely digital transmission system which acts as a two-way communication link among intelligent field level as well as control devices. As such, it is a replacement of the analogue 4–20 mA current loop. A fieldbus named *Foundation* has already been introduced by Rosemount Inc.

16.3 Recovery of Signals

More often than not, signals from transducers are buried in noises. As the title suggests, in this section we will consider a couple of methods of recovery of signals rejecting noises.

Of these methods, the lock-in amplifier and phase-locked loops are important as well as ingenious. We now consider them.

Lock-in Amplifier

Filters that we have so far discussed are frequency-selective and therefore, they can be utilised to reject noise or unwanted signal if the desired and spurious signals are located in different positions in the frequency spectrum. Obviously, such filters are of little use if the signal and the noise have nearly equal frequencies. Even in such a situation the signal can be recovered if the following criteria are satisfied:

1. The noise is random
2. The desired signal is, or can be made, periodic

The device which can do the trick in such circumstances is called a *lock-in amplifier* which is basically a deft combination of a phase-sensitive detector (PSD) and a low-pass filter (LPF).

Schematically, it looks like Fig. 16.56. The periodic signal (mixed with noise) is fed to the two throws of an electronic SPDT in two ways—one directly and another through an inverting voltage-follower. The SPDT, in turn, is actuated by a square-wave reference type signal.

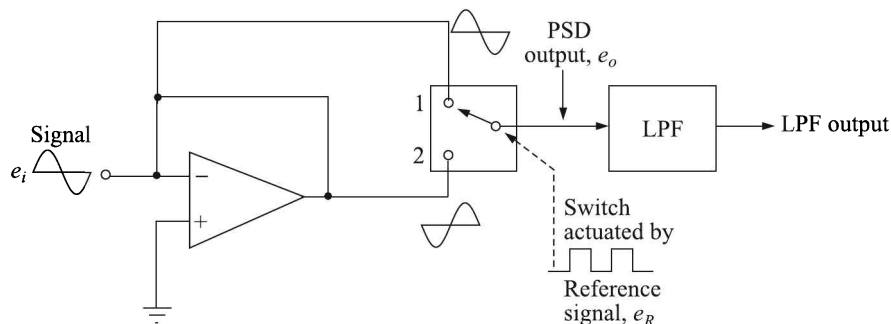


Fig. 16.56 Lock-in amplifier—circuit scheme.

If the reference signal has the same frequency and the phase as those of the desired signal, then the switch will be in position 1 to sample the direct signal for the first-half (i.e. for phase-angle 0° to 180°) of the wave, and at position 2 to sample the inverted signal for the next half. As a result, the PSD output will resemble that of a full-wave rectifier, and the LPF, acting as an averaging circuit, will give a steady dc output (Fig. 16.57).

If, however, the input signal is of different frequency, or of the same frequency but different phase from the reference signal, the diagrams in Fig. 16.58 will show that the LPF output will be zero or negligibly small if averaging is done over a sufficient time.

A simple mathematical analysis will perhaps clarify the operation. The PSD, in fact, acts as a multiplier, i.e. if e_i is the input signal, e_R the reference signal and e_o the PSD output, then

$$e_o = e_i \times e_R \quad (16.43)$$

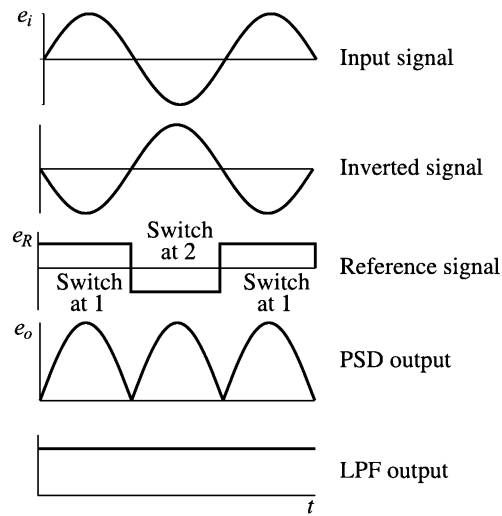


Fig. 16.57 Lock-in amplifier—signals at different stages.

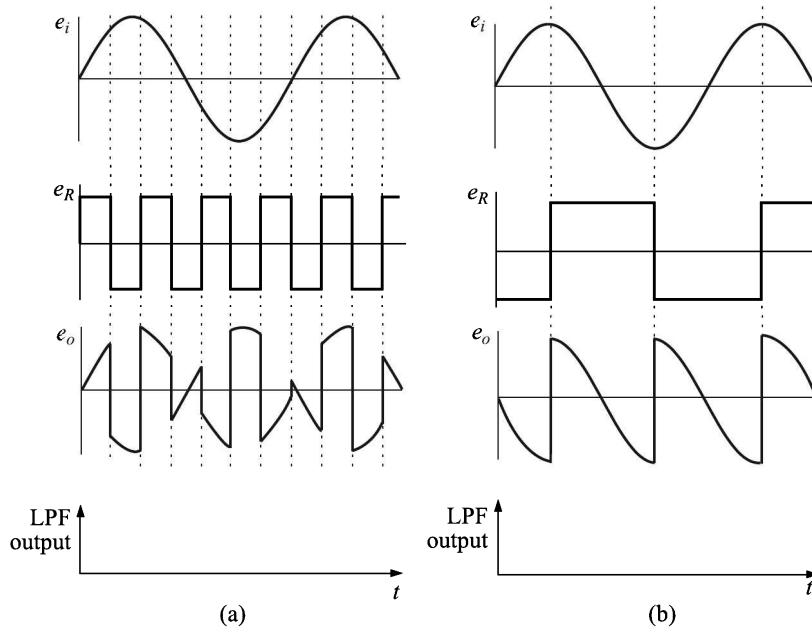


Fig. 16.58 Lock-in amplifier operation: (a) For an input signal of different frequency, and (b) for an input signal of different phase.

Now, e_i is sinusoidal whereas e_R is a square wave. But, with the help of Fourier analysis e_R can be written as sum of a fundamental sinusoidal component and its over-harmonics. To simplify matters, let us deal with the fundamental component and write

$$e_R = E_R \sin \omega_R t$$

where ω_R is the fundamental frequency of the reference signal.

And, since $e_i = E_i \sin \omega_i t$, from Eq. (16.43), we get

$$\begin{aligned} e_o &= E_R E_i \sin \omega_R t \sin \omega_i t \\ &= E_i \sin \omega_R t \sin \omega_i t \quad [\text{if } E_R = 1, \text{ which is the general practice}] \\ &= \frac{1}{2} E_i [\cos(\omega_R - \omega_i)t - \cos(\omega_R + \omega_i)t] \end{aligned} \quad (16.44)$$

Now, if the cut-off frequency of the LPF is ω_c and $\omega_c \ll \omega_R$, the frequency $(\omega_R + \omega_i)$ will be suppressed by the LPF while the frequency-band characterised by $|\omega_R - \omega_i| < \omega_c$ will pass through the LPF unattenuated. And, for $\omega_R = \omega_i$, the output will be a dc voltage which is proportional to the amplitude of the input signal.

So far we have not considered the phase factor. Let us now consider that for a signal of frequency ω_R . That is, let

$$e_i = E_i \sin(\omega_R t + \phi)$$

Then,

$$e_o = E_i \sin \omega_R t \sin(\omega_R t + \phi) = \frac{1}{2} E_i [\cos \phi - \cos(2\omega_R t + \phi)]$$

which when averaged over t yields

$$\bar{e}_o = \frac{1}{2} E_i \cos \phi \quad (16.45)$$

From this discussion it is clear that the signal-to-noise ratio depends on

1. Distribution of noise
2. Bandwidth of the LPF

The block diagram of a typical lock-in amplifier is shown in Fig. 16.59.

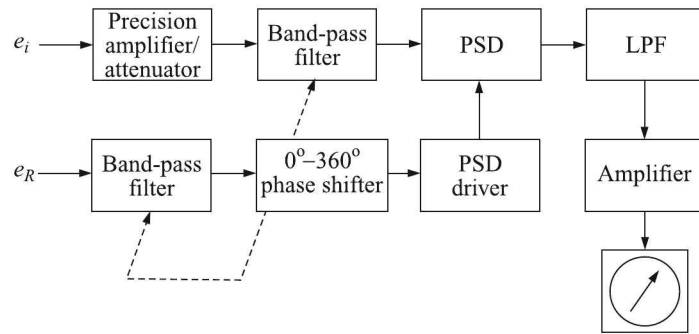


Fig. 16.59 Block diagram of a typical lock-in amplifier.

A lock-in amplifier automatically tracks a change in signal frequency since the reference frequency changes simultaneously. It is this automatic tracking feature which narrows the band-width of the lock-in amplifiers. Typical commercial lock-in amplifiers may detect a $1 \mu\text{V}$ signal rejecting a 1 V noise and may operate in the frequency range from 1 Hz to 1 MHz .

For precision measurements, such as in bridge measurements, it is better to use ac excitation because dc is normally associated with noise and ac noise contribution is all but eliminated by a lock-in amplifier.

We have already pointed out⁸ that for a push-pull type transducer, such as LVDT, a PSD in combination with an LPF (i.e. a lock-in amplifier) is used to provide a dc output whose amplitude is proportional to displacement, and polarity is related to direction of displacement from the null position.

Phase-locked Loop (PLL)

An interesting application of the lock-in amplifier concept is the PLL. Apart from FM stereo decoders, motor speed controls, tracking filters, frequency synthesised signal generators and receivers, FM demodulators, etc., PLL has found wide application in the generation of local oscillator frequencies in household TV and FM tuners as *automatic frequency control* (AFC). Indeed, PLL has emerged one of the fundamental building blocks in electronics today and it is commercially available as a single package (e.g. SE/NE 560 series of Signetics).

Basically, a PLL is a lock-in amplifier in which the reference signal is provided by its own output, converted to frequency by a VCO or voltage-to-frequency converter⁹. The block diagram of Fig. 16.60 will help understand its operation.

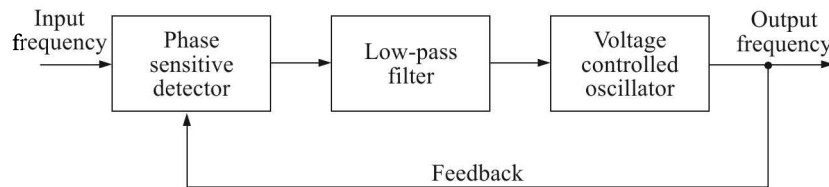


Fig. 16.60 Block diagram of PLL.

We have seen that the output of a PSD and LPF combination is a dc proportional to the phase difference between ω_i and ω_R when they are equal [see Eq. (16.45)]. When locked to the input frequency, the dc output is small but sufficient to drive the VCO to produce a frequency which is equal to that of the signal. In this tracking situation, the input signal and the VCO output are almost in phase quadrature (i.e. $\phi \simeq 90^\circ$, $\cos \phi \simeq 1$) and the PSD-LPF combination produces a small dc voltage which is often referred to as *error voltage*.

When there is no input, the PLL is in the *free-running* state. The moment an input signal is fed, the VCO frequency starts changing and the PLL is said to be in the *capture mode*. The VCO continues to change its frequency until it equals that of the input and stays put there; the PLL is then in the phase-locked state. In this state if there is any change in the input frequency, the loop automatically tracks it through its repetitive action.

16.4 Signal Conversion

With the advent of digital electronics, it is advantageous to convert signals to their digital form in order that they can be handled and analysed with the help of computers. Let us consider a few methods of conversion of analogue signals to their digital forms and vice versa.

⁸See Section 16.2 at page 773.

⁹See Section 16.4 at page 806.

Sample-and-hold Unit

After the analogue signal is processed, it may go to a sample-and-hold (S/H) circuit, followed by an analogue multiplexer, and then to an analogue-to-digital converter (ADC) [Fig. 16.61(a)].

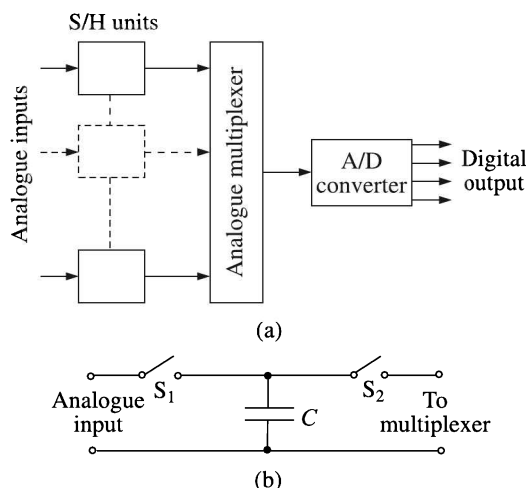


Fig. 16.61 Sample-and-hold process.

As the name indicates, these units sample analogue data inputs at a specified time and then hold the voltage levels at their outputs for the analogue multiplexer to perform a time-division-multiplexing (TDM) operation between different data inputs. Schematically, the process is as follows [Fig. 16.61(b)]:

- Step 1. Open S_2 and close S_1 . The capacitor gets charged to a voltage equal to the input from the signal conditioner (*Sample Mode*).
- Step 2. Open S_1 . The capacitor holds the voltage (*Hold Mode*).
- Step 3. The multiplexer sends signal to close S_2 when the sampled voltage is sent to the ADC [Fig. 16.61(a)].

Aperture time and *acquisition time* are the two parameters which are of primary importance in a S/H circuit. The former is the delay between the transition from sample mode to hold mode while the latter is the delay for the capacitor to acquire the new value of the input when the command changes from *hold* to *sample*. A S/H circuit using two op-amps and two FET switches is shown in Fig. 16.62(a).

The acquisition time depends on how fast the capacitor C can be charged which, in turn, depends on the current that the amplifier A_1 can provide.

An acquisition time of less than $10 \mu\text{s}$ and an aperture time of less than $1 \mu\text{s}$ are easily achieved with a $0.01 \mu\text{F}$ capacitor. If the input signal varies rapidly, there is some error in the value held owing to a finite aperture time [Fig. 16.62(b)].

Analogue-to-digital Conversion

Most transducers generate analogue signals, i.e. outputs which are continuous functions of time. In many cases, signals from transducers, if necessary after conditioning, are indicated

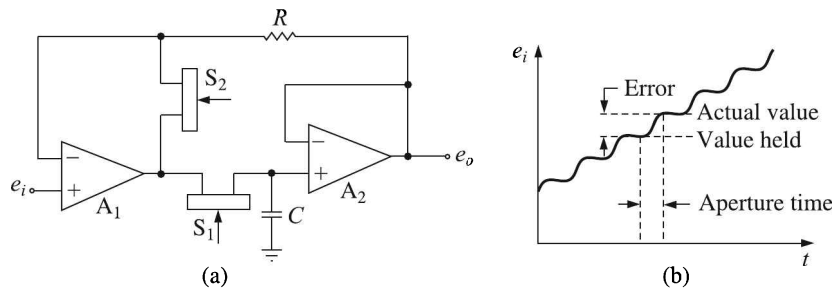


Fig. 16.62 Sample-and-hold unit: (a) schematic circuit, and (b) the error owing to finite aperture time.

or recorded using analogue instrumentation. Although theoretically an analogue output has an infinite resolution, the practical difficulties of readout devices limit the accuracy to usually not better than 0.1%.

Digital quantities on the other hand, are discrete and vary in equal steps. They can be read exactly through various read-out devices.

Before we consider methods of conversion of analogue signals to digital ones, we need to know a few parameters which we now discuss.

Quantisation and decision level

Each digital number is a fixed multiple of a minimum number which can be termed a *quantum*. Therefore, an analogue quantity has to be quantised to convert it to a digital one.

An analogue to digital conversion characteristic, or what is known as quantiser characteristic, is shown in Fig. 16.63(a).

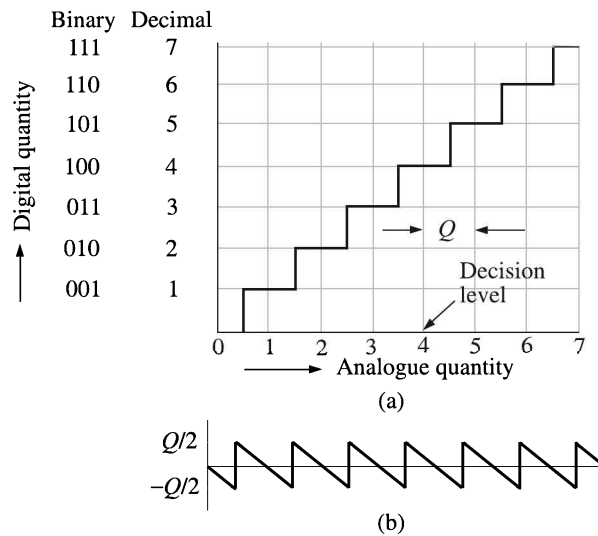


Fig. 16.63 Analogue to digital conversion: (a) quantiser characteristic, and (b) quantisation error.

While the analogue quantity (shown along the abscissa) is best expressed in the usual decimal representation, the digital quantity (shown along the ordinate) is best expressed in a

binary code. A quantum jump of 0.5 has been adopted in the conversion. Which implies that the digital value 1 represents the range 0.5 to 1.5 (i.e. 1 ± 0.5) of the analogue value. These quantum jump levels, e.g. 0.5, 1.5, 2.5 . . . , etc. are called *decision levels*. Decision levels are set at values which lie about the true levels and the distance Q between the decision levels is the quantum size. Thus analogue values between two decision levels cannot be obtained in the digital output.

Suppose, we have a 3-bit ADC which can generate $2^3 = 8$ numbers or 7 subdivisions ranging from 0 to 7 in decimal. If we want to convert an analogue quantity ranging from 0 to 7, the decision levels have to be set at 0.5, 1.5, 2.5, 3.5, 4.5, 5.5 and 6.5 so that we have 7 subdivisions which are offered by 3 bits. Thus the minimum step we can generate is 1 which is the quantum size. In other words, a 3-bit ADC has 8 or 2^3 discrete levels and 7 or $(2^3 - 1)$ analogue decision levels. Obviously, for n -bits, the number of discrete levels is 2^n and the number of analogue decision levels is $(2^n - 1)$.

Resolution

The resolution of an ADC is determined by the number of bits in the digital value and corresponds to the analogue value of the right-most bit or the least significant bit (LSB).

Thus, for a 3-bit ADC the binary code 101 represents

$$\left(1 \times \frac{1}{2}\right) + \left(0 \times \frac{1}{4}\right) + \left(1 \times \frac{1}{8}\right) = \frac{5}{8} \text{ of full-scale of the converter}$$

The full-scale analogue value for an ADC can be set at any convenient voltage. In this case, if the full-scale analogue value is 2 V, the minimum value of the ADC can generate is $2 \times (1/8) = 1/4$ V, which is the resolution. For an n -bit conversion the LSB weightage is $1/2^n$ of full-scale and therefore,

$$\text{Resolution} = \frac{\text{Full-scale analogue value}}{2^n}$$

Often, the resolution is expressed as percentage of the full-scale. For example, a 7-bit ADC which has a resolution of $1/(2^7) = 1/128$ of the full-scale is said to have a resolution of 1%. Table 16.6 gives an idea of the number of bits required to obtain different per cent resolutions.

Table 16.6 Bit requirement for different resolutions

| <i>Bits required</i> | <i>Resolution (Fraction of the full-scale)</i> | <i>Resolution (%)</i> |
|----------------------|--|-----------------------|
| 7 | $\frac{1}{128}$ | 1 |
| 10 | $\frac{1}{1024}$ | 0.1 |
| 14 | $\frac{1}{16384}$ | 0.01 |

Since a factor of 2 corresponds to 6.02 dB, the resolution of an ADC in dynamic range in dB is given by the expression (the number of bits) \times 6.02. Thus, a 7-bit ADC has a dynamic range of 42.14 in dB.

So far we have discussed only the binary code for the conversion, because it is the most commonly used code. Other convenient codes can also be used.

Quantisation error

From Fig. 16.63(a) one can see that our ADC will generate a digital value 1.0 for the analogue value 0.5. The error is 0.5. For analogue value 1.0 the error is 0, and for 1.5 it is -0.5 . Thus the error decreases linearly from 0.5 to -0.5 in this range. The picture repeats itself in the range 1.5 to 2.5 and so on, generating a saw-tooth curve as shown in Fig. 16.63(b). For a quantum size Q , the error amplitude varies between $+Q/2$ and $-Q/2$.

The rms value for a saw-tooth curve of period 2 and amplitude $E = E_m - E_m t$ is given by

$$\begin{aligned} E_{\text{rms}} &= \sqrt{\frac{1}{2} \int_0^2 (E_m - E_m t)^2 dt} \\ &= \sqrt{\frac{E_m^2}{2} \int_0^2 (1 - 2t + t^2) dt} \\ &= \frac{E_m}{\sqrt{3}} \end{aligned}$$

The output of a quantiser, therefore, is in effect the input analogue signal combined with a quantisation error E_Q given by

$$E_Q = \frac{Q}{2\sqrt{3}} \quad (16.46)$$

S/H unit as a simple ADC

We have already discussed the S/H unit which can be conveniently used for ADC by suitably adjusting the aperture time and minimising the acquisition time. Figure 16.64 illustrates the process.

In Fig. 16.64, (a) shows the analogue signal to be digitised, (b) shows a train of sampling pulses of very small ON- or aperture-time, (c) shows the result of that sampling process which can be thought of as a multiplication of the analogue value with the pulse value of unity magnitude at every sampling pulse. In other words, it means that an amplitude modulation of the sampling pulses takes place. These modulated values of pulse-amplitudes are held by the capacitor of the S/H circuit producing a digitised signal [Fig. 16.64 (d)].

It is necessary to control the aperture time t_a for a successful A/D conversion. Considering t_a as the uncertainty in amplitude for the time rate of change of signal, one can write

$$\Delta E_o = \left[\frac{d}{dt} (E_o \sin \omega t) \right]_{t=0} t_a = E_o \omega t_a$$

which yields

$$t_a = \frac{\Delta E_o}{E_o} \cdot \frac{1}{2\pi f} \quad (16.47)$$

Equation (16.47) can be used to determine the aperture time necessary to digitise a signal of frequency f . The following example will make the point clear.

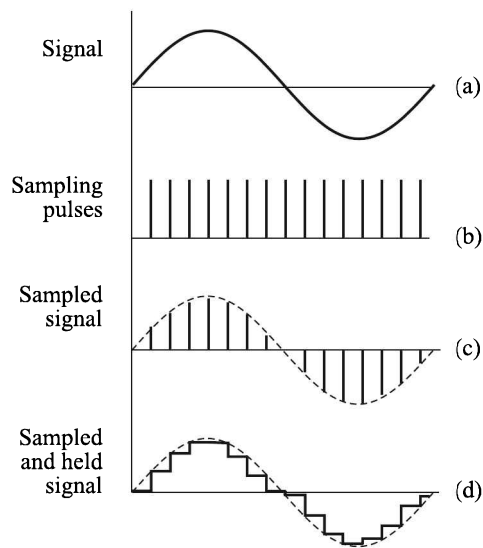


Fig. 16.64 Analogue to digital conversion process by S/H unit.

Example 16.8

Determine the aperture time required to digitise a 500 Hz signal to 10 bits resolution.

Solution

10 bits resolution is equivalent to 0.1% resolution, which means, $\frac{\Delta E_o}{E_o} = 0.001$. Therefore,

$$t_a = \frac{0.001}{2 \times 3.1416 \times 500} \text{ s} \simeq 318 \text{ ns}$$

Aliasing and sampling frequency

Consider two sinusoidal signals of frequencies 0.5 kHz and 3.5 kHz respectively. If they are sampled at a 4 kHz rate (i.e. $t_a = 0.25$ ms), the sampled values are those indicated by \odot in Fig. 16.65.

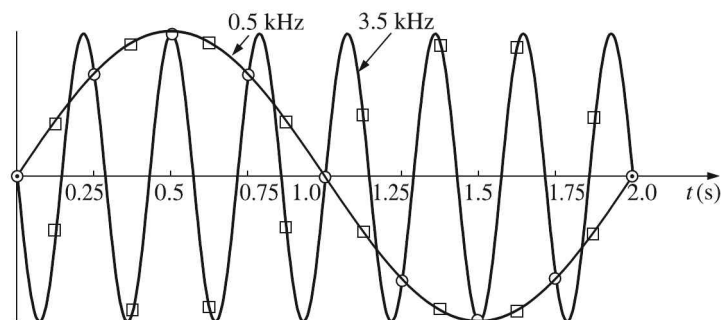


Fig. 16.65 Aliasing: \odot indicates sampled value at 4 kHz rate, \square indicates sampled value at 8 kHz rate.

From Fig. 16.65 it will be clear that both the signals will have the same sampled values, rendering unique reconstruction of original signals impossible. This ambiguity is called ‘aliasing’. If however, the sampling rate is 8 kHz (i.e. $t_a = 0.125$ ms), there will be no ambiguity as will be evident from Fig. 16.65 where the sampling is indicated by \square .

Though all signals are not sinusoidal, we can determine by a Fourier analysis the highest frequency f_h present in a signal. Once that is determined, Nyquist theorem states that the original signal can be completely recovered without distortion if it is sampled at the minimum rate of $2f_h$ samples per second. Sampling at lower frequencies generates aliasing.

Thus an S/H system can faithfully generate discrete values for analogue signal only up to a frequency of $f_s/2$ where f_s is the sampling frequency. In other words, effectively it acts like a low-pass filter with a cut-off frequency of $f_s/2$. It is also clear from Fig. 16.64 (d) that since a sampled value is held for a period of $1/f_s$ s, the generated signal has a phase delay of $1/(2f_s)$ s.

ADC techniques

Four methods are generally used in ADC. They are:

1. Successive approximation method (or potentiometric type)
2. Voltage-to-time conversion method (or ramp type)
3. Dual-slope integration method
4. Voltage-to-frequency conversion method (or integration type)

Most digital voltmeters are based on one of these or similar methods.

Successive approximation method. As the name implies, in this method a reference voltage is repeatedly divided by 2 and the result is compared with the analogue signal at each step. If the result is higher, the bit is set to 1 starting from the MSB (most significant bit, the left-most bit). Or else, the bit is set to 0. Let us consider an example from which the basics of the method will be clear.

Example 16.9

What will be the output of a successive approximation type 8-bit ADC if the input is 491 mV and the reference is 512 mV?

Solution

In the following table, d indicates binary digit. For example, d_7 indicates the digit at the seventh place from the right. We have an 8-bit ADC. Therefore, the MSB is d_7 .

| Step | Setting | Calculation | Comparison | Conclusion |
|------|-----------|------------------------------|-------------|-----------------------|
| 1. | $d_7 = 1$ | $512 \div 2 = 256$ | $491 > 256$ | d_7 remains a 1 |
| 2. | $d_6 = 1$ | $256 + (512 \div 2^2) = 384$ | $491 > 384$ | Retain a 1 at d_6 . |
| 3. | $d_5 = 1$ | $384 + (512 \div 2^3) = 448$ | $491 > 448$ | Retain a 1 at d_5 . |
| 4. | $d_4 = 1$ | $448 + (512 \div 2^4) = 480$ | $491 > 480$ | Retain a 1 at d_4 . |
| 5. | $d_3 = 1$ | $480 + (512 \div 2^5) = 496$ | $491 < 496$ | Set d_3 to a 0. |
| 6. | $d_2 = 1$ | $480 + (512 \div 2^6) = 488$ | $491 > 488$ | Retain a 1 at d_2 . |
| 7. | $d_1 = 1$ | $488 + (512 \div 2^7) = 492$ | $491 < 492$ | Set d_1 to a 0. |
| 8. | $d_0 = 1$ | $488 + (512 \div 2^8) = 490$ | $491 > 490$ | Retain a 1 at d_0 . |

Thus, the output of the ADC is 11110101.

Note: This output denotes the fraction of the reference voltage. The digital output equals

$$\left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{64} + \frac{1}{256}\right) \times 512 \text{ mV} = 490 \text{ mV}$$

which is in error by 1 mV. This is expected because the LSB has a value of 2 mV ($= 512 \div 2^8$) in this case. The error in the conversion equals $\pm \frac{1}{2}$ LSB, which is the inherent quantisation uncertainty.

A block diagram of the ADC is shown in Fig. 16.66. It incorporates a digital-to-analogue converter (DAC), a programmer, a comparator and a clock input. The programmer may contain a shift register, a control logic unit and an output register.

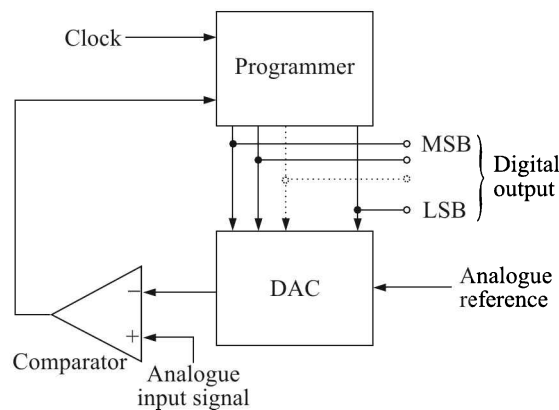


Fig. 16.66 Successive approximation type ADC.

At the outset, the programmer sets the MSB to 1 with all other bits to 0, and the comparator compares the DAC output with the analogue input signal. If the analogue input is larger, the 1 is retained. Otherwise, the 1 is replaced by a 0 and the next MSB is tried. This process continues, in order of descending bit weight, until the last bit has been tried and the binary equivalent obtained. Thus, for an N -bit system, the conversion time is N clock periods.

Note: If the analogue input signal changes considerably during the conversion process, the converter goes awry. That is why it is a general practice to use an S/H device ahead of the ADC.

Commercially available successive-approximation ADCs include the 12-bit AD 7589 of the Analog Devices, 10-bit ADC 80 AG-10 of Burr-Brown and 12-bit ADCH×12 BMM of Datel Intersil among others.

Voltage-to-time conversion method. In this type, a staircase waveform is generated and compared with the input analogue signal at each step. A digital counter measures the time period required to match the levels of the input and the staircase. This time period is calibrated with voltage. The arrangement is shown schematically in Fig. 16.67(a).

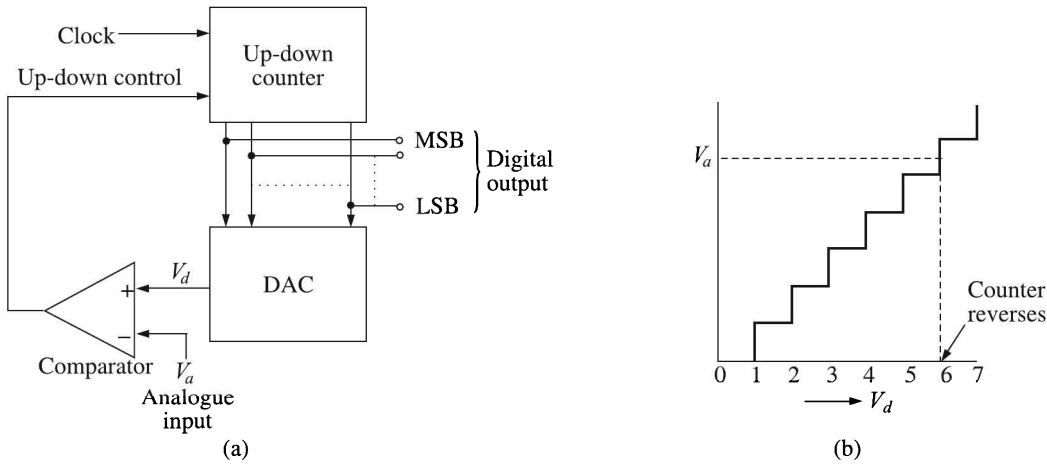


Fig. 16.67 Voltage-to-time conversion type ADC: (a) schematic circuit, and (b) the staircase output of the DAC.

The clock provides pulses at regular intervals. The pulse count increases linearly with time and therefore, the DAC output V_d generates a staircase waveform [Fig. 16.67(b)]. Suppose initially, $V_d < V_a$. Then the positive comparator output causes the counter to count UP. But V_d goes on increasing with every pulse from the clock until it exceeds V_a when the UP-DOWN control line changes and the counter starts counting DOWN. No sooner is V_d lowered by 1 step than the control changes state for counting UP and the count increases by 1 LSB. The process keeps repeating with the result that the digital output reads ± 1 LSB around the correct value.

Note: This method can be used as a tracking or servo converter because the conversion time is small for small changes in the sampled analogue signal.

Dual-slope integration method. In this method an integrator is used to integrate first an input signal voltage for a fixed period of time and then an accurate voltage reference with the reverse slope, and the time required to return to the starting voltage is measured. The input signal voltage is obtained from the ratio of the time periods and the value of the reference voltage. If T_1 is the fixed time period for which the integrator integrated input signal voltage, it is clear from Fig. 16.68 that

$$V'_o = -\frac{V_a T_1}{RC} \quad (16.48)$$

assuming that the capacitor was completely discharged at the beginning. The negative sign indicates that the input voltage is positive. Next the switch is thrown to the reference voltage V_R . Then the integrator will begin to ramp towards zero at a rate of V_R/RC , assuming that V_R is of opposite polarity as V_a . Here, since the integrator starts from V'_o , we can write

$$V_o = V'_o + \frac{V_R T_2}{RC} \quad (16.49)$$

where T_2 is the time required to reach zero voltage level. From Eqs. (16.48) and (16.49) it is easy to find

$$V_a = \frac{T_2}{T_1} |V_R| \quad (16.50)$$

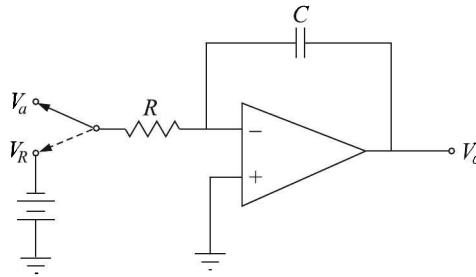


Fig. 16.68 Dual slope integration.

- Note:*
1. Equation (16.50) does not include R or C of the integrator.
 2. It includes the two time periods in the form of a ratio. Hence a stable rather than an accurate clock is needed.
 3. Because the input is integrated over a time period, the variation of the input signal is averaged and, therefore, no S/H device at the input is necessary.

A system that is widely used is shown in Fig. 16.69 (a).

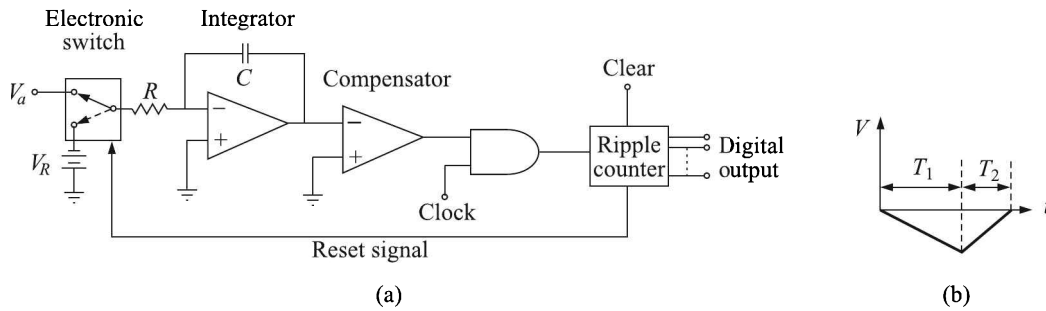


Fig. 16.69 Dual slope integration: (a) a widely used system, and (b) comparator output.

Initially, the ripple counter is cleared and the switch is connected to V_a . The analogue signal is integrated for a fixed number n_1 of clock pulses. If the clock period is T , after $T_1 = n_1 T$, the ripple counter reads 0. This change of state sends a control signal to the electronic switch and V_R gets connected to the input of the integrator. V_R is negative (or of opposite polarity to that of V_a), and $|V_R| > V_a$. Hence the integration time T_2 is less than T_1 . As long as V is negative, the output of the comparator remains high, the AND gate remains enabled and the counter keeps counting pulses [Fig. 16.69(b)]. Say, after n_2 counts, i.e. $T_2 = n_2 T$, the voltage falls to the level of V_a . At that moment V falls to zero, the AND gate is disabled and the ripple counter stops counting pulses.

Equation (16.50) can now be rewritten as

$$V_a = \frac{n_2 T}{n_1 T} |V_R|$$

Since $|V_R|$ and n_1 are constants, $V_a \propto n_2$ which is the count displayed at the ripple counter.

Datel Intersil ICL7109 is a dual-slope ADC which is commercially available.

Example 16.10

In a dual-slope ADC, the reference voltage is 100 mV and the first integration period is set at 50 ms. The input resistor of the integrator is 100 k Ω and the integrating capacitor 0.047 μ F.

- (a) For an output voltage of 120 mV, the second integration (de-integration) period will be
 (i) 50 ms (ii) 60 ms (iii) 100 ms (iv) 120 ms
- (b) If the input of 120 mV is corrupted by power supply interference of 50 Hz having peak amplitude of 3π mV, the worst-case error introduced by the interference in the reading is
 (i) 0% (ii) 1% (iii) 3% (iv) π %

Solution

Given: $V_R = 100$ mV, $V_a = 120$ mV and $T_1 = 50 \times 10^{-3}$ s.

- (a) From Eq. (16.50),

$$T_2 = \frac{V_a}{|V_R|} T_1 = \frac{120}{100} (50 \times 10^{-3}) = 60 \text{ ms}$$

\therefore Ans. (ii).

- (b) Equation (16.48) can be written as

$$V_o' = -\frac{1}{RC} \int_0^{T_1} (V_a + V_{\text{int}}) dt$$

where V_{int} is the interference voltage. Here, $V_{\text{int}} = 3\pi \sin[2\pi(50)t]$. So, its contribution to V_o' is

$$\int_0^{50\text{ms}} 3\pi \sin(100\pi t) dt = 0$$

because $100\pi T_1 = (100)(50 \times 10^{-3})\pi = 5\pi$. So, the error is 0%.

\therefore Ans. (i)

Voltage-to-frequency converter. Figure 16.70 shows a voltage-to-frequency converter.

It consists of an integrator that feeds a comparator which, in turn, drives a mono¹⁰. One output of the one-shot feeds a counter and another actuates an electronic switch which discharges the integrator via a current source.

The input voltage V_a to the integrator outputs a ramp with a negative slope. The comparator compares the output with a reference voltage V_R . No sooner the integrator outputs a zero than the one-shot produces an output pulse. Let T_1 be the time required to reach zero voltage. Then,

$$\frac{1}{C} \frac{V_a}{R} T_1 = V_R$$

or
$$T_1 = \frac{V_R}{V_a} RC \quad (16.51)$$

¹⁰Monostable multivibrator or *one-shot*.

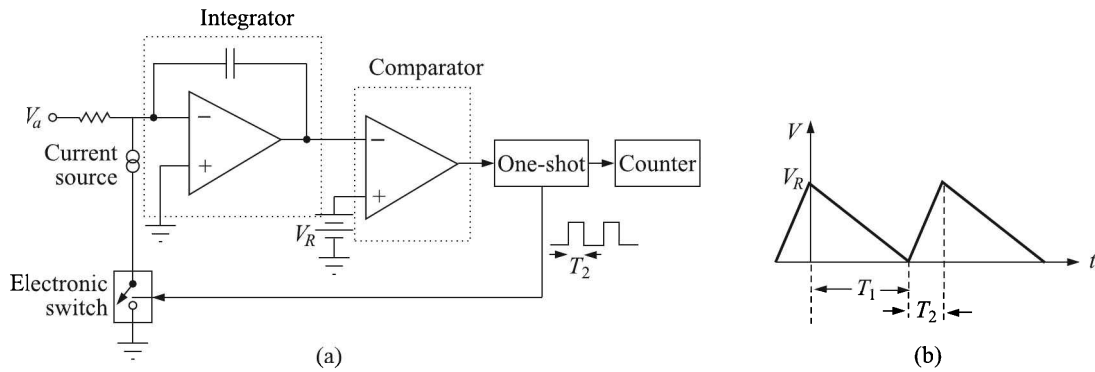


Fig. 16.70 Voltage-to-frequency converter: (a) schematic circuit, and (b) comparator input.

The output pulse of the one-shot actuates the electronic switch to drain current and for the duration of the pulse T_2 the output of the integrator ramps with a positive slope till the voltage level reaches V_R . In fact, during this period the capacitor discharges to the desired level of V_R . Hence,

$$V_R = \frac{1}{C} \left(I - \frac{V_a}{R} \right) T_2$$

or

$$T_2 = \frac{V_R C}{I - (V_a/R)} \quad (16.52)$$

where I is the current source output.

Eliminating V_R from Eq. (16.51) with the help of Eq. (16.52) we can write

$$\begin{aligned} T = T_1 + T_2 &= \frac{RC}{V_a} \left[\frac{I - (V_a/R)}{C} \right] T_2 + T_2 \\ &= \frac{IRT_2}{V_a} \end{aligned}$$

Hence the frequency f of the output signal, which the counter counts, is

$$f \equiv \frac{1}{T} = \frac{V_a}{IRT_2} \quad (16.53)$$

It is clear from Eq. (16.53) that $f \propto V_a$, other factors being constants of the circuitry.

Note: Since this method involves integration, it provides a true average of the signal over the ramp duration. The accuracy of the method depends on the stability of the one-shot and the current source and is comparable to that of the voltage-to-time conversion method.

Example 16.11

An analogue voltage signal, whose highest significant frequency is 1 kHz, is to be digitised with a resolution of 0.01 per cent covering a range of 0 to 10 V. Determine

- (a) the minimum number of bits in a digital code
- (b) analogue value of the LSB
- (c) rms value of the quantisation error
- (d) minimum sampling rate, and
- (e) aperture time required

Solution

Given: Highest significant frequency $f_{\max} = 1 \text{ kHz}$
 Resolution = 0.01%
 Range = 0 to 10 V

- (a) Since $V_{\max} = 10 \text{ V}$,

$$V_{\text{rms}} = \sqrt{2}V_{\max} \cong 14 \text{ V}$$

For 0.01% resolution, we need to resolve

$$14 \times \frac{0.01}{100} = 1.4 \text{ mV}$$

Therefore, the required number of levels is

$$\frac{14}{1.4 \times 10^{-3}} = 10^4$$

These levels can be provided by 14 bits because $2^{14} = 16384$

- (b) Analogue value of the LSB is

$$\frac{V_{\max}}{2^{14}} = \frac{10}{16384} = 610.4 \mu\text{V}$$

- (c) From Eq. (16.46), the rms value of the quantisation error is

$$\frac{610.4}{2\sqrt{3}} = 176.2 \mu\text{V}$$

- (d) Minimum sampling rate should be $2 \times f_{\max} = 2 \text{ kHz}$

- (e) Aperture time t_a is given by

$$\begin{aligned} t_a &= \frac{1}{(2^{14})(2\pi f_{\max})} \\ &= \frac{1}{16384 \times 2\pi \times 10^3} \\ &= 9.71 \times 10^{-9} \text{ s, i.e. } 9.71 \text{ ns} \end{aligned}$$

Digital-to-analogue Conversion

The principle of digital-to-analogue conversion involves developing voltage across precision resistors according to the value of the digital input data. We present below two digital-to-analogue converters (DAC) .

Simple DAC

A simple DAC arrangement is shown in Fig. 16.71.

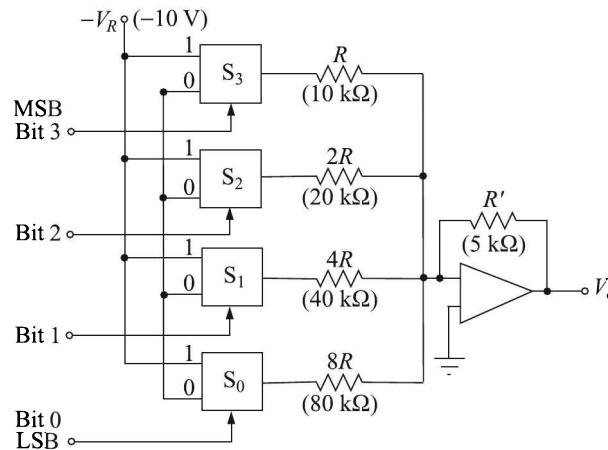


Fig. 16.71 Simple DAC.

It consists of

1. A voltage reference V_R
2. A network of precision resistors $R, \dots, 8R$
3. A set of digitally controlled switches S_0, \dots, S_3
4. An op-amp which acts as a current-to-voltage converter

The switches are electronic single-pole double-throw type which either connect or disconnect a resistor to the voltage source according to the input data.

- Note:*
1. The voltage source is negative because the op-amp is of inverting type.
 2. The resistors are weighted inversely according to the positional weight of the binary input.

Thus, if the MSB is a 1 and the rest 0s, the current passes only through the resistor R , and its value is $-V_R/R$. Then, the output V_o equals $V_R R'/R$. If all the 4 bits are 1s, the output is

$$V_o = \left(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8}\right) \frac{V_R R'}{R} = (8 + 4 + 2 + 1) \frac{V_R R'}{8R} \quad (16.54)$$

Equation (16.54) shows that the analogue output is proportional to the decimal value of the digital input.

The reliability of this kind of DAC mainly depends on the accuracy and thermal stability of the value of the resistors as discussed below.

1. If the number of bits is high, values of the resistors near the LSB become excessively high. For example, the LSB weighting resistor for a 12-bit DAC will be $10 \text{ k}\Omega \times 2^{12-1} = 20.48 \text{ M}\Omega$ if that for the MSB is $10 \text{ k}\Omega$. It is very difficult and expensive to obtain thermally stable, precision resistors of such values.
2. Since the value of this resistor determines the output voltage V_o [see Eq. (16.54)], the accuracy will be affected.
3. If, to obviate this difficulty, the highest resistance is chosen of a reasonable value, say $51.2 \text{ k}\Omega$, the value of the smallest becomes $25 \text{ }\Omega$ which becomes comparable to the output resistance of the electronic switch, thus affecting the accuracy.

The ladder-type DAC, described below, is free from this difficulty and therefore, is generally used for DACs of more than 4 bits.

Ladder-type DAC

The circuit arrangement in the ladder-type DAC is shown in Fig. 16.72. Because of its appearance, it is known as R - $2R$ network.

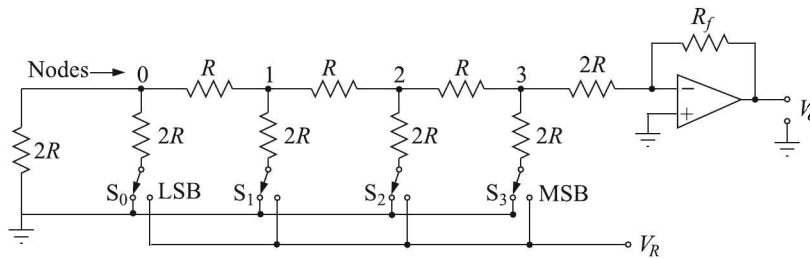


Fig. 16.72 Ladder type DAC.

The ladder basically splits the current according to the need. We have designated nodes in the ladder by 0, 1, 2 and 3. Consider the case when S_0 is connected to V_R and all other switches are connected to ground. The resulting resistive part of the circuit is shown in Fig. 16.73(a). We can say, here the input is 1 to the LSB and 0 to others.

The successive application of Thevenin's theorem reduces the network to one having the voltage of $V_R/16$ and a series resistance of $3R$ as shown in Fig. 16.73(e). If now S_1 is connected to V_R and the rest grounded, or in other words inputting 1 to the last but one bit and 0 to the rest, it can be seen in the same way that the resulting circuit consists of a voltage of $V_R/8$ and a series resistance of $3R$.

In this way, inputting 1 to other bits—one at a time—and 0 to others, and superposing the result, we arrive at the following expression for the current I

$$I = \frac{V_R}{3R} \left(\frac{S_0}{16} + \frac{S_1}{8} + \frac{S_2}{4} + \frac{S_3}{2} \right) \quad (16.55)$$

where the values of S_i 's will be either 0 or 1 depending on the input.

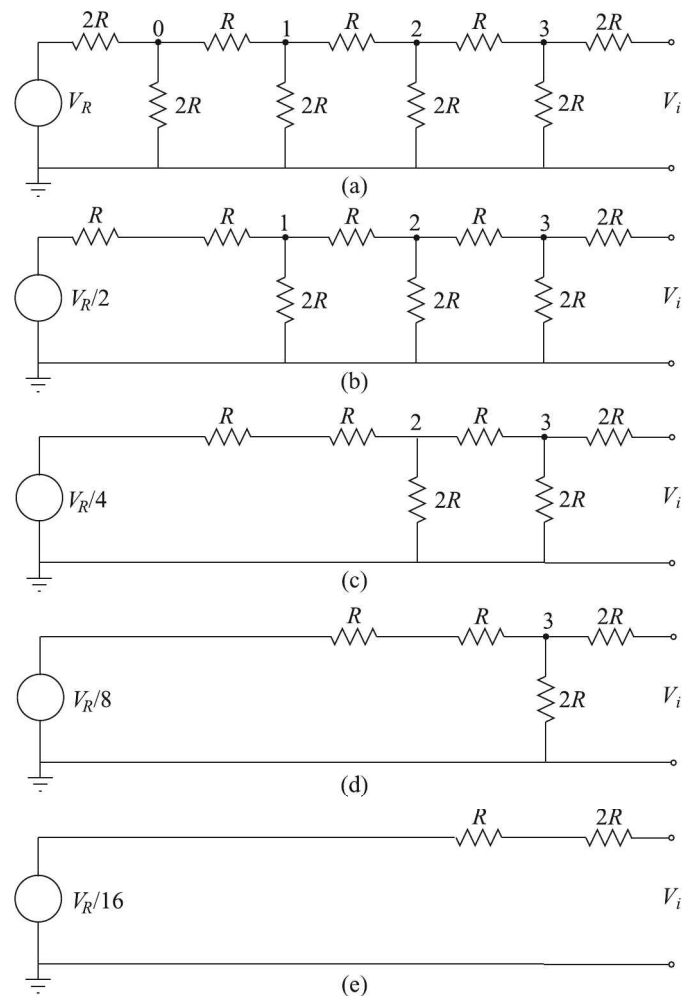


Fig. 16.73 Reduction of the R - $2R$ network to its Thevenin equivalent when the LSB is 1 and the rest 0.

Since the input to the op-amp is a virtual ground, the output voltage of the op-amp is given by

$$V_o = -IR_f = -R_f \frac{V_R}{3R} \left(\frac{S_0}{16} + \frac{S_1}{8} + \frac{S_2}{4} + \frac{S_3}{2} \right) \quad (16.56)$$

Equation (16.56) can also be written in the following convenient form

$$V_o = -\frac{V_R R_f}{48R} (2^3 S_3 + 2^2 S_2 + 2^1 S_1 + 2^0 S_0)$$

Example 16.12

A 3-bit resistance ladder D/A (R – $2R$ network) has resistor values of $R = 10 \text{ k}\Omega$ and $20 \text{ k}\Omega$. V_{ref} equals 8 volts. What is I_{out} for a digital input of 111?

Solution

Figure 16.72 shows a 4-bit DAC, ours is a 3-bit one. For an 111 input, the current I_{out} can be obtained from Eq. (16.55) as

$$I_{\text{out}} = \frac{V_R}{3R} \left(\frac{1}{8} + \frac{1}{4} + \frac{1}{2} \right) = \frac{8}{3(10 \times 10^3)} (0.875) \cong 0.2 \text{ mA}$$

Example 16.13

Consider the 4-bit digital-to-analogue converter shown in Fig. 16.72 where the logic levels 0 and 1 are 0.4 V and 2.4 V respectively and $R_f = 2R$. Find the analogue output V_o for an input of 1001.

Solution

From Eq. (16.56), we get on substituting 0.4 for 0 and 2.4 for 1

$$V_o = \frac{2}{3} \left(\frac{2.4}{2^4} + \frac{0.4}{2^3} + \frac{0.4}{2^2} + \frac{2.4}{2} \right) = 1 \text{ V}$$

We reiterate what we have stated at the outset that a complete discussion on signal conditioning encompasses the gamut of analogue and digital electronics. Such a discussion is beyond the scope of the present book.

We now move on to the next aspect of instrumentation, namely the display and recording devices.

Review Questions

- 16.1 Draw the circuit diagram of a non-inverting amplifier for a gain of 10 using an operational amplifier. Use minimum resistance value as 10 k Ω .
- 16.2 Draw the circuit diagram of an instrumentation amplifier.
- 16.3 State the problems associated in the design of dc amplifiers.
- 16.4 Define CMRR and explain its significance in difference amplifiers. Show the circuit diagram of a differential amplifier which can provide high CMRR and justify the same.
- 16.5 List the ideal characteristics of an op-amp.
- 16.6 What are the ideal conditions to be met by an op-amp? Explain the term 'CMRR', 'offset voltage', 'offset current' and 'slew rate' as referred to an op-amp.
How can you make unity gain amplifier with an op-amp?
- 16.7 (a) How will you realise a unity gain buffer using an op-amp?
(b) Determine the input impedance of a non-inverting op-amp.
(c) What is the common mode rejection ratio of an op-amp? What is the range of CMRR values for commercial op-amps?
- 16.8 What are the properties of an ideal op-amp? Explain how an op-amp may be used to perform addition of two signals, and integration of a signal.
- 16.9 Make a block diagram of a carrier-type ac signal conditioning system.

-
- 16.10 How can a simple ac amplifier be used to amplify a dc input signal? Explain with the help of suitable diagrams.
- 16.11 Explain the concept of amplitude modulation. Show how this concept can be utilised to measure dynamic strains of strain gauges.
- 16.12 Explain, with the help of block diagrams only, the construction and operation of chopper-stabilised dc amplifiers.
- 16.13 (a) Write down two advantages of smart-type of pressure transmitter over that of analogue-type.
- (b) Why is a 2-wire transmitter preferred to a 4-wire transmitter?
- (c) What is the advantage of 4–20 mA signal over a 0–20 mA signal as a standard for transmission?
- (d) What is the function of an equalising value in connection with the installation of a differential pressure transmitter?
- (e) If 4–20 mA is the standard for transmission of signal in electrical form, what is the standard for the same in pneumatic form?
- 16.14 Draw the block schematic diagram of a phase locked loop and explain its operation bringing out clearly the concept of lock range and capture range.
- 16.15 What overall accuracy can one expect from the construction of a 10-bit A/D converter?
- 16.16 Suppose that the counter type A/D converter is 8-bit and driven by 500 kHz clock. Find
- (a) the maximum conversion time, and
- (b) maximum conversion rate
- 16.17 What are the various applications of A/D and D/A converters in the field of instrumentation? Describe the operation of a D/A converter with the help of its circuit diagram.
- 16.18 What do you mean by the resolution of a digital-to-analogue converter? ‘A 10-bit D/A converter has 0.1% resolution’. Explain.
- 16.19 Why do you need a sample-and-hold circuit? Explain with diagram a sample-and-hold circuit.
- 16.20 What is successive approximation logic of analogue-to-digital conversion?
- 16.21 Pick the right statement
- (a) A 10-bit successive approximation type A/D converter working in the range 0–10 V takes 50 μs to convert an analogue signal of 10 volts. How much time would it take to convert 5 volts?
- (i) 25 μs
- (ii) slightly more than 25 μs
- (iii) slightly less than 25 μs
- (iv) 50 μs

- (b) The following statements refer to A/D converters
- (i) successive approximation converters provide only an approximate conversion of analogue signals whereas the counter-type provides an exact value.
 - (ii) in all converters the time for conversion depends on magnitude of the input.
 - (iii) a 10-bit converter with an input range of ± 10 V will have a quantisation error of about 5 mV.
 - (iv) a 12-bit converter cannot be interfaced to an 8-bit computer.
- (c) A band limited signal with the highest frequency content of 1000 Hz is undergoing sampling at uniform intervals. For recovery of the signal in an unambiguous way, the sampling frequency should necessarily greater than
- (i) 500 Hz
 - (ii) 100 Hz
 - (iii) 1500 Hz
 - (iv) 2000 Hz
- (d) A $3\frac{1}{2}$ digit DVM has a lowest measuring range of 200 mV. Hence it can be concluded that
- (i) its best resolution is 0.1 mV
 - (ii) its poorest resolution is 0.2 mV
 - (iii) its accuracy is at least 0.05%
 - (iv) the maximum voltage that can be measured in this range is 199.9 mV
- (e) An 8-bit ADC outputs all 1's when $V_{in} = 1.275$ volts. The quantisation error is
- (i) +5 mV
 - (ii) -5 mV
 - (iii) 10 mV
 - (iv) ± 2.5 mV
- (f) The differential input resistance of the circuit shown in Fig. 16.74 is

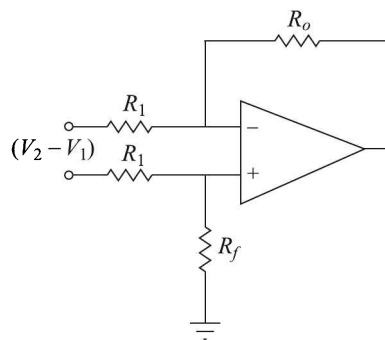


Fig. 16.74

- (i) R_1
 - (ii) $R_1/2$
 - (iii) $2R_1$
 - (iv) R_f
- (g) In the bridge circuit shown in Fig. 16.75, when $(X_C/R) = 1$, the voltmeter reads
- (i) 5 V
 - (ii) 0.0 V
 - (iii) 2.5 V
 - (iv) 10 V

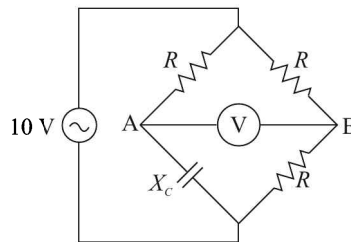


Fig. 16.75

- (h) The operational amplifiers use a differential input stage with a constant current source mainly to obtain
- (i) very low common mode gain
 - (ii) very high differential gain
 - (iii) very low input noise
 - (iv) very high input resistance
- (i) A sample-and-hold circuit is normally required before the following type of A/D converter
- (i) successive approximation
 - (ii) flash (parallel) converter
 - (iii) voltage-to-frequency converter
 - (iv) dual slope integrator
- (j) Percent resolution of an 8-bit D/A converter is
- (i) 0.39
 - (ii) 0.78
 - (iii) 2.56
- (k) The full scale input voltage to an ADC is 10 V. The resolution required is 0.5 mV. The minimum number of bits required for ADC is
- (i) 8
 - (ii) 10
 - (iii) 11
 - (iv) 12

- (l) A phase locked loop can be employed for demodulation of
- pulse amplitude modulated signals
 - pulse code modulated signals
 - frequency modulated signals
 - single side-band amplitude modulated signals
- (m) The advantage of a dual slope converter over successive approximation converter is that the dual slope converter
- is faster
 - eliminates error due to drift
 - can reduce the error due to power supply
 - does not require a stable voltage reference
- (n) A twisted pair of wires is used for connecting the signal source with the instrumentation amplifier, as it helps reducing
- the effect of external interference
 - the error due to bias currents in the amplifier
 - the loading of the source by the amplifier
 - the common mode voltage
- (o) Majority of digital voltmeters are built with a Dual-slope ADC because
- dual slope ADCs are less complex than other types of ADCs
 - dual slope ADCs are faster than other types of ADCs
 - dual slope ADCs can be designed to be insensitive to noise and interference
 - dual slope ADCs provide BCD outputs
- (p) The wiper of the $20\text{ k}\Omega$ potentiometer in Fig. 16.76 is positioned half way. Then the voltage V_o of the circuit is

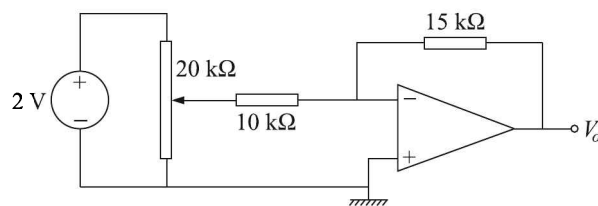
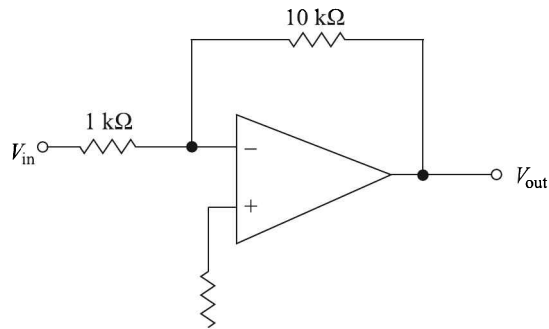


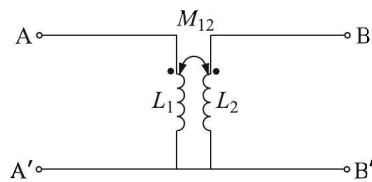
Fig. 16.76

- -1.5 V
- -1.0 V
- -0.75 V
- $+2.5\text{ V}$

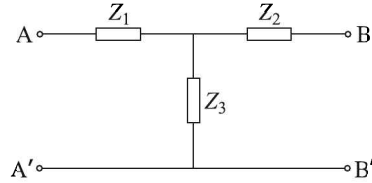
- (q) For N -bit successive approximation ADCs, other parameters such as clock frequency remaining constant, the conversion time is proportional to
- N^2
 - \sqrt{N}
 - $\log N$
 - N
- (r) If the value of the resistance R in the following figure is increased by 50%, then voltage gain of the amplifier shown in the figure will change by



- 50%
 - 5%
 - 50%
 - negligible amount
- (s) The potential difference between the input terminals of an op-amp may be treated to be nearly zero if
- the two supply voltages are balanced
 - the output voltage is not saturated
 - the op-amp is used in a circuit having negative feedback
 - there is a dc bias path between each of the input terminals and the circuit ground
- (t) Consider the coupled circuit shown below.

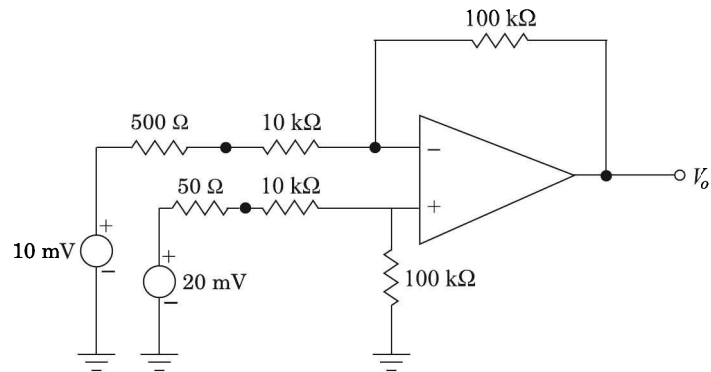


At angular frequency ω , this circuit can be represented by the equivalent T-network shown below.



Indicate the correct set of expressions for the impedances of the T-network.

- (i) $Z_1 = j\omega(L_1 - M_{12})$, $Z_2 = j\omega(L_2 - M_{12})$, $Z_3 = j\omega M_{12}$
 - (ii) $Z_1 = j\omega(L_1 + M_{12})$, $Z_2 = j\omega(L_2 + M_{12})$, $Z_3 = j\omega M_{12}$
 - (iii) $Z_1 = j\omega L_1$, $Z_2 = j\omega L_2$, $Z_3 = -j\omega M_{12}$
 - (iv) $Z_1 = j\omega(L_1 - M_{12})$, $Z_2 = j\omega(L_2 - M_{12})$, $Z_3 = j\omega(L_1 + L_2 + M_{12})$
- (u) An ideal op-amp has the characteristics of an ideal
- (i) voltage controlled voltage source
 - (ii) voltage controlled current source
 - (iii) current controlled voltage source
 - (iv) current controlled current source
- 16.22 Discuss the principle of operation of a D/A converter with special resistive divider known as $R-2R$ ladder network. What are its advantages compared to other DACs?
- 16.23 Draw the circuit diagram of a slope A/D converter and explain its operation.
- 16.24 Explain the operation of a basic digital-to-analogue converter using a resistive divider network.
- 16.25 Describe, with a neat circuit diagram, the operation of a counter-type A/D converter.
- 16.26 In a voltage-controlled oscillator, 550 pulses stored in a counter for the input V after passing through AND gate during 100 ms gating pulse. Calculate the input voltage if the ratio of the input voltage to the output frequency for the voltage controlled oscillator is 0.2.
- 16.27 Explain the performance of a dual slope analogue-to-digital converter.
- 16.28 Why is the dual slope integrating type A/D converter preferred for the digital multimeter?
- 16.29 Obtain the binary representation of an analogue signal of magnitude 2.875 V using the SAR method. Assume that successive approximation A-to-D converter has a reference voltage of 5 V.
- 16.30 The figure shows a single op-amp differential amplifier circuit.



Which one of the following statements about the output is correct?

- (a) $V_o \leq 95\text{ mV}$
- (b) $95\text{ mV} < V_o \leq 98\text{ mV}$
- (c) $98\text{ mV} < V_o \leq 101\text{ mV}$
- (d) $V_o > 101\text{ mV}$

Display Devices and Recording Systems

Display devices constitute the interface between the instrument and the human observer and are, therefore, the final stages of instruments.

17.1 Classification and Comparison

These devices can broadly be divided into two categories

1. Analogue
2. Digital

The display of a moving-coil voltmeter—pointer and scale—is of the former category while that of a pocket-calculator—seven segment display—belongs to the latter category. Digital displays have many plus points as will be evident from the following comparison (Table 17.1).

Table 17.1 Comparison between analogue and digital displays

| <i>Property</i> | <i>Analogue</i> | <i>Digital</i> |
|----------------------------|---|--|
| Resolution | 1 part in several hundreds | 1 part in several thousands |
| Precision | Around 0.1% of the FSD | Can be much higher |
| Range | Generally single range. For multiple ranges, the selection is done manually | Generally auto-ranging |
| Polarity | Needs to be ascertained before connecting the instrument | Automatic selection of polarity |
| Observational error | Susceptible | Free |
| Loading the previous stage | May load | Negligible because of their high input impedance |
| Power input | Higher | Lower |
| Display | Continuous, having moving parts with their associated disadvantages, like getting stuck | Discrete, but no moving parts |

17.2 Characteristics of Digital Display

The display of analogue instruments consists generally of mechanical devices which are familiar to us, while that of digital instruments comprises certain display elements which are excited by electrical means. Before we consider these display elements, let us consider a few aspects of digital displays like

1. Specification
2. Resolution
3. Sensitivity
4. Accuracy

which are different to some extent from their analogue counterparts because of their discrete nature.

Specification. For convenience in presentation of data, the display generally shows readings in decimal format. Therefore, if it is 4-digit display, the maximum number which it can display is 9999. Most of the displays, however, have one more digit on the left which can display an overflow of 1. In a 4-digit display, the maximum displayable quantity is thus 19999. Such a display is called a $4\frac{1}{2}$ -digit display.

Resolution. Suppose, we are measuring voltage with a $4\frac{1}{2}$ -digit display and our selected range is 1 volt. The instrument now can display up to 1.9999 V (this is why some prefer to call it a 2 V range). Here the resolution is 0.0001 V. If the range is 10 V, the maximum displayable voltage is 19.999 V, with a resolution of 0.001 V. Thus, the resolution R can be defined as

$$R = \frac{\text{Range}}{10^n}$$

where n is the number of full digits in the display.

Sensitivity. It is the resolution of the instrument in the lowest range. In our preceding example, if 1 V is the lowest range, the sensitivity of the voltmeter is 0.0001 V or 0.1 mV.

Accuracy. The accuracy of digital instrument is generally specified as $\pm x\%$ of the reading of $\pm n$ digits. Thus, if 8 V is being measured in the 10 V range of $4\frac{1}{2}$ -digit display instrument, and if the accuracy of the instrument has been specified by the manufacturer as $\pm 1\%$ of the reading ± 1 , the total error in the measurement can be calculated as follows:

$$\begin{array}{r} \pm 0.1\% \text{ of } 8 \text{ V} = \pm 0.008 \text{ V} \\ \pm 1 \text{ digit} = \pm 0.001 \text{ V} \\ \hline \text{Total error} = \pm 0.009 \text{ V} \end{array}$$

Because of this kind of specification of accuracy, one has to be careful in the selection of ranges while measuring by a digital instrument. The following example will make the point clear.

Example 17.1

What are the errors in the measurement of 0.1 V in the (a) 10 V and (b) 1 V ranges of a $4\frac{1}{2}$ -digit voltmeter having an accuracy of $\pm 0.1\%$ of the reading ± 1 ?

Solution

(a) Consider measurement in the 10 V range. The error can be calculated as,

$$\begin{array}{r} \pm 0.1\% \text{ of } 0.1 \text{ V} = \pm 0.0001 \text{ V} \\ \pm 1 \text{ digit} = \pm 0.001 \text{ V} \\ \hline \text{Total error} = \pm 0.0011 \text{ V} \end{array}$$

$$\text{Thus error} = \frac{\pm 0.0011 \times 100}{0.1} = \pm 1.1 \%$$

(b) Next, consider the same measurement in the 1 V range. Here, the error can be calculated as:

$$\begin{array}{r} \pm 0.1\% \text{ of } 0.1 \text{ V} = \pm 0.0001 \text{ V} \\ \pm 1 \text{ digit} = \pm 0.0001 \text{ V} \\ \hline \text{Total error} = \pm 0.0002 \text{ V} = \pm 0.2\% \end{array}$$

Note: The first measurement is 5.5 times worse than what the instrument can achieve.

17.3 Digital Display Elements

Digital displays can be subdivided into two categories

1. Alphanumeric display
2. Graphic and pictorial display

according to the requirement of what is to be displayed. We discuss them individually.

Alphanumeric Display

These displays, as the name suggests, can generate letters of alphabet (A to Z), decimal numerals (0 to 9), punctuation marks and many symbols. The displays are made with the help of *display elements*. After a considerable experimentation with different kinds of elements, two kinds, namely

1. Light emitting diode
2. Liquid crystal display

seem to have gained acceptance because of their low power consumption.

Light emitting diode (LED)

It is well known that a *p*-type semiconductor contains excess positive carriers or holes and an *n*-type, excess electrons.

If a *p-n* junction is forward-biased, i.e. positive voltage is applied to the *p*-material and negative to the *n*-material, the excess electrons in the *n*-material will recombine with the holes of the *p*-material. The process is akin to the transition of an electron from the conduction band to the valence band releasing energy in the form of light quanta. If the *p*- and *n*-materials

are transparent and if the energy difference $\Delta E = E_c - E_v$ corresponds to a wavelength in the visible region of the spectrum through the Planck relation

$$\Delta E = h\nu = \frac{hc}{\lambda}$$

where symbols have their usual meaning, then the released energy will be visible in the form of light (Fig. 17.1).

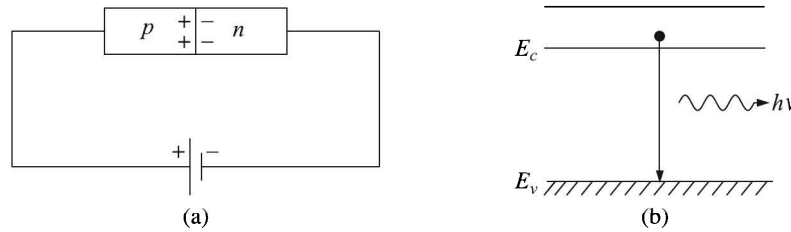


Fig. 17.1 A *p-n* junction: (a) forward-biasing, and (b) emission of radiation. E_c and E_v indicate conduction and valence band energies respectively.

This principle is utilised in the construction of LEDs where gallium arsenide (GaAs) or gallium arsenic phosphide (GaAsP) crystals fit the bill. Gallium arsenide being a Group III–IV compound, it can be made *p*-type by doping it with a Group IV element like sulphur, selenium or tellurium. Normally the emitted light is red, but yellow or green light may be generated with suitable dopants.

Since holes have a lower mobility in semiconductors, more so in GaAs, the recombination takes place in the *p*-region. Hence, the anode is made in the form of a ring while the *n*-material is completely covered with a thin film of the cathode material¹, which also acts as a mirror and reflects the emitted light (Fig. 17.2).

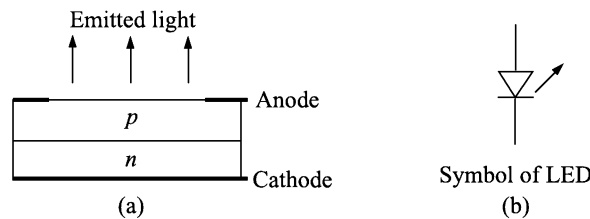


Fig. 17.2 LED: (a) Construction, and (b) symbol.

LEDs can be fabricated in very small sizes. About 1.2 V dc is sufficient to excite them and a typical forward current might be 5 mA. This makes them suitable to be driven directly by the transistor transistor logic (TTL) in contrast to earlier gas-discharge devices which needed an intermediate 5 V transistor stage. LEDs are very fast devices, the switching time being of the order of nanoseconds, and they can tolerate temperature variation from 0 to 70°C.

Liquid crystal display (LCD)

A liquid crystal is a complex organic liquid that shares certain properties of a crystal—such as its molecules are arranged in an ordered, repetitive pattern as occurring in a crystal lattice.

¹Usually gold which establishes ohmic contact with the semiconductor.

At the same time it behaves as a liquid—its molecular arrangement changes easily when an electric field is applied to it. A few hundred such liquids are known. A simple liquid crystal is ammonium oleate, $C_{17}H_{33}COONH_4$.

Based on the processes which produce displays, LCDs can be classified into two types—dynamic scattering type and field effect type. They are made by sandwiching a 10 to 12 μm layer of a liquid crystal fluid between two glass plates. A transparent, electrically conducting film, or *backplane* is coated on the rear glass plate. On the front glass plate transparent sections of conductive film in the shape of desired characters are coated. An electric field is produced when a voltage is applied between a segment and the backplane.

In the former type, the field scrambles the molecules changing the liquid crystal from being transparent to being milky opaque. This produces etched-glass-looking character on a dark background.

In the field-effect type (Fig. 17.3), light falling on the top gets polarised at a certain plane by the polariser, passes through the liquid crystal only to change its plane of polarisation by 90° and so easily passes through the crossed polariser (called *analyser*) to reach the bottom mirror. The light beam then reflects its original path to make the segment appear bright. But once an electric field is applied to a segment, the beam is no longer twisted and, therefore, blocked by the analyser to make the segment appear dark.

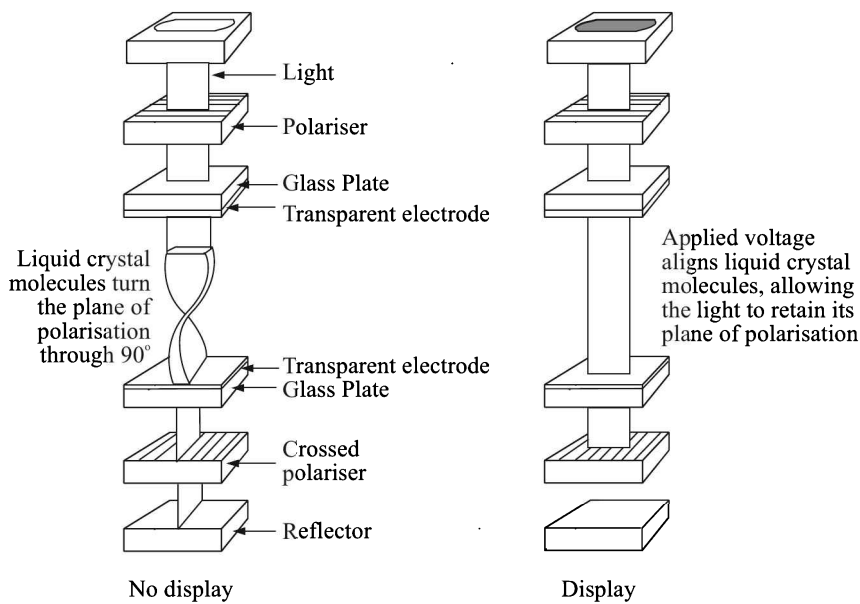


Fig. 17.3 Construction and principle of operation of a field-effect type LCD.

Thus the brightness of an LCD depends upon the incident light, since it does not generate any light itself. It cannot be excited by dc because of its degradation by electrolytic action and, therefore, it cannot be driven by the outputs of a TTL decoder, connecting the backplane to ground. To prevent a dc build-up on the segments, the drive signals should be square waves with a frequency of 30 to 150 Hz and less than 10 V peak to peak. CMOS gates are often used to drive LCDs. An LCD has a slower switching, typical time required for ON and OFF being 1 ms and 10 ms. Also it should be kept around $20 \pm 10^\circ\text{C}$ to have maximum lifespan.

Nevertheless, LCDs are popular because they are cheap. The relative advantages and disadvantages of LEDs and LCDs will be apparent from the comparison given in Table 17.2.

Table 17.2 Comparison between LED and LCD displays.

| <i>Property</i> | <i>LED</i> | <i>LCD</i> |
|-----------------------------|---|--|
| Material | Solid (GaAs, GaP) | Liquid (organic) |
| Power consumption | About 40 mW/numeral | About 140 μ W/seven-segment |
| Power supply | dc (1.2 V) | ac (10 V p-p, 50 Hz) |
| Switching time | Fast, < 1 ns | Slow, ON 1 ms, OFF 10 ms |
| Cost | Higher | Lower |
| Size | Smaller, each diode about 0.4 mm square, typical height of each character is 7 mm | Bigger, typical height of each character is 10–30 mm |
| Operating temperature range | Wide, 0–70°C | Restricted, 10–30°C |
| Luminosity | Self-illuminated. Hence visible in the dark | To be illuminated |

Display systems

Display systems are constructed with the help of display elements, and are generally of two types

1. Dot-matrix
2. Seven-segment

Since LEDs occupy smaller areas, they are preferred to construct dot-matrix displays while seven-segment displays are preferably constructed of LCDs which occupy larger areas.

Dot-matrix systems. As the name implies, in this display system tiny dots are arranged in regular arrays, and letters of alphabet or digits are generated by activating suitable dots.

Generally 3×5 dot-matrix is used to display numbers and 5×7 to display alphanumeric characters. The circuit arrangement and display pattern for a 5×7 dot-matrix is shown in Fig. 17.4. LEDs are arranged in rows, their connections having been shown in Fig. 17.4(a). Here, the 5th row and 2nd column switches are ON and hence the LED in the corresponding row and column glows.

In this way any LED in the matrix can be *selected* by energising the corresponding row and column. But an entire character cannot be generated at a time. It can be generated by selectively making rows ON of a particular column and then energising that column. The process may be repeated for the next column and so on sequentially and in rapid succession such that the display does not flicker.

Displaying something by a dot-matrix is an old concept. Previously block-makers used to construct half-tone blocks for printing pictures by converting them to arrays of tiny dots. We will find later in this chapter that this concept has been utilised to construct a host of printers including laser and inkjet printers which can print not only alphanumeric characters but also colourful graphics.

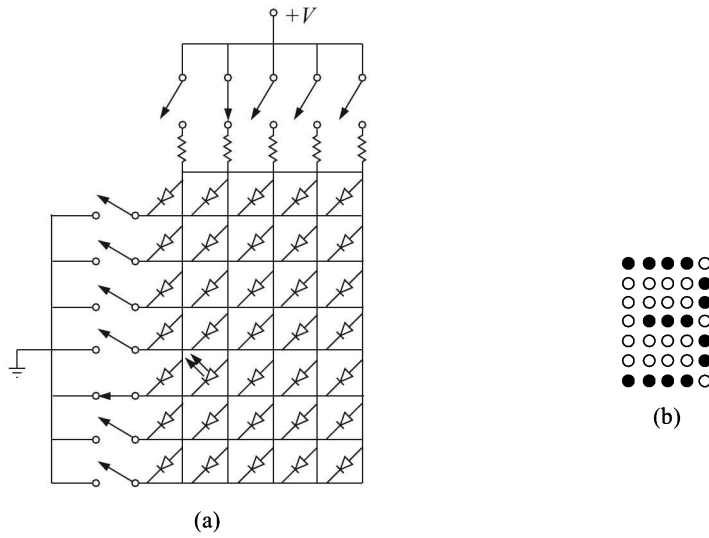


Fig. 17.4 A 5 × 7 dot-matrix display: (a) circuit arrangement, and (b) display pattern.

Seven-segment systems. The appearance of a seven-segment display is shown schematically in Fig. 17.5(a).

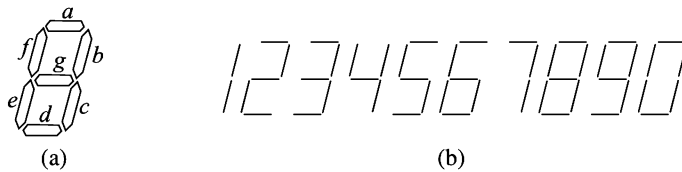


Fig. 17.5 (a) Seven-segment display, and (b) display of digits by energising appropriate segments.

Seven-segment displays are generally constructed with LCDs. Segments are named by letters of alphabet *a* to *g* in the order shown in the diagram. By energising appropriate segments, different digits can be displayed as shown in Fig. 17.5(b). TTL IC decoders are available to convert BCD numbers to seven-segment display. IC 7447 is one such device. The wiring of the display is simplified by having one electrode—*anode* [Fig. 17.6(a)] or *cathode* [Fig. 17.6(b)]—common.

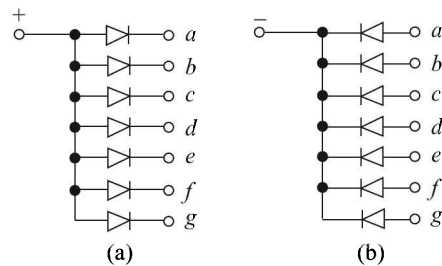


Fig. 17.6 Seven-segment systems: (a) common anode connection, and (b) common cathode connection.

The decoder and the display can be interconnected as shown in Fig. 17.7. However, in this arrangement one decoder is needed for each digit. This can be economised by using a

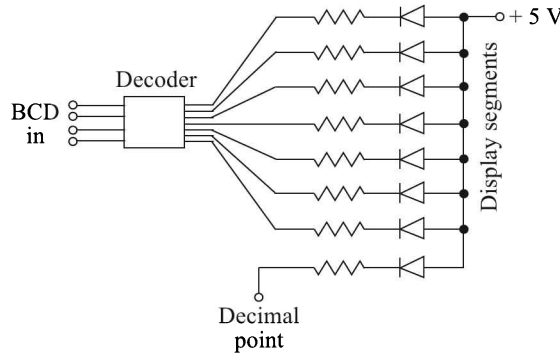


Fig. 17.7 Interconnection of the decoder and the display segments.

multiplexer and energising each digit in rapid succession as shown in Fig. 17.8. At any instant, only one of the digits is selected and the corresponding BCD code is decoded and fed to the appropriate display. A refresh rate of 50 to 100 Hz makes the display flicker free.

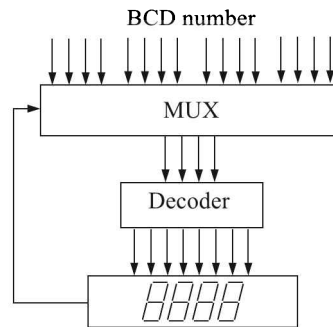


Fig. 17.8 Use of a multiplexer to reduce the number of decoders.

It may be mentioned here that we should not develop an idea that LCDs are suitable for constructing only seven-segment displays of pocket calculators and that they can only generate a black-and-white display. LCDs have been successfully utilised to construct dot-matrix displays for computers as well as TV screens where colourful graphics and video pictures can be displayed.

Graphic Display

In principle a graphic display can be made either by energising phosphors on a screen with the help of a scanning electron beam or by supplying power to individual LEDs or LCDs arranged in large arrays on a screen. In general, such a display system is called a video display unit (VDU) or *monitor*. The former kind is called cathode ray tube (CRT) which is still in use, but is steadily being phased out.

Cathode ray tube

A cathode ray tube is a special vacuum tube in which images are produced when an electron beam strikes a phosphorescent screen. Most desktop computer displays make use of CRTs. The CRT in a computer display is similar to the ‘picture tube’ in a television receiver.

The cathode ray tube consists of several basic components, as illustrated in Fig. 17.9. A narrow beam of electrons, generated by the electron gun, is accelerated by applying a high voltage at the anode.

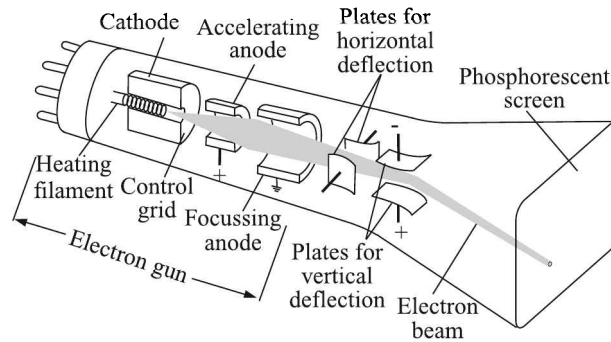


Fig. 17.9 Schematic diagram of a cathode ray tube. The conductive coating on the inside wall of the tube is not shown.

Two sets of deflecting plates—horizontal and vertical—produce an extremely low frequency electromagnetic field that allows for constant adjustment of the direction of the electron beam. The intensity of the beam can be varied.

Motion of an electron in a CRT. To analyse the electron motion, let us first calculate the velocity v_x of the electrons as they leave the electron gun. Since the initial velocities at which the electrons are emitted from the cathode are very small compared to their final velocities, we assume that the initial velocities are zero. Then the velocity of the electrons as they leave the electron gun is given by

$$\frac{1}{2}m_e v_x^2 = eV$$

or

$$v_x = \sqrt{\frac{2eV}{m_e}} \quad (17.1)$$

where e and m_e are the mass and the charge of the electron, and V is the accelerating anode potential. We note that the kinetic energy of an electron leaving the anode depends only on the potential difference between the anode and the cathode, and not on the details of the fields or the electron trajectories within the electron gun. To have an idea of the velocities of electrons, if $V = 2000$ volt,

$$v_x = \sqrt{\frac{2(1.6 \times 10^{-19} \text{ C})(2 \times 10^3 \text{ V})}{9.11 \times 10^{-31} \text{ kg}}} = 2.65 \times 10^7 \text{ m/s}$$

This is about 9% of the velocity of light in a vacuum.

If there is no electric field between the horizontal-deflection plates, the electrons enter the region between the vertical-deflection plates with velocity v_x . If there is a potential difference V' between these plates, with the upper plate at higher potential, there is a downward electric field with magnitude $E_y = V'/d$ between the plates, where d is the distance between the plates. A constant upward force of magnitude eE_y then acts on the electrons. Therefore, their upward acceleration is

$$f_y = \frac{eE_y}{m_e} = \frac{eV'}{m_e d} \quad (17.2)$$

The horizontal component of velocity v_x is constant. Because they move with a constant velocity in the x -direction and a constant acceleration in the y -direction, the electrons describe a parabolic path between the plates. After the electrons emerge from this region, their paths again become straight lines, and they strike the screen at a point a distance d_y above its centre. We can prove that this distance is directly proportional to the deflecting potential difference V' .

Time t required for the electrons to travel the length L of the plates is

$$t = \frac{L}{v_x} \quad (17.3)$$

During this time, they acquire an upward velocity component v_y given by

$$v_y = f_y t \quad (17.4)$$

Combining Eqs. (17.2), (17.3) and (17.4), we get

$$v_y = \frac{eV'}{m_e d} \frac{L}{v_x} \quad (17.5)$$

When the electrons emerge from the deflecting field, their velocity \mathbf{v} makes an angle θ with the x -axis given by

$$\tan \theta = \frac{v_y}{v_x} \quad (17.6)$$

Ordinarily, the length L of the deflection plates is much smaller than the distance D from the plates to the screen (Fig. 17.10).

Therefore, the angle θ is also given approximately by $\tan \theta = d_y/D$. Substituting this in Eq. (17.6), we get

$$\frac{d_y}{D} = \frac{v_y}{v_x}$$

Eliminating v_x and v_y by using Eqs. (17.1) and (17.5), we obtain

$$d_y = \left(\frac{LD}{2d} \right) \frac{V'}{V} \quad (17.7)$$

The factor in parentheses depends only on the geometry of the system, and therefore, it is constant for a particular tube. So, we observe that the deflection d_y is proportional to the deflecting voltage V' . It is also inversely proportional to the accelerating voltage V . This is obvious because the faster the electrons are going, the less they are deflected by the deflecting voltage.

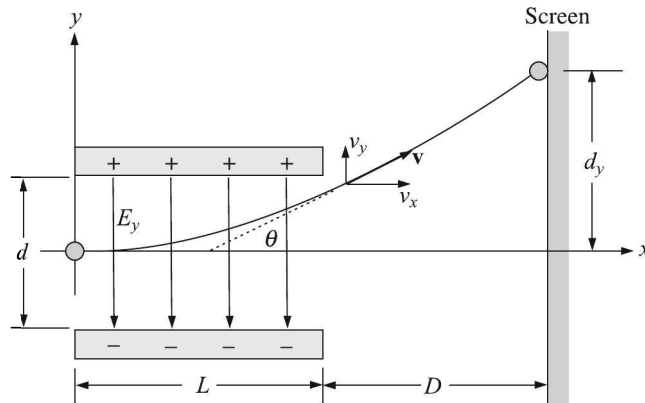


Fig. 17.10 Electrostatic deflection of an electron in a CRT by the vertical-deflection plates.

If there is also a field between the horizontal deflecting plates, the beam is also deflected in the horizontal direction, perpendicular to the plane of Fig. 17.10. The coordinates of the luminous spot on the screen are then proportional to the horizontal and vertical deflecting voltages, respectively. This is the principle of the cathode ray oscilloscope. If the horizontal deflection voltage sweeps the beam from left to right at a uniform rate, the beam traces out a graph of the vertical voltage as a function of time.

The picture tube in a television set is similar, but the beam is deflected by magnetic fields rather than electric fields. In some systems, the electron beam traces out the area of the picture 30 times per second in an array of 525 horizontal lines, and the intensity of the beam is varied to make bright and dark areas on the screen. The accelerating voltage V in TV picture tubes is typically about 20 kV. Computer displays and monitors operate on the same principle, using a magnetically deflected electron beam to trace out images on a phosphorescent screen.

Example 17.2

The vertical deflection plates of a CRT are 5 cm long and are situated 1 cm apart. A deflecting voltage of 50 V is applied between them. The screen is 20 cm away from the plates and the accelerating voltage is 1 kV. Calculate

- transverse acceleration of electrons moving between the plates
- deflection of the spot on the screen from its centre

Solution

Given, $V = 1000$ V, $L = 5$ cm = 0.05 m, $d = 1$ cm = 0.01 m, $V' = 50$ V, and $D = 20$ cm = 0.2 m.

- Substituting these values in Eq. (17.2), we get

$$\text{Transverse acceleration } f_y = \frac{1.6 \times 10^{-19} \text{ C}}{9.11 \times 10^{-31} \text{ kg}} \cdot \frac{50}{0.1} = 8.78 \times 10^{14} \text{ m/s}^2$$

- Similarly from Eq. (17.7), we get

$$d_y = \left(\frac{0.05 \times 0.2}{2 \times 0.01} \right) \cdot \frac{50}{1000} = 0.025 \text{ m} = 2.5 \text{ cm}$$

Example 17.3

What is the minimum distance that will allow full deflection of 5 cm at the CRT screen with a deflection factor of 100 V/cm and with an acceleration potential of 2000 V?

Solution

The deflection factor G is defined as the deflecting voltage requires to deflect the spot on the screen by unit distance. Thus from Eq. (17.7),

$$G = \frac{2V'd}{LD}$$

or

$$D = \frac{2Vd}{LG}$$

Now, for maximum possible deflection, we have from the geometry of Fig. 17.10

$$\frac{D}{d_y} = \frac{L}{d}$$

Multiplying these two equations and on rearranging, we get

$$\frac{D^2}{d_y} = \frac{2V}{G}$$

or

$$D = \sqrt{\frac{2Vd_y}{G}} = \sqrt{\frac{2 \times 2000 \text{ (V)} \times 0.05 \text{ (m)}}{10000 \text{ (V/m)}}} = 0.141 \text{ m} = 14.1 \text{ cm}$$

Figure 17.9 shows only one electron gun which is typical of a monochrome, or single-colour, CRT. However, virtually all CRTs today produce multicolour images. These devices have three electron guns for three primary colours—red (R), green (G) and blue (B). The CRT thus produces three overlapping images in red, green and blue. This is the so-called RGB colour model.

Scanning. We have discussed how the electron beam produces a tiny, bright visible spot when it strikes the phosphor-coated screen. To produce an image on the screen, complex signals are applied to the deflecting plates, and also to the circuitry that controls the intensity of the electron beam. The process is called *scanning*. The scanning technique can be either of the two types

1. Raster² scan
2. Vector scan

Raster scan. In the raster scan technique, the electron beam is made to move horizontally along a line across the screen, then displaced downwards to the next line, and to scan it horizontally again. The process is repeated until the entire screen is covered within a time which is less than the persistence time of human vision, i.e. 1/16 second. As viewed from the front of the CRT, the spot moves in a pattern similar to the way our eyes move when we read

² *Raster* is a German word which means screen.

a page of text. But the scanning takes place at such a rapid rate that our eye sees a constant image over the entire screen. Raster scan is done in TV screens and is universally adopted in CRT displays used in the PCs.

The CRT screen is made of dots of phosphors called *pixels*³. Any dot is illuminated when the electron beam of sufficient intensity falls on it (Fig. 17.11).

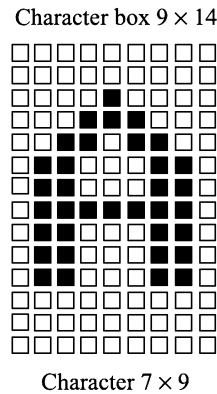


Fig. 17.11 Character generation by pixels or picture elements.

During the raster scan, appropriate pixels are illuminated by increasing the intensity of the beam when it strikes them and others are kept invisible by reducing the intensity when the beam traverses them, and thus a pattern is generated on the screen. The retrace portion of the raster scan is suppressed by reducing the intensity of the beam during retrace.

Typical horizontal sweep or scan rate of the electron beam for TV CRT monitor is 15.75 kHz and the vertical sweep frequency (or refresh rate) is 60 Hz. The figures may be arrived at from the following calculation.

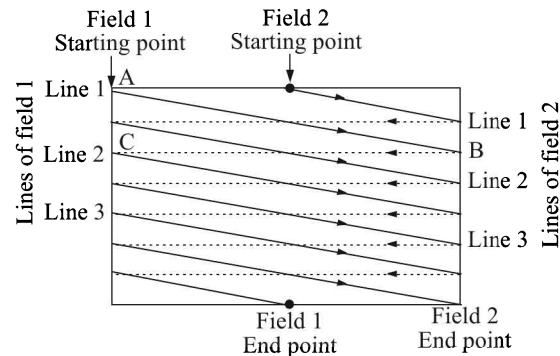


Fig. 17.12 Raster scan of a CRT.

Assume that the beam strikes at A (Fig. 17.12). Then deflecting fields make the beam appear at B. At this moment, a blanking pulse cuts off the video signal and a synchronising pulse causes the beam to move rapidly to point C. The steps are repeated. Note, when the beam reaches C, it skips a line. This is done intentionally and is followed throughout the

³Picture (pix) elements.

scanning. As a result, in one vertical sweep it scans, say, odd lines and in the next, even lines. This is called *interlaced* scanning. The odd or even scanning is done in $(1/60)$ s, so that the full image is scanned in $(1/30)$ s. A CRT contains 525 lines. Therefore, the horizontal scanning rate should be $30 \times 525 = 15750$ lines per second or 15.75 kHz.

The CRT units used for computer VDUs usually use non-interlaced scanning. Here, the number of sweep lines per frame is 260. Therefore, for a vertical sweep rate of 60 Hz, the required horizontal sweep rate is 15.6 kHz.

Vector scan. For certain applications, such as displaying a graph which consists of an array of straight lines, the raster scan is rather a slow and wasteful procedure. Also the diagonal lines drawn on such displays look like stair steps when examined closely.

The vector scan is suitable for such situations. Suppose, a line has to be drawn connecting points X and Y on the screen. In vector scan, this is done by moving the electron beam directly from X to Y . One way of doing it is by connecting one analogue input to the vertical deflection circuitry of the CRT. If the inputs are digital, two DACs may be used.

The vector scan is, therefore, an adaptation of the well-known method of displaying output waveforms in a cathode ray oscilloscope where the analogue input is applied to plates while a sweep voltage of desired frequency is applied to the horizontal deflection plates.

The vector display works well where the information to be displayed consists mostly of straight lines. A circle, for example, may be drawn by joining short line segments. But then, each segment has to be drawn 60 times a second to keep the display refreshed. This puts a limit on using it for displays that has many curves and shaded areas.

Plasma display panel (PDP)

A plasma display panel (PDP) is a type of flat panel display now commonly used for large TV displays (typically above 940 mm or 37 inch). A schematic diagram of the PDP is shown in Fig. 17.13.

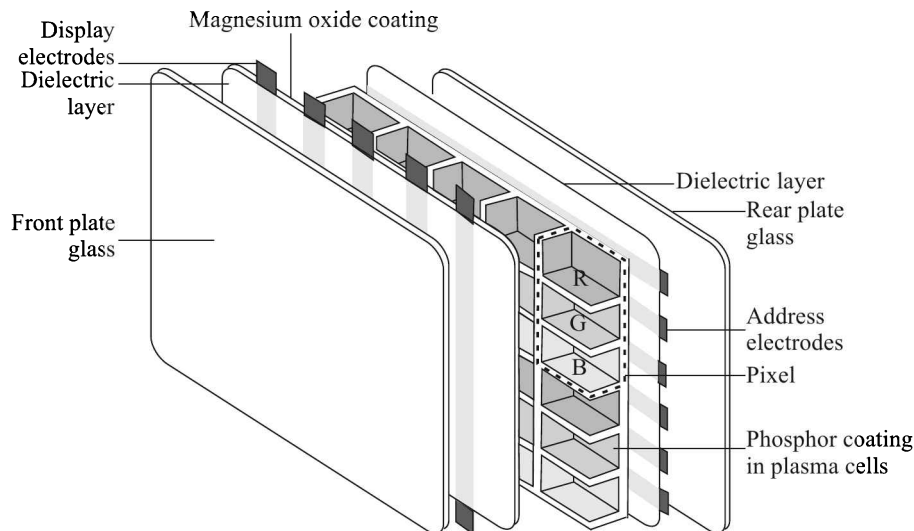


Fig. 17.13 Schematic diagram of a plasma display panel.

Hundreds of thousands tiny cells, located between two plates of glass, hold an inert mixture of noble gases neon and xenon. Long electrodes are sandwiched between the glass plates, in front of and behind the cells. The address electrodes sit behind the cells, along the rear glass plate. The transparent display electrodes, which are surrounded by an insulating dielectric material and covered by a magnesium oxide protective layer, are mounted in front of the cell, along the front glass plate. Control circuitry charges the electrodes that cross paths at a cell, creating a voltage difference between front and back and causing the gas to ionise and form a plasma. As the gas ions rush to the electrodes and collide, photons are emitted.

In colour panels, the back of each cell is coated with a phosphor. The ultraviolet photons emitted by the plasma excite these phosphors to give off coloured light. The operation of each cell is thus comparable to that of a fluorescent lamp. Every pixel is made up of three cells with phosphors for red (R), green (G) and blue (B) colours. These colours blend together to create the overall colour of the pixel.

Plasma displays are bright, have a wide colour range, and can be produced in fairly large sizes, up to 381 cm (150 inches) diagonally. The black level of a PDP is dark compared to the grey of the unilluminated parts of an LCD screen. The display panel is only about 6 cm (~2.5 inches) thick, while the total thickness, including electronics, is less than 10 cm (~4 inches). Plasma displays use as much power per square metre as a CRT or an LCD television. Power consumption varies greatly with picture content, with bright scenes drawing significantly more power than darker ones.

Nominal power rating is typically 220 to 400 watts for a 50-inch (~127 cm) screen.

Liquid crystal display monitor

The liquid crystal display (LCD) is the technology used for displays in notebook and other smaller computers. Like light-emitting diode and plasma display technologies, LCDs allow displays to be much thinner than CRT technology. LCDs consume much less power than LED and plasma displays because they work on the principle of blocking light rather than emitting it. We have already discussed at length in Section 17.3 about their principle of working.

The display grid of an LCD can be either a passive matrix or an active matrix. The passive matrix LCD has a grid of conductors with pixels located at each intersection in the grid. A current is sent across two conductors on the grid to control the light for any pixel.

An *active matrix*, also known as a *thin film transistor* (TFT) display, has a transistor located at each pixel intersection. This requires less current to control the luminance of a pixel. For this reason, the current in an active matrix display can be switched on and off more frequently, improving the screen refresh time. Some passive matrix LCD scan the grid twice with current in the same time that it took for one scan in the original technology. This is called *dual scanning*. However, active matrix is still a superior technology.

A more recent development is the organic thin film transistor (OTFT) technology, which makes it possible to have flexible display surfaces. In the OTFT technology, organic semiconducting compounds are used to construct computer displays. Apart from their low cost, such displays are bright, the colours are vivid, and they are easy to read in most ambient lighting environments. But until recently, they have proven slow in terms of carrier mobility. Slow carrier mobility gives birth to a poor response time. However, the most exciting element of the OTFT technology is that organic substrates allow for displays to be fabricated on flexible surfaces, rather than on rigid materials as is necessary in traditional

TFT displays. A piece of flexible plastic might be coated with OTFT material and made into a display that can be handled like a paper document. It is likely that in the near future such displays with good response time will be available.

Modes of Display

The term display mode refers to the characteristics of a CRT display, in particular the maximum number of colours and the maximum image resolution (in pixels horizontally \times pixels vertically). There are several display modes, or sets of specifications according to which the CRT operates. The most common specification for CRT displays is known as SVGA (super video graphics array). Notebook computers typically use liquid crystal display. The technology for these displays is much different than that for CRTs.

The early display was called the *colour graphics adapter* (CGA). Suitable for monochrome monitors, it was in use in personal computers that were used in word processors and text-based computer systems. This display system was capable of rendering 4 colours, and had a maximum resolution of 320 (H) \times 200 (V) pixels vertically. Next came the *enhanced graphics adapter* (EGA) display. It allowed up to 16 different colours and offered resolutions of up to 640 \times 350 pixels. Though this improved the appearance over earlier displays, and made it possible to read text easily, EGA did not offer sufficient image resolution for full graphic display of images. A few years later, the *video graphics array* (VGA) display system was introduced. In VGA, the maximum resolution depends on the number of colours displayed. One can choose between 16 colours at 640 \times 480 pixels, or 256 colours at 320 \times 200 pixels. Winding its way through XGA and XGA-2, now the standard display has reached at what is called *super video graphics array* (SVGA) display. Typically, an SVGA display can support a palette of up to 16 million colours.

Of late, new specifications—*super extended graphics array* (SXGA) and *ultra extended graphics array* (UXGA)—have come into being. Table 17.3 shows display modes and the resolution levels most commonly associated with each.

Table 17.3 Display modes and their resolutions

| <i>Mode</i> | <i>Resolution</i> (in H \times V pixels) |
|-------------|--|
| VGA | 640 \times 480 |
| SVGA | 800 \times 600 |
| XGA | 1024 \times 768 |
| SXGA | 1280 \times 1024 |
| UXGA | 1600 \times 1200 |

17.4 Recording

In contrast to temporary display, a permanent recording of data is sometimes necessary. All such devices which can record data can broadly be divided into three categories:

1. Chart recorders and plotters
2. Printers
3. Magnetic recorders

Chart Recorders

Chart recorders, as the name implies, record data as graphs on papers and hence the record is analogue. Figure 17.14 explains the principle of operation of such a recorder. The chart

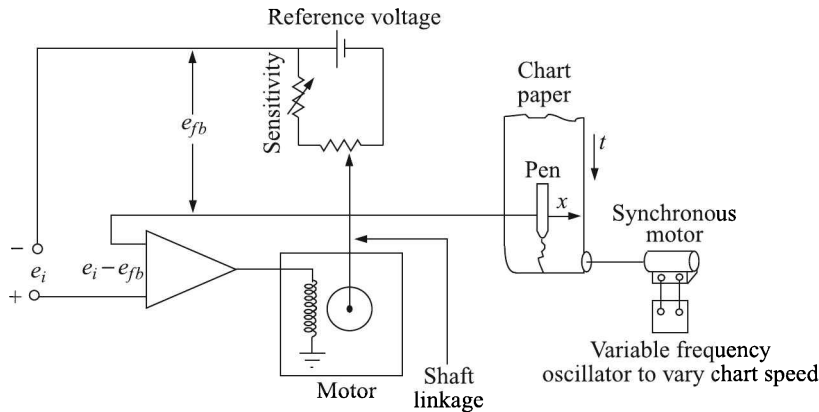


Fig. 17.14 Schematic diagram of a chart recorder.

drive is a synchronous motor actuated by a stable oscillator whose frequency can be varied to alter the chart speed. The input voltage to the motor e_i is compared with a reference voltage e_{fb} and the difference between them is amplified and fed to the field coil of a dc motor. A potentiometric wiper, mechanically connected to the armature of the motor moves over the potentiometer to minimise the error signal ($e_i - e_{fb}$) so that a null is obtained. A pen, mechanically connected to the wiper, records the movement of the wiper on a chart paper.

The record on the chart paper is thus an indication of the magnitude of the error signal which is essentially related to the magnitude of the input signal.

We have already pointed out that a null measurement gives almost the true value of the measurand because the input impedance of the measuring instrument nears infinity at the balance condition.

In *XY*-recorders a cross-plot of two variables is obtained. This is achieved by replacing the moving the chart with a stationary piece of paper and the pen is simultaneously driven in two perpendicular directions by two servomechanisms each having input from one variable.

A plotter is a variant of the *XY*-recorder where digital data can be fed to obtain analogue records. This is done by using DACs at the *X* and *Y* inputs and a stepper motor in lieu of the dc motor.

Simple analogue recorder

A very simple form of analogue recorders is the galvanometric recorder which utilises a d'Arsonval type galvanometer with a mirror arrangement (Fig. 17.15). The input, fed to the galvanometer, produces an angular displacement which is recorded on light-sensitive paper by the lamp and mirror arrangement.

From the analysis point of view, a galvanometric recorder and a potentiometric chart recorder are equivalent because both generate displacement by the motor principle. Hence

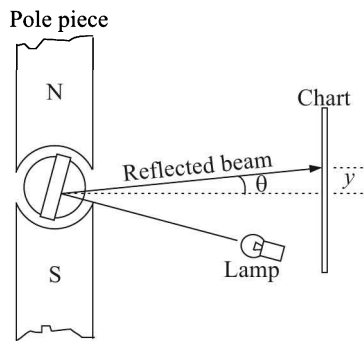


Fig. 17.15 Schematic diagram of a galvanometric recorder (top view).

both are linear second-order systems. We choose the galvanometer and analyse its behaviour to gain an insight into its different controlling factors (Fig. 17.16).

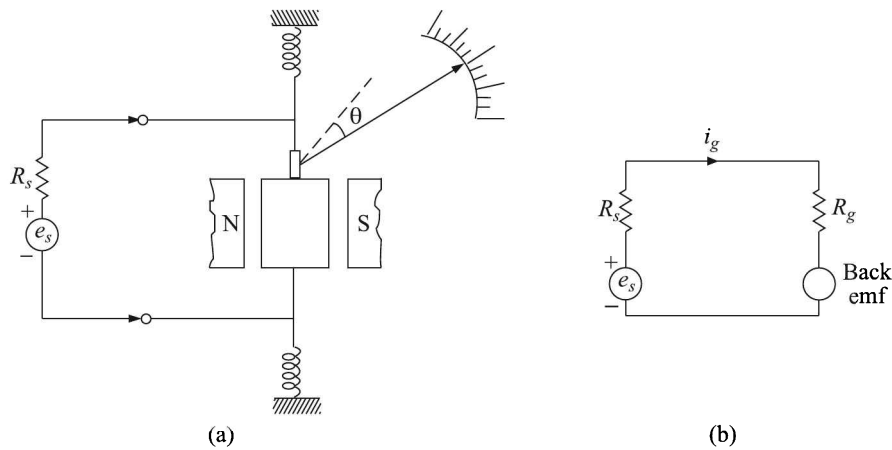


Fig. 17.16 d'Arsonval galvanometer: (a) schematic diagram, and (b) the equivalent circuit.

The differential equation of motion of the recorder can be obtained by applying the Newtonian law to the rotational motion of the coil:

$$\sum \text{Torques} = J \frac{d^2\theta}{dt^2} \tag{17.8}$$

where J is the moment of inertia of the movement and θ is the angle of rotation of the suspension. Two torques act on the movement:

1. The applied current i_g through the conductor, which is placed in a magnetic field, produces an electromagnetic force and causes a torque of magnitude $K_1 i_g$ where K_1 is a constant whose value depends on the magnetic flux density and the cross-sectional area of the coil.
2. The spring of the suspension of the coil produces a force of restitution and the magnitude of the corresponding torque is $K_2 \theta$ where K_2 is the spring constant.

Thus,

$$\sum \text{Torques} = K_1 i_g - K_2 \theta \quad (17.9)$$

To find an expression for the applied current, we note that because of the motion of a conductor in a magnetic field, a back emf will be generated in it. If e_s and R_s are the source emf and the source resistance respectively, the KVL for the electric circuit can be written as

$$i_g(R_g + R_s) + K_1 \frac{d\theta}{dt} - e_s = 0$$

which gives

$$i_g = \frac{e_s - K_1 \frac{d\theta}{dt}}{R_s + R_g} \quad (17.10)$$

From Eqs. (17.8), (17.9) and (17.10) we get

$$K_1 \left[\frac{e_s - K_1 (d\theta/dt)}{R_s + R_g} \right] - K_2 \theta = J \frac{d^2\theta}{dt^2}$$

or

$$\frac{d^2\theta}{dt^2} + \frac{K_1^2}{J(R_s + R_g)} \frac{d\theta}{dt} + \frac{K_2}{J} \theta = \frac{K_1}{J(R_s + R_g)} e_s \quad (17.11)$$

or

$$\frac{d^2\theta}{dt^2} + 2\zeta\omega_0 \frac{d\theta}{dt} + \omega_0^2 \theta = \frac{K_1}{J(R_s + R_g)} e_s$$

where damping factor of the galvanometer, $\zeta = \frac{K_1^2}{2(R_s + R_g)\sqrt{JK_2}}$ (17.12)

its undamped natural frequency $\omega_0 = \sqrt{\frac{K_2}{J}}$ (17.13)

From this analysis, it is clear that a galvanometer or for that matter, a galvanometric recorder is a linear second order instrument and therefore, its response to a step input or to a sinusoidal input is influenced by the damping factor ζ and natural frequency ω_0 .

The following inferences can be drawn from the above analysis:

1. Putting $d^2\theta/dt^2 = d\theta/dt = 0$, we get from Eq. (17.11) that the steady-state sensitivity of the system is given by

$$S \equiv \frac{\theta}{e_s} = \frac{K_1}{K_2} \frac{1}{R_s + R_g} \quad (17.14)$$

K_1 depends on the area of the coil and the field strength of the magnet and K_2 is the spring constant of the suspension system. K_1 and K_2 remaining constant for a system, it may appear that the sensitivity can be increased by lowering the source resistance. The source resistance, in turn, can be brought down by adding a parallel resistance. But then, the Thevenin equivalent of the source voltage goes down, the deflection decreases and therefore, the sensitivity is lowered. On the other hand, if the source resistance is increased, the sensitivity is lowered. Thus, to adjust sensitivity by adjusting the source resistance, we are left with Hobson's choice.

2. Equation (17.12) shows that the damping factor can be altered by adjusting the source resistance. From our analysis of second-order instruments in Chapter 4 we know that for the best frequency response, ζ should have a value around 0.7. Also, such a value of ζ restricts the overshoot for a step input.
3. Equation (17.13) shows that the bandwidth, i.e. undamped natural frequency of the galvanometric recorder can be increased by
 - (a) increasing the spring constant K_2
 - (b) decreasing the moment of inertia J of the suspension system
4. Any increase in the value of the spring constant lowers sensitivity [see Eq. (17.14)] as well as the damping factor [see Eq. (17.12)].
5. The moment of inertia can be brought down by making the coil thin. But a decrease in coil thickness or coil area lowers the value of K_1 which, in turn, lowers S and ζ .

Printers

Printers are quite popular as recording devices. But unlike chart recorders, printers are digital devices. Therefore, a printer can be hooked to an instrumentation system only via a printer controller which will receive analogue input and convert it to suitable digital instruction for the particular printer. Nowadays it is common to interface a PC to the instrumentation system and the received data are processed and presented in a suitable format through a printer which acts as an output device of the PC.

Generally available printers are of the following three kinds, namely

1. Dot-matrix printer
2. Laser printer
3. Ink-jet printer

Dot-matrix printers

Dot-matrix printers print anything by printing a matrix of dots. In impact-type printers pins are arranged in a vertical column in a print-head and each pin can be driven by a solenoid located at the rear of the print-head. Usually 9- or 24-pin printers are available.

Suppose we have to print the letter 'A'. If it is a 9×4 printer, each character is formed by a matrix of 9 rows and 4 columns. So, to print 'A', the head is positioned at the first column and all the required dots for that column are formed by actuating appropriate solenoids. The head then moves on to the next column, and the process is repeated. When all the 4 columns are covered, the character is printed. Since graphics designs can be formed by dots, a dot-matrix printer can print graphics by printing appropriate dots sequentially, column by column.

A variant of dot-matrix printer is the thermal printer where pins are not driven back and forth by solenoids, but are heated by coils. This kind of printer requires a paper which has a special heat-sensitive coating. When a pin is heated, the corresponding spot on the paper is heated and the spot turns dark.

A second form of thermal printers has heating elements arranged on a metal bar which extends up to the width of the paper. Thus, such a printer can print an entire line of dots at a time. The metal bar acts as a heat sink for quick recovery of the pin temperature.

Some of the thermal printers use heat-sensitive ribbons which transfer dots of ink on plain paper.

In comparison to impact printers, thermal ones have less moving parts and are less noisy but they are more expensive as they use sensitised paper or ribbon and slower because pins require time to recover temperature.

Laser printers

In a laser printer a laser beam is turned ON or OFF and swept back and forth across a photo-sensitive drum according to the image to be printed. As a result, the drum develops static charge at certain spots where laser beam strikes while the other spots remain neutral. Next, a blast of powdered ink mixed with resin (the mixture is called *toner*) is applied to the drum when the charged spots attract the ink powder leaving the neutral spots clean. Thus a replica of the image is created on the drum and this image is transferred electrostatically on a paper by applying a reverse field between the paper and the drum. Subsequently, the paper is heated to fuse the toner spots on the paper.

The process is akin to that of xerox copiers except that a xenon flash lamp and optical image-formation are not necessary there.

Figure 17.17 illustrates the mechanism of the laser printer. The figure is self-explanatory. The resolution of laser printer is specified by DPI⁴. The 600 DPI printers are quite common though higher DPI printers are preferred for quality work.

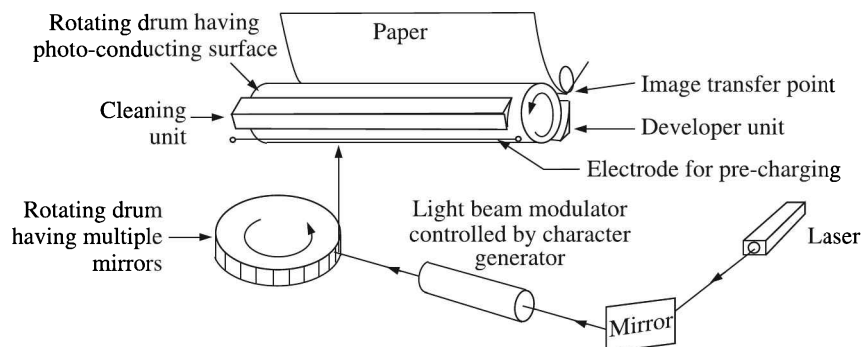


Fig. 17.17 Laser printer mechanism.

Ink-jet printers

Ink-jet printers are now popular because of their lower price and quality of printing. They work on the principle of dot-matrix printers. Some use ink cartridges which contain a column of tiny heaters. A blob of ink explodes onto the paper when one of the heaters is switched on. Another kind uses a special ribbon from which tiny ink bubbles explode onto the paper on application of an electric field on the stylus.

Like laser printers, 600 and 1200 DPI ink-jet printers are common nowadays. An advantage of ink-jet printers is that colour printing can be done without much increase in the cost of the printer.

⁴Dots per inch.

Magnetic Recorders

Magnetic recording has a distinct advantage over other forms because the recorded data or signal can be retrieved at any time to reproduce the original data or signal. The other advantage is, of course, the magnetic record can be erased and the medium can be re-used to record data or signal afresh.

Recording can be made on either a tape or a diskette. Whatever the form, the magnetic substance is thin film of iron oxide (Fe_2O_3) particles deposited on mylar film. A read/write head retrieves data from or records data on the magnetic material. The construction of the read/write heads resembles that of transformer having a toroidal core with a coil. The tape or disc is placed about $10\ \mu\text{m}$ below the head to avoid friction (Fig. 17.18). A current in the coil magnetises the iron oxide film in the immediate vicinity of the head. The pattern of magnetisation of the magnetic film stores the image of the signal.

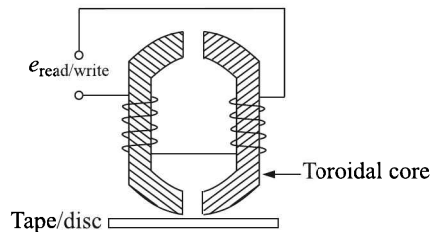


Fig. 17.18 Recording on a magnetic tape by a solenoid head.

The point to note is that when a recorded signal is read, not the original signal but its differentiated form is retrieved. This is because of the fact that the magnetically recorded pattern while passing by the read head generates an induced emf in its coil, and Faraday's law of induction states that the induced emf is

$$|e_{\text{read}}| = \frac{d\Phi}{dt}$$

where Φ is the flux. Therefore, if the recorded signal is given by

$$e_{\text{record}} = A \sin \omega t$$

the reproduced signal will be

$$|e_{\text{read}}| = C\omega \cos \omega t$$

where A and C are constants. It may be noted that e_{read} is

1. 90° out of phase with e_{record} , and
2. proportional to the signal frequency

While the first factor is of little consequence in this case, the second factor implies that the higher the signal frequency the higher will be the reproduced amplitude, thus completely upsetting the fidelity in the reproduction of signal containing a spectrum of frequencies. The problem is overcome by passing the retrieved signal through an amplifier whose amplification varies inversely with the frequency of its input. This process of compensation is known as *equalisation*.

Magnetic recording can be of two forms—analogue or direct recording and digital recording. We discuss the two forms here.

Analogue or direct recording

As the name implies, the direct recording should be feeding the signal as it is to the record-head and reproducing the same by the read head coupled with a compensating amplifier. But, in reality, it is not so simple because of the fact that the current vs. magnetising field curve (i.e. $I-H$ curve) is not linear for a magnetic film. Because of this nonlinearity, a sine-wave input, in which the current varies from negative to positive values passing through zero, the recorded waveform will be completely distorted as shown in Fig. 17.19(a).

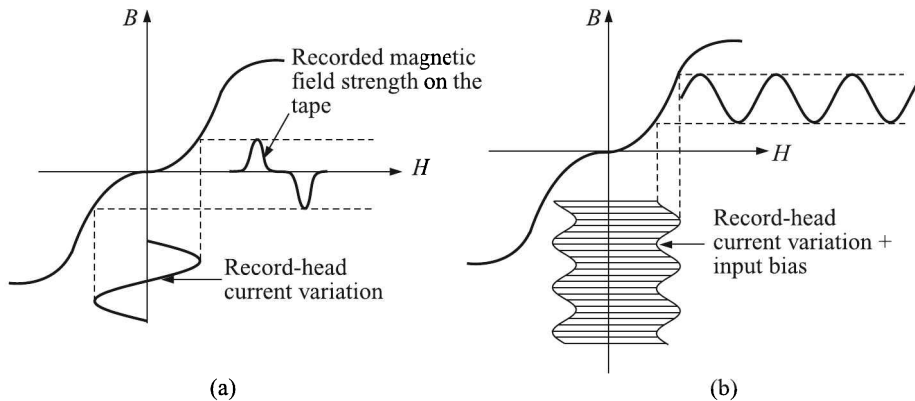


Fig. 17.19 (a) Distortion in magnetic recording, and (b) result of high-frequency bias to the record-head to avoid distortion in magnetic recording.

This distortion can be avoided by mixing a high frequency bias of constant amplitude with the signal to ensure that the variation of the signal current corresponds to the linear region of the $I-H$ characteristic [Fig. 17.19(b)]. Note that this is not amplitude modulation. To facilitate final filtering of the bias, its frequency is chosen to be at least 4-times the maximum signal frequency. And its amplitude should be such that the knee region of the $I-H$ characteristic is avoided.

We have already pointed out that the amplitude of the retrieved signal is proportional to the frequency of the original signal. The frequency is zero for dc, and hence a dc signal cannot be reproduced from direct recording. Equalisation is of no help, because a zero frequency demands an infinite amplification which is absurd.

Nor a very high frequency can be reproduced by this method. Because if the wavelength corresponding to a high frequency signal is smaller than the gap length between the pole pieces of the read head, the coil of the head cannot sense the variation of the signal [Fig. 17.20 (a)].

The high-frequency limit of direct recording is determined by the relation

$$f_{\text{high}} = \frac{v}{l}$$

where v is the tape speed and l is the gap length of the read head. Our previously determined relation that the amplitude of the reproduced waveform is proportional to the frequency of the recording signal is true as long as

$$l = \frac{\lambda}{2}$$

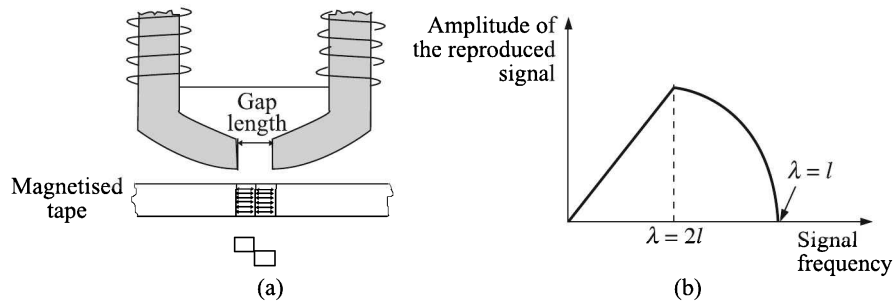


Fig. 17.20 High-frequency limit of direct recording: (a) situation when the gap length equals the signal wavelength, and (b) amplitude vs. signal frequency characteristics.

where λ is the signal wavelength. Beyond that, the amplitude of the reproduced signal starts falling. Ultimately it reduces to zero when the recorded wavelength equals the gap length [see Fig. 17.20(b)].

Digital recording techniques

Digital signals consist of 0's and 1's and although the process of recording a 0 or a 1 on a magnetic surface may appear straightforward, two basic necessities have given rise to a few techniques. The requirements are:

1. The packing density should be as high as possible. This means that, each bit or cell of magnetisation should occupy as small space as possible. Obviously, this will increase the information storage capacity of the tape or disc.
2. The read/write process should be as reliable as possible.

Clearly, the two requirements contradict each other, because the smaller the cell of magnetisation the higher the possibility of its distortion.

To achieve these ends, various digital recording techniques have been devised which can be divided into three basic categories, namely

1. Return-to-zero (RZ) technique
2. Return-to-bias (RB) technique
3. Non-return-to-zero (NRZ) technique

Return-to-zero technique. In this method the recorded tape consists of magnetised cells separated by nonmagnetised cells. For magnetised cells, the directions of magnetisation for recording a 0 and 1 are opposite. The method is illustrated in Fig. 17.21(a). For recording a 1 or a 0, a positive or negative pulse is applied to the write head. In either case, the current through the write head returns to zero after the pulse and remains there until the next pulse arrives. Hence is the name.

The amplified playback (or read) output can be ANDed with the CLK and then fed to a monostable multivibrator (or one-shot) to retrieve the original signal [Fig. 17.21(b)].

In this method, when there is no recording bit, the magnetic field returns to zero flux. This means that the positioning of the read/write head is critical if a record is to be modified. That is why this method is rarely used now.

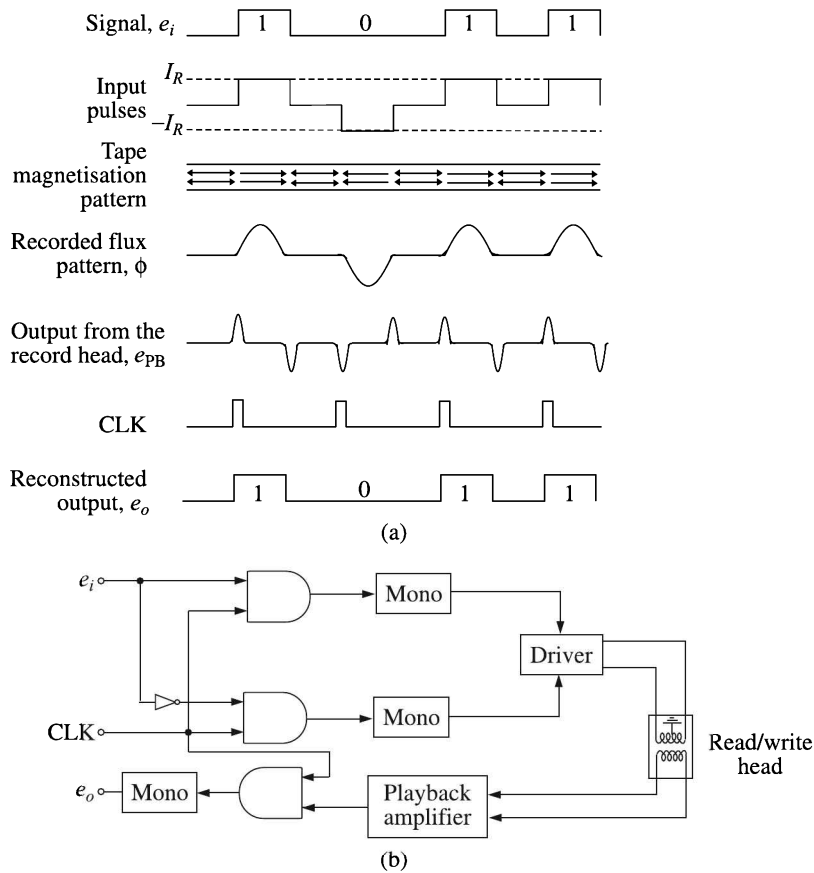


Fig. 17.21 Return-to-zero (RZ) technique. (a) signal at different stages of recording and reproduction, and (b) the circuitry involved.

Return-to-bias technique. This method (Fig. 17.22) is similar to the RZ technique, except that no opposite current is sent through the write head for recording a 0. Hence, the read head only senses 1s and, therefore, the timing is not so critical because except at 1s there is always a negative current flowing through the write head. So, the previous recording will not affect the modified new recording.

However, a primary problem here concerns the sequence of 0's. A parallel recording of CLK signals may solve the problem.

Non-return-to-zero technique. Figure 17.23 illustrates three variants of the NRZ technique. In the simple NRZ method, the current through the write head is negative for 0 bits and positive for 1 bits. In the NRZ-M⁵ method, the current through the write head flips or changes polarity whenever a 1 bit is encountered. Otherwise, the current maintains its previous polarity. In the third method, which is variously called *biphase-mark*, *phase-encoded*, *Harvard*, *Manchester* or *split-frequency* system, a 0 is recorded as negative-to-positive-going pulse and a 1 as a positive-to-negative-going pulse, the change of polarity taking place at the mid-bit time.

⁵M for 'mark'.

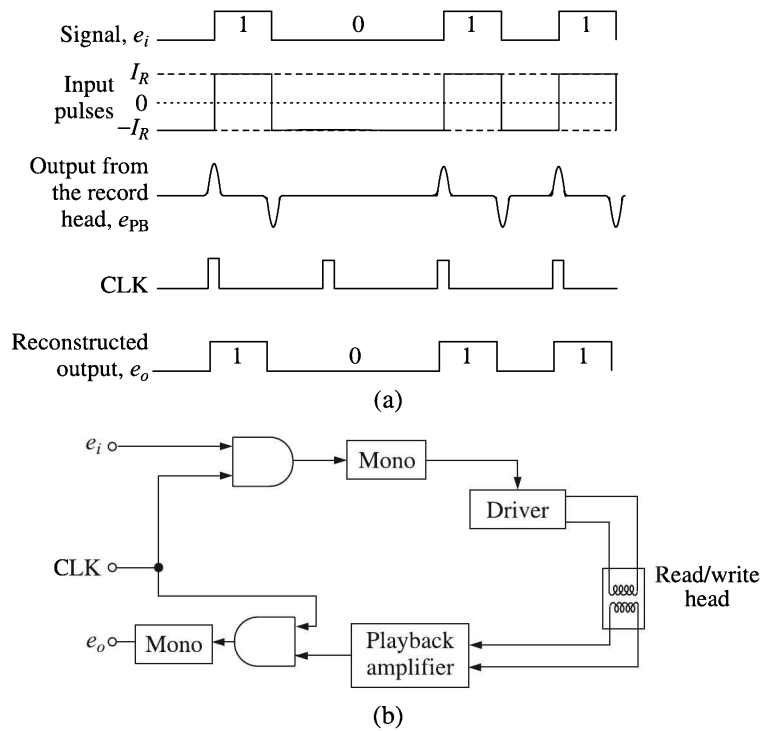


Fig. 17.22 Return-to-bias (RB) technique. (a) signal at different stages of recording and reproduction, and (b) the circuitry involved.

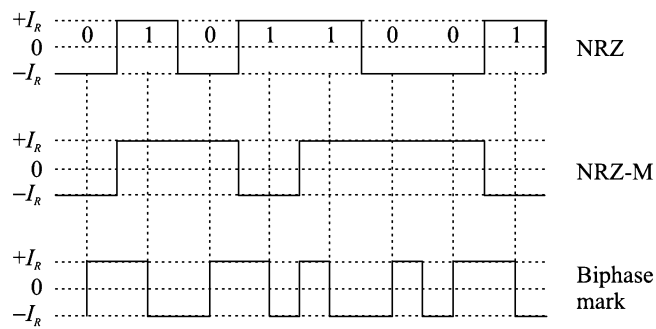


Fig. 17.23 Non-return-to-zero (NRZ) signals of three types.

The NRZ-M technique is schematically shown in Fig. 17.24. Of all the NRZ techniques, the third one is used for high-speed systems.

The comparison made in Table 17.4 will show its superiority.

The other two interesting variants of the NRZ method are:

1. Frequency modulation (FM or F2F), and
2. Modified frequency modulation (MFM)

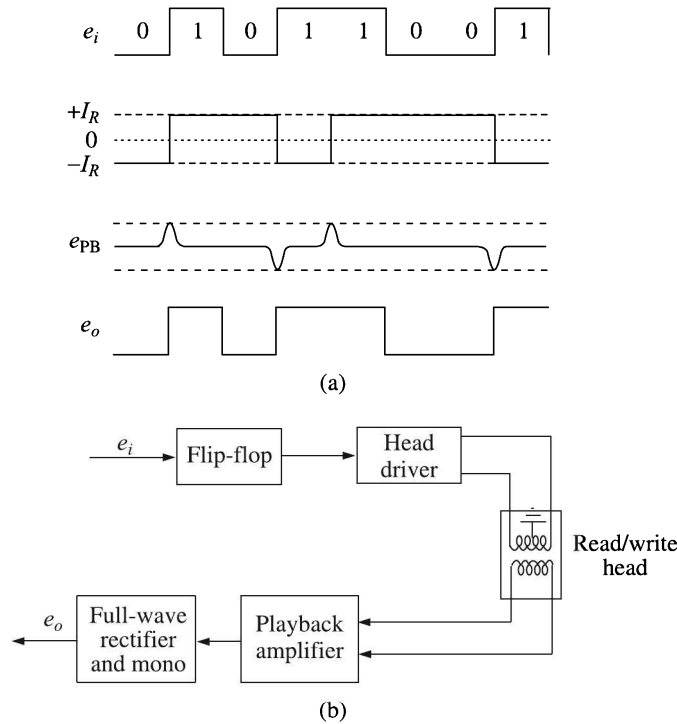


Fig. 17.24 NRZ-M technique: (a) signals, and (b) the circuitry.

Table 17.4 Comparison between NRZ and biphase-mark methods

| Property | NRZ | Biphase-mark |
|----------------------------|-----|--------------|
| Pacing density in bits/cm | 320 | 600 |
| Read/write rate in kbits/s | 120 | 300 |

FM. In FM, every bit cell is marked with a clock pulse. If a data bit is a 1, another pulse is recorded between clock pulses, whereas 0 data bits are represented by no pulse between clock pulses. Putting in extra pulses for 1s modifies the frequency, hence the name frequency modulation (the third row in Fig. 17.25).

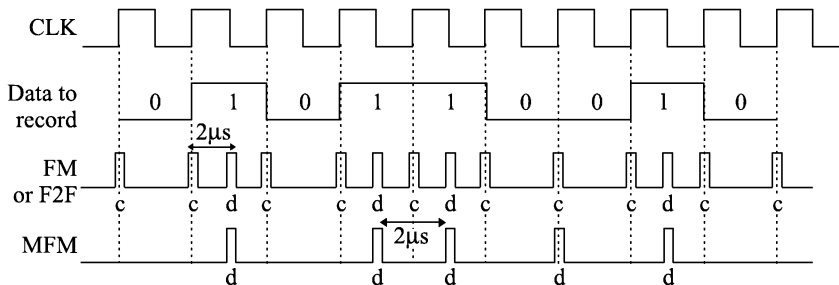


Fig. 17.25 F2F and MFM formats.

MFМ. In MFМ (the last row in Fig. 17.25), normally no clock pulses are recorded. Only pulses for 1 bits are recorded. However, if more than one 0 comes in succession, clock pulses are recorded to differentiate bit cells. As a result, MFМ records double the number of bits that can be recorded within the same length in a track in FM. This is why FM is known as *single density* and MFМ as *double density* recording. For retrieving data, the read head can be synchronised by a PLL with the help of clock pulses and 1 bits. Floppy diskettes in PCs used these types of recording.

Magnetic recording, its formats and read/write principles constitute a subject by itself. We presented here only a bird's eye view. A comprehensive discussion of these topics is beyond the scope of this book.

17.5 Data Acquisition Systems

Traditionally, measurements are displayed on stand-alone instruments of various types—oscilloscopes, PMMC meters, counters, etc. However, the need to record the measured data and process them for visualisation has become rather easy and elegant with the advent of the digital computer. The result is the data acquisition system.

Abbreviated as DAS (or DAQ), data acquisition system, in practice, processes signals obtained from measurement systems and generates some desired information. DAS components, therefore, convert electrical measurement signals generated by appropriate sensors in such a way that they can be fed to a computer port. Acquired data is displayed, analysed, and stored on the computer, either using vendor supplied software, or custom software developed using various programming languages such as C, Java, Lisp, etc.

Vendor supplied software can be grouped under one of the following categories:

1. Software that provides instant results for dedicated functions such as datalogging and display.
2. Icon-based point-and-click software for interactively developing more advanced custom test applications.
3. Comprehensive programming environments that provide flexibility for creating complex algorithms and custom operator interfaces.

Let us elaborate the differences between the three types of software.

Instant-result DAS

An instant-result DAS software can provide nearly effortless *dedicated functions* such as basic signal verification, real-time data display and datalogging. One such package is DaqView⁶, which allows one to perform the said functions within moments of setting up the supplied hardware.

But since they offer very limited manoeuvrability, this variety is best suited for some routine measurements that do not demand customisation.

⁶Product of ioTech, see www.iotech.com

Icon-based point-and-click software

Interactive icon-based data acquisition software does not require one to write or understand programmes. Instead, it offers a point-and-click GUI so that one can build on-screen block diagrams via a series of functional icons. DASyLab of ioTech, for example, is one such software.

Most interactive icon-based data acquisition software also has extensive configuration, importing, analysis, graphing, control and reporting capabilities, all of which can be configured via intuitive dialogue boxes. A simple application takes minutes to complete.

Graphical programming environment

Some applications demand a unique data acquisition and control capability which has to be custom-built. Once the application is created, it may be necessary to design a turnkey application suitable for use by a nontechnical operator. Graphical programming environment is the right kind of software for this purpose. It helps one to create demanding system requirements, such as algorithms and sophisticated graphical interfaces.

A few of such software are presented in Table 17.5:

Table 17.5 Graphical programming environment generating software

| <i>Software</i> | <i>Availability</i> | <i>What it does</i> |
|-----------------|--|--|
| LabVIEW | Product of National Instruments, USA. (www.ni.com) | A commercial product, it offers a graphical programming environment optimised for data acquisition as well as virtual instrumentation. |
| MATLAB | Product of The Mathworks, USA (www.mathworks.com) | Another commercial product, it is a programming language having built-in graphical tools and libraries for data acquisition and analysis. |
| EPICS | Acronym of Experimental Physics and Industrial Control System. Developed by Argonne National Laboratory, USA (www.aps.anl.gov/epics/) | This is a set of open source software. It offers tools, libraries and applications that can be used to build large scale data acquisition and control systems for scientific instruments such as particle accelerators, telescopes and similar large scientific experiments. |

Supervisory control and data acquisition

EPICS, in fact, is a step forward for data acquisition on a large scale and using the same for the control of large scientific instrument installations. Much in the same way, the process in an industry is controlled by a category of data acquisition system which is called Supervisory Control And Data Acquisition system or SCADA.

SCADA systems consist of software and hardware components. The hardware collects data and feeds them to a computer that has the SCADA software installed. The computer processes the data and issues necessary signals to appropriate hardware such that the process goes on satisfying pre-scheduled parameters. Over and above, the computer records and logs all events

into a file and sounds alarms in case the conditions become hazardous. SCADA is used in power plants, refineries, telecommunication systems, transportation and so on.

We will, however, confine ourselves to a discussion on DASs only.

How Data are Acquired

Data exchange between an instrument and the computer can take place in several ways:

1. Some instruments are equipped with a serial port that allows exchange of data between a computer or another instrument.
2. Others may have a GPIB⁷ board which allows transfer of data in a parallel format along with an identity tag for the instrument in a network of instruments.
3. The instrument may not be equipped with any port, but its output may be fed to a DAS hardware which is used for data acquisition.

The DAS hardware could be in the form of modules that can be connected to the computer's ports (parallel, serial, USB, etc...) or cards connected to slots in the mother board. DAS-cards often contain multiple components (multiplexer, ADC, DAC, TTL-IO, high speed timers, RAM). The block diagram of a typical DAS card is shown in Fig. 17.26.

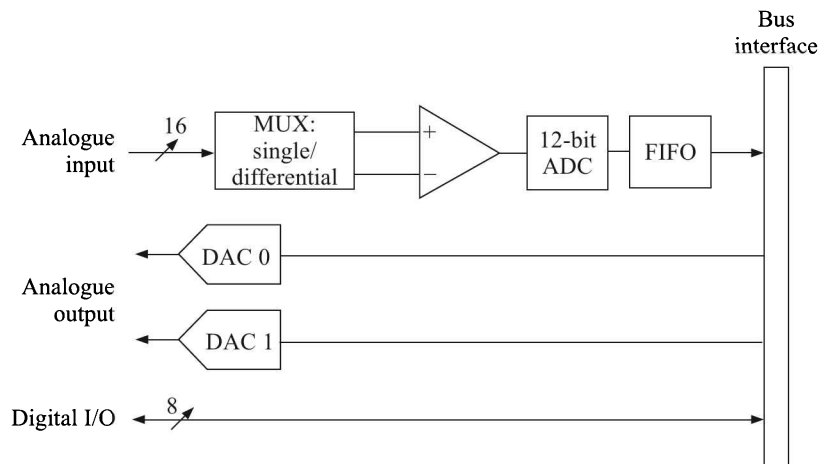


Fig. 17.26 Block diagram of a typical DAS card. 16 and 8 indicate number of inputs or I/O.

It has 16 analogue channels which can either be configured as 16 single-ended inputs, or 8 differential inputs. This is accomplished by the multiplexer, or switching circuit and is software configurable.

The output of the multiplexer feeds into an amplifier whose gain is programmable through software. This circuit allows the programmer to select an amplification, say from 0.5 to 100, appropriate to the signal that is to be measured. For example, consider a bipolar (both positive and negative) input signal. The ADC has an input voltage range of ± 5 V. Hence a gain of 0.5 would enable the board to handle voltages ranging between ± 10 V ($= 5/0.5$). Similarly, a gain of 100 would result in a maximum range of ± 50 mV ($= 5/100$) at the input to the board.

⁷General Purpose Instrumentation Bus.

In addition to the ADC, there are 2 DACs which allow one to generate analogue signals. There are 8 general purpose digital I/O lines which allow the board to control external digital circuitry or monitor the state of external devices such as switches or buttons.

Low level communication with the data acquisition board is handled through drivers provided by vendors. These drivers allow the programmer to perform all the necessary tasks such as initialising, configuring, and sending/receiving data from the board.

Characteristics of DASs

Resolution

The resolution of the conversion of analogue input signal into digital format depends upon the number of bits the ADC uses. The higher the resolution, the higher the number of divisions the voltage range is broken into, and therefore, the smaller the detectable voltage change. An 8-bit ADC gives $2^8 = 256$ levels while a 12-bit ADC will give $2^{12} = 4096$ levels. Hence, the 12-bit ADC will be able to detect smaller increments of the input signals than its 8-bit counterpart. The LSB is the minimum increment of the voltage that an ADC can convert. Hence, the LSB varies with the operating input voltage range of the ADC. We have discussed this at length in Section 17.2. Thus, if one needs to detect smaller changes, one has to use a higher resolution ADC. Clearly, the resolution is an important characteristic of the DAS board.

Nonlinearity

Ideally, if the voltage applied to the input of an ADC is increased linearly, we would expect the digital codes to increment linearly as shown in Fig. 17.27(a).

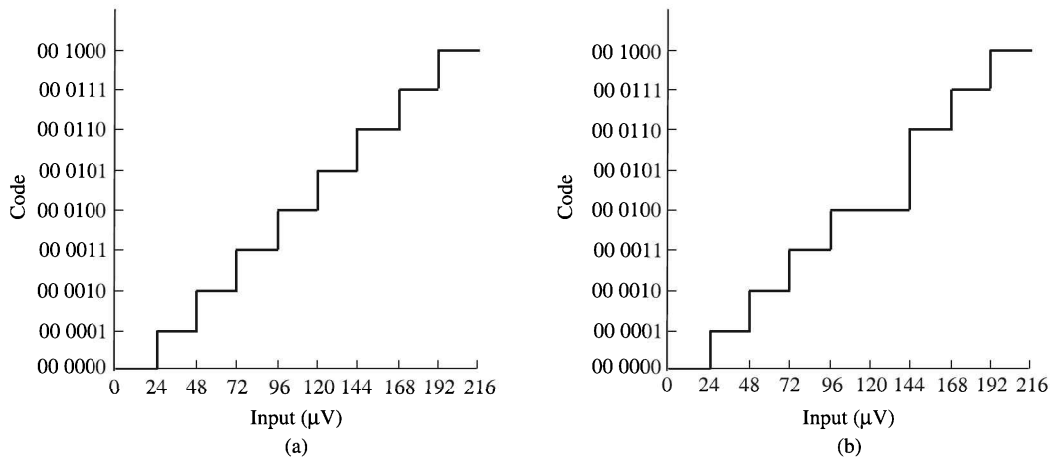


Fig. 17.27 Transfer characteristic of (a) an ideal ADC, and (b) an ADC showing differential nonlinearity.

While a perfect DAS board will have no nonlinearity, most of the commercially available boards display some nonlinearity which is specified as differential nonlinearity. Figure 17.27(b) shows the result of nonlinearity.

Settling time

On a DAS board, a multiplexer first selects the analogue signal and then it is amplified before its conversion by the ADC. The amplifier located between the multiplexer and the ADC must be able to track the output of the multiplexer. Or else, the ADC will continue to convert the signal that is still being transferred from the previous channel and add the value to the current channel value. Therefore, the settling time, which changes with the sampling rate and gain of the DAS board, is a major characteristic.

Data transfer speed

DAS boards are installed in a PC with high speed data bus. Depending on the speed of the motherboard of the PC, the data transfer speed between the microprocessor and memory is 20–40 MHz [Fig. 17.28(a)]. It can be enhanced by allowing the DAS board to communicate with the RAM directly [Fig. 17.28(b)].

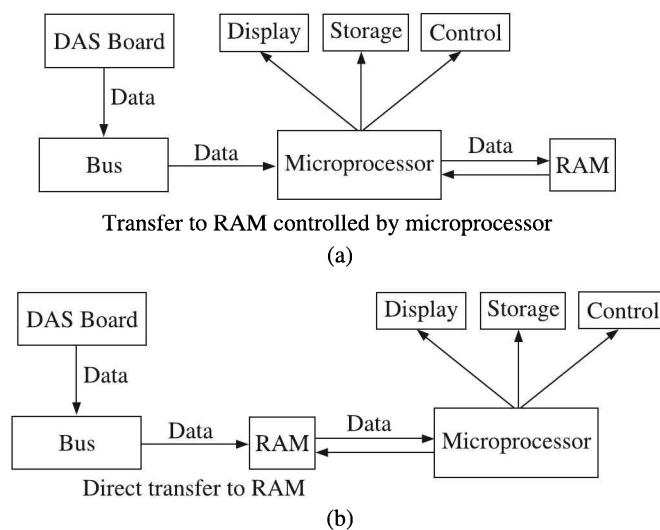


Fig. 17.28 Data transfer from a DAS board: (a) normal, and (b) bus mastering.

The technique of direct communication with the RAM is known as *bus mastering*. Expensive DAS boards incorporate this facility.

17.6 Virtual Instrumentation

Using high performance DAS cards, fast computers, and data processing software like LabVIEW, one can achieve performance similar to expensive bench top instruments. A whole new instrumentation paradigm, known as virtual instrumentation, has evolved over the years by which one can control an output, process input signals and log data.

What It Is

Till recently, instrumentation systems generally consisted of individual instruments. For example, a temperature measurement system comprised the following sequence (Fig. 17.29):

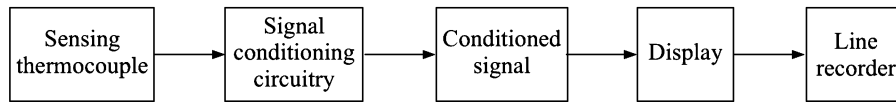


Fig. 17.29 Traditional instrumentation for temperature measurement system.

The recorder inked a trace of changing conditions to create a time record of temperature variations. Even complex systems, such as chemical process control applications, typically used sets of individual physical instruments wired to a central control panel that comprised an array of physical data display devices such as dials and counters, together with sets of switches, knobs and buttons for controlling the instruments.

The introduction of computers into the field of instrumentation began as a way to couple an individual instrument, such as a temperature sensor, to a computer, and enable the display of measurement data on a simulated instrument panel, displayed on the computer monitor and containing buttons or other means for controlling the operation of the sensor. Thus, such instrumentation enabled the creation of a simulated physical instrument, having the capability of controlling physical sensing components.

A variety of data collection instruments designed specifically for computerised control and operation through software were developed and made commercially available, ushering the field of virtual instrumentation⁸. Virtual instrumentation thus refers to the use of general purpose computers in combination with data collection hardware devices and virtual instrumentation software, to construct an integrated instrumentation system.

The functions of the system may be:

- To control external measurement hardware devices, which incorporate sensing elements for detecting changes in the conditions of test subjects, from the computer,
- To display test or measurement data collected by the external device on a computer screen simulating in appearance of the physical dials, meters and other data visualisation devices of traditional instruments, and
- To control processes based on data collected and processed by a computerised instrumentation system.

There is another kind of virtual instrumentation systems that comprise pure *software instruments*, such as oscilloscopes or spectrum analysers, for collecting the sensor data and presenting them in user-friendly formats. The VisualSCOPE⁹ or VirtualBench¹⁰ help create such vi systems.

Problems to Tackle

Many difficulties need to be tackled to develop a virtual instrumentation system. Some of them are:

- Existence of many types of electronic interfaces by which external data collection devices could be coupled to a computer.
- The variety in the *command sets* used by different hardware device vendors to control their respective products.

⁸Often abbreviated as *vi*.

⁹Product of Scientific Software Tools Inc., USA see www.drverlinx.com.

¹⁰Product of National Instruments, USA see www.ni.com.

- Difference in the internal structures and functions of data collecting hardware devices, requiring virtual instrumentation systems to take the differences into account. For example, some DASs are so-called ‘register-based’ instruments, controlled by streams of 0s and 1s sent directly to control components within the instrument, while others are ‘message-based’ instruments controlled by strings of ASCII characters effectively constructing written instructions that must be decoded within the instrument.
- Different instruments use different protocols—for example, some as electrical frequencies and others as variations in a base voltage—to output data.
- Requirements may vary for different applications. For example, one process may need only to collect a single value—e.g. outside temperature—once an hour, and to have all collected values stored in a file. Another process possibly requires that several related process temperatures be monitored continuously, and that a shut-off valve be activated in the event of relative temperature variation between two process steps, say, by more than five degrees for a time period of five seconds or more, during a 10 minute period, and to store only data concerning such events.

Any virtual instrumentation system intended for integration to a variety of commercially available data collection hardware devices must address these problems so that it contains software tools that make it capable of communicating effectively with the disparate types of hardware devices.

Until 1990s, only professional programmers would write codes using languages such as BASIC, PASCAL, C or C++ for the virtual instrumentation systems. Apparently, development of vi’s using such programming languages is not only tedious and time-consuming but also needs considerable code writing skill. Such a code is virtually unreadable by non-programmers and thus cannot be modified to suit their specific needs.

During the last decade, a few commercial software products have been made available with the help of which vi systems can be developed providing a graphical development environment within which a custom vi system can be built. Typically, the user is presented with a design desktop environment, generally having the look-and-feel familiar to users of Windows-based graphical applications, in which a variety of software options and tools are accessible from toolbars and dialogue boxes featuring drop-down menus, and may be accessed by manipulating an on-screen cursor using a computer mouse.

The diagram of such a vi can look rather complex but it is not very difficult to learn. Consider, for example, the LabVIEW software which is widely used for developing vi systems. There is no syntax to be learnt in LabVIEW¹¹.

It provides the user with tools for designing the so-called data flow diagrams. The user is required to place icons representing desired system components onto a design desktop and then to effect wiring connections between components. Of course, to design a data flow diagram for a measurement system, one is required to have a good understanding of the specific data paths and element combinations that will be required to attain one’s objective. Also, one has to have the knowledge of components that are functionally compatible. Recent developments of relevant software offer adequate help to ensure mutual compatibility of components.

¹¹Read “My first vi” in *LabVIEW—Graphical Programming for Instruments, User Manual*, National Instruments (1996) to understand how a simple vi can be developed in LabVIEW.

LabVIEW also requires one to develop a GUI that interacts with the data flow diagram. It features a large selection of procedural icon libraries. Each procedural icon within a library represents a conventional programming construct (for example, FOR or WHILE) or a conditional statement (for example, IF) similar to those found in languages such as C.

The appearance of the GUI of a typical virtual instrument for temperature measurement and control, generated by LabVIEW, is presented for illustration in Fig. 17.30.

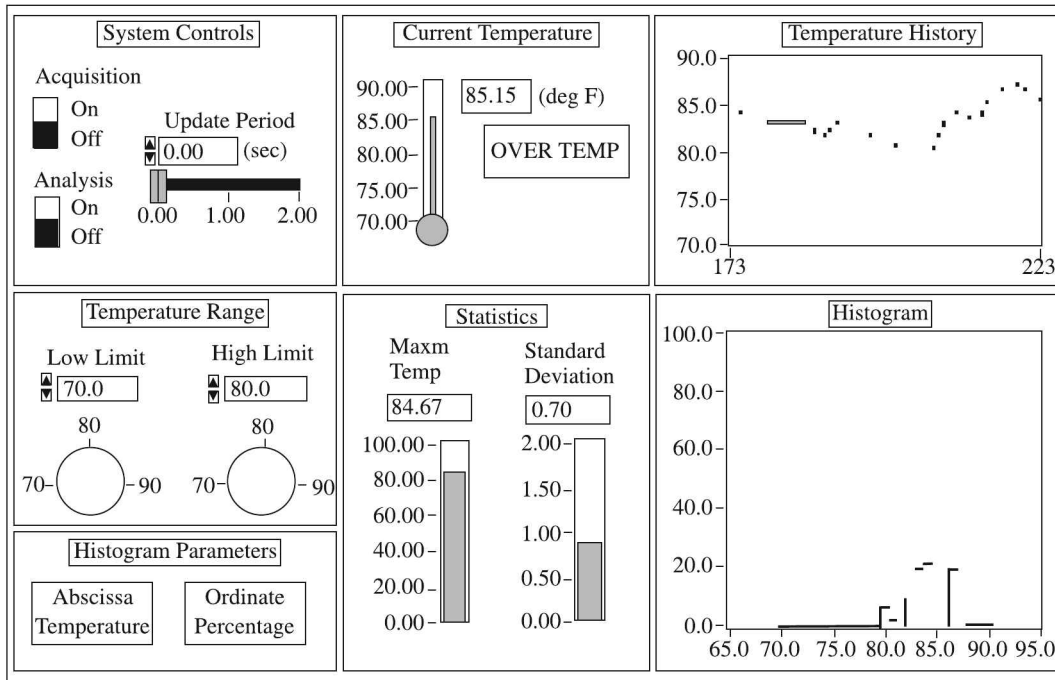


Fig. 17.30 Representative illustration of a temperature measurement and control virtual instrument generated by LabVIEW.

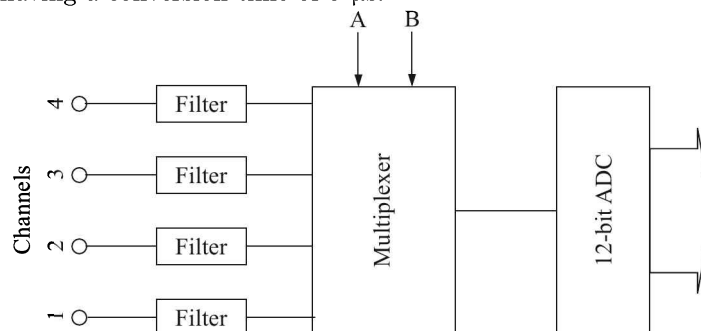
In this example, a temperature monitoring system consists of acquiring data from a temperature sensor and plotting them on a graph. The vi also calculates the mean and standard deviation of the data and plots them. Alarms have been incorporated so that if the temperature falls below or above the set point, the alarm goes ON.

Review Questions

- 17.1 Explain the working of an X-Y recorder with the help of a diagram. What are the important features of this instrument?
- 17.2 What are the various means of permanent recording? Briefly discuss the principles of working of a magnetic type recording system.

With neat sketches describe galvanometric, null-type and potentiometric strip-chart recorders.

- 17.3 Compare LED and LCD in respect of (i) construction material, (ii) construction principle, (iii) power consumption, (iv) speed, (v) voltage and current, (vi) cost, and (vii) size of the display area.
- 17.4 Define the sensitivity of a DVM. The lowest range of a $4\frac{1}{2}$ digit DVM is 10 mV full-scale. What is the sensitivity of this meter?
- 17.5 A 3-digit DVM has an accuracy specification of $\pm 0.5\%$ of reading ± 2 digits. (a) What is the possible error, in volts, when the instrument is reading 5.00 V on its 10 V range? (b) What is the possible error, in volts, when reading 0.10 V on the 10 V range? (c) What percentage of the reading is the possible error in the case of (b)?
- 17.6 Explain with the help of schematic diagrams the magnetic recording and reproduction of a sine-wave signal by the method of Pulse Duration Modulation. Elucidate the NRZ (Non-return-to-zero) method of magnetic recording and reproduction of the digital signal 1011001.
- 17.7 Name the various modulation techniques used in data recording. What will be the data format to record 1011010010 in full-binary unipolar RZ form?
- 17.8 A $3\frac{1}{2}$ -digit multimeter has an accuracy of $\pm 0.5\%$ reading ± 5 digits. If the meter reads 2 mA on a full-scale of 20 mA, what is the worst-case error in the reading?
- 17.9 Indicate the correct choice:
- An amplitude modulated signal can be observed on a CRO by applying the following waveform to the 'external trigger' input:
 - the modulated waveform itself
 - derivative of the modulated waveform
 - the modulating waveform
 - the carrier waveform
 - The post-deflection acceleration in a CRT is used in high frequency oscilloscopes mainly to
 - increase the deflection sensitivity
 - improve the beam focussing
 - improve the deflection linearity
 - improve the spot brightness
- 17.10 A data acquisition system (DAS) shown below employs a successive approximation type 12-bit ADC having a conversion time of 5 μ s.



-
- (a) The quantisation error of the ADC is
- (i) 0% (ii) $\pm 0.012\%$ (iii) $\pm 0.024\%$ (iv) $\pm 0.048\%$
- (b) The system is used as a single channel DAS with channel 1 selected as input to the ADC which is set in the continuous conversion mode. For avoiding aliasing error, the cutoff frequency f_c of the filter in channel 1 should be
- (i) $f_c < 100$ kHz
(ii) $f_c = 100$ kHz
(iii) $100 \text{ kHz} < f_c < 200$ kHz
(iv) $f_c = 200$ kHz
- (c) If the multiplexer is controlled such that the channels are sequenced every $5 \mu\text{s}$ as 1, 2, 1, 3, 1, 4, 1, 2, 1, 3, 1, 4, 1, ..., the input connected to channel 1 will be sampled at the rate of
- (i) 25k samples/s (ii) 50k samples/s
(iii) 100k samples/s (iv) 200k samples/s

Variance of Combinations

Suppose, the variable X is dependent upon the experimental variables p, q, r, \dots , where each variable varies in a random and independent way, i.e.

$$X = f(p, q, r, \dots)$$

Then,

$$\delta X = \frac{\partial X}{\partial p} \delta p + \frac{\partial X}{\partial q} \delta q + \frac{\partial X}{\partial r} \delta r + \dots \quad (\text{A.1})$$

$$\begin{aligned} (\delta X)^2 &= \left(\frac{\partial X}{\partial p} \right)^2 (\delta p)^2 + \left(\frac{\partial X}{\partial q} \right)^2 (\delta q)^2 + \left(\frac{\partial X}{\partial r} \right)^2 (\delta r)^2 + \dots \\ &+ 2 \left(\frac{\partial X}{\partial p} \right) \left(\frac{\partial X}{\partial q} \right) \delta p \delta q + 2 \left(\frac{\partial X}{\partial q} \right) \left(\frac{\partial X}{\partial r} \right) \delta q \delta r + \dots \end{aligned} \quad (\text{A.2})$$

In Eq. (A.2), the cross terms, i.e. terms with a coefficient of 2, cancel, because $\delta p, \delta q, \delta r \dots$ are error terms which means for each $+\delta p$ there is a $-\delta p$ and so on. Thus,

$$(\delta X)^2 = \left(\frac{\partial X}{\partial p} \right)^2 (\delta p)^2 + \left(\frac{\partial X}{\partial q} \right)^2 (\delta q)^2 + \left(\frac{\partial X}{\partial r} \right)^2 (\delta r)^2 + \dots \quad (\text{A.3})$$

If we denote variance of X as $V_X = (\delta X)^2$, and those of $p, q, r \dots$ as $V_p = (\delta p)^2, V_q = (\delta q)^2, V_r = (\delta r)^2 \dots$, then Eq. (A.3) may be re-written as

$$V_X = \left(\frac{\partial X}{\partial p} \right)^2 V_p + \left(\frac{\partial X}{\partial q} \right)^2 V_q + \left(\frac{\partial X}{\partial r} \right)^2 V_r + \dots$$

Since the standard deviation is square root of variance,

$$\begin{aligned} \sigma_X &= \sqrt{\left(\frac{\partial X}{\partial p} \right)^2 V_p + \left(\frac{\partial X}{\partial q} \right)^2 V_q + \left(\frac{\partial X}{\partial r} \right)^2 V_r + \dots} \\ &= \sqrt{\left(\frac{\partial X}{\partial p} \right)^2 \sigma_p^2 + \left(\frac{\partial X}{\partial q} \right)^2 \sigma_q^2 + \left(\frac{\partial X}{\partial r} \right)^2 \sigma_r^2 + \dots} \end{aligned}$$

One may further define weighted standard deviations as

$$\sigma_{pX} = \frac{\partial X}{\partial p} \sigma_p, \quad \sigma_{qX} = \frac{\partial X}{\partial q} \sigma_q, \quad \sigma_{rX} = \frac{\partial X}{\partial r} \sigma_r \dots$$

whence

$$\sigma_X = \sqrt{\sigma_{pX}^2 + \sigma_{qX}^2 + \sigma_{rX}^2 + \dots} \quad (\text{A.4})$$

Linear Time-invariant Systems

B.1 Linear Systems

A system is called *linear* if it possesses the following two properties:

- (a) Additivity
- (b) Homogeneity

The two properties are elaborated below.

Additivity property

If x and y belong to the domain of the function f , additivity demands

$$f(x + y) = f(x) + f(y)$$

It may appear, this definition looks more mathematical than real-life control problems. But, we have to remember, all control systems should be amenable to mathematical modelling, else they cannot be tackled systematically. So, this definition, though mathematical, is applicable to real-life control situations.

Let us now examine a few functions and see if they satisfy the additivity criterion.

Example B.1

Let a function be defined by

$$u(y) = \frac{d^2 y(t)}{dt^2} + 8 \frac{dy(t)}{dt} + y(t) \tag{i}$$

Does it satisfy the additivity property?

Solution

Let us check what happens to the function if we substitute $x + y$ for y in Eq. (i).

$$\begin{aligned} u(x + y) &= \frac{d^2}{dt^2}[x(t) + y(t)] + 8 \frac{d}{dt}[x(t) + y(t)] + [x(t) + y(t)] \\ &= \left[\frac{d^2 x(t)}{dt^2} + 8 \frac{dx(t)}{dt} + x(t) \right] + \left[\frac{d^2 y(t)}{dt^2} + 8 \frac{dy(t)}{dt} + y(t) \right] \\ &= u(x) + u(y) \end{aligned}$$

Therefore, $u(y)$ is an additive function.

Example B.2

Let a function be defined by

$$f(x) = x^2 \quad (i)$$

Does it satisfy additivity criterion?

Solution

As before, we substitute $x + y$ for x in Eq. (i). Then,

$$f(x + y) = (x + y)^2 \neq x^2 + y^2$$

Thus, $f(x)$ is not an additive function.

Example B.3

Let

$$u(y) = y \frac{d^2 y}{dt^2} + 7 \frac{dy}{dt} \quad (i)$$

Is it an additive function?

Solution

On substituting $x + y$ for y in Eq. (i), we get

$$u(x + y) = (x + y) \frac{d^2}{dt^2} (x + y) + 7 \frac{d}{dt} (x + y) \neq \left[x \frac{d^2 x}{dt^2} + 7 \frac{dx}{dt} \right] + \left[y \frac{d^2 y}{dt^2} + 7 \frac{dy}{dt} \right]$$

Thus, $u(y)$ is not an additive function.

Homogeneity property

The homogeneity property demands that for any y belonging to the domain of the function f , if α is a scalar constant, then

$$f(\alpha y) = \alpha f(y)$$

Let us now examine whether the functions, that we tested for additivity, satisfy the homogeneity criterion.

Example B.4

Is the function defined by

$$u(y) = \frac{d^2 y(t)}{dt^2} + 8 \frac{dy(t)}{dt} + y(t) \quad (i)$$

homogeneous?

Solution

We substitute αy for y in Eq. (i) and observe

$$\frac{d^2(\alpha y)}{dt^2} + 8 \frac{d(\alpha y)}{dt} + \alpha y = \alpha \left[\frac{d^2 y}{dt^2} + 8 \frac{dy}{dt} + y \right]$$

Therefore, $u(y)$ defined by Eq. (i) is a homogeneous function.

Example B.5

Does the function defined by

$$f(x) = x^2 \quad (i)$$

satisfy the homogeneity test?

Solution

We see the result of substitution of x by αx in Eq. (i) that

$$f(\alpha x) = (\alpha x)^2 \neq \alpha x^2$$

Therefore, $f(x)$ defined in Eq. (i) does not stand the homogeneity test.

Example B.6

What about the homogeneity of the function defined in the following equation?

$$u(y) = y \frac{d^2 y}{dt^2} + 7 \frac{dy}{dt} \quad (i)$$

Solution

The same test of substitution of y by αy yields

$$(\alpha y) \frac{d^2(\alpha y)}{dt^2} + 7 \frac{d(\alpha y)}{dt} = \alpha^2 y \frac{d^2 y}{dt^2} + 7\alpha \frac{dy}{dt} \neq \alpha \left[y \frac{d^2 y}{dt^2} + 7 \frac{dy}{dt} \right]$$

Hence, the function $u(y)$ defined in Eq. (i) is not homogeneous.

Considering all the three functions cited above, we arrive at the conclusion that only the one given in Example B.4 is linear because it satisfies both additivity and homogeneity properties while the functions given in Examples B.5 and B.6 do not and hence are nonlinear.

Time-varying and Time-invariant Systems

If the parameters of a control system vary with time, it is a *time-varying* system. If they do not vary with time, the system is *time-invariant*.

A simple rocket is a time-varying system because while it is moving in its trajectory, one parameter of its mathematical modelling—mass—is continuously changing with time because of burning of its fuel. However, a simple pendulum of mass m , executing a simple harmonic motion about its mean position of rest, is a time-invariant system, and therefore, its trajectory can be described by a simple equation

$$m \frac{d^2 x}{dt^2} = -\alpha x$$

where its parameters— m and α —are independent of time.

An LR circuit, within its current tolerance, is a time-invariant system. It is described by

$$e = L \frac{di}{dt} + Ri$$

where L and R are constants. But if the current is high, L and R become $L(t)$ and $R(t)$ because they get hot with time and their values change thereby. Obviously then, the LR circuit is not time-invariant.

Laplace Transform

Let $f(t)$ be a function of t specified for $t > 0$. Then, Laplace transform of $f(t)$, denoted by $\mathcal{L}\{f(t)\}$, is defined as

$$\mathcal{L}\{f(t)\} \equiv F(s) = \int_0^{\infty} e^{-st} f(t) dt \quad (\text{C.1})$$

where s may be real or complex. It is more useful to consider s complex as $s = \sigma + j\omega$. Table C.1 lists Laplace transforms of a few elementary functions.

Table C.1 Laplace transforms of a few elementary functions

| $f(t)$ | $\mathcal{L}\{f(t)\} \equiv F(s)$ | $f(t)$ | $\mathcal{L}\{f(t)\} \equiv F(s)$ |
|--------------------------------|-----------------------------------|------------|-------------------------------------|
| 1 | $\frac{1}{s} \quad s > 0$ | $\sin at$ | $\frac{a}{s^2 + a^2} \quad s > 0$ |
| t | $\frac{1}{s^2} \quad s > 0$ | $\cos at$ | $\frac{s}{s^2 + a^2} \quad s > 0$ |
| $t^n \quad n = 0, 1, 2, \dots$ | $\frac{n!}{s^{n+1}} \quad s > 0$ | $\sinh at$ | $\frac{a}{s^2 - a^2} \quad s > a $ |
| $\exp(at)$ | $\frac{1}{s - a} \quad s > a$ | $\cosh at$ | $\frac{s}{s^2 - a^2} \quad s > a $ |

C.1 Some Important Properties of Laplace Transforms

Laplace Transform of a Constant

$$\mathcal{L}\{k\} = \frac{k}{s} \quad (\text{C.2})$$

Linearity Property

If c_1 and c_2 are constants and $\mathcal{L}\{f_1(t)\} \equiv F_1(s)$, $\mathcal{L}\{f_2(t)\} \equiv F_2(s)$, then

$$\mathcal{L}\{c_1 f_1(t) + c_2 f_2(t)\} = c_1 \mathcal{L}\{f_1(t)\} + c_2 \mathcal{L}\{f_2(t)\} = c_1 F_1(s) + c_2 F_2(s)$$

Because of this property, \mathcal{L} is called a linear operator.

Example C.1

$$\begin{aligned}\mathcal{L}\{3t^2 - 2\cos 2t + 4\exp(-t)\} &= 3\mathcal{L}\{t^2\} - 2\mathcal{L}\{\cos 2t\} + 4\mathcal{L}\{\exp(-t)\} \\ &= 3\left(\frac{2!}{s^3}\right) - 2\left(\frac{s}{s^2 + 4}\right) + 4\left(\frac{1}{s + 1}\right) = \frac{6}{s^3} - \frac{2s}{s^2 + 4} + \frac{4}{s + 1}\end{aligned}$$

First Translation or Shifting Property

If $\mathcal{L}\{f(t)\} \equiv F(s)$ then

$$\mathcal{L}\{\exp(at)f(t)\} = F(s - a)$$

Example C.2

Since $\mathcal{L}\{\cos 2t\} = \frac{s}{s^2 + 4}$,

$$\mathcal{L}\{\exp(-t)\cos 2t\} = \frac{s + 1}{(s + 1)^2 + 4} = \frac{s + 1}{s^2 + 2s + 5}$$

Second Translation or Shifting Property

If $\mathcal{L}\{f(t)\} \equiv F(s)$ and $g(t) = \begin{cases} f(t - a) & t > a \\ 0 & t < a \end{cases}$

then

$$\mathcal{L}\{g(t)\} = \exp(-as)F(s)$$

Example C.3

Since $\mathcal{L}\{t^3\} = \frac{3!}{s^4} = \frac{6}{s^4}$,

$$\mathcal{L}\{g(t)\} = \frac{6\exp(-2s)}{s^4} \quad \text{when} \quad g(t) = \begin{cases} (t - 2)^3 & t > 2 \\ 0 & t < 2 \end{cases}$$

Change of Scale Property

If $\mathcal{L}\{f(t)\} \equiv F(s)$, then

$$\mathcal{L}\{f(at)\} = \frac{1}{a}F\left(\frac{s}{a}\right)$$

Example C.4

Since $\mathcal{L}\{\sin t\} = \frac{1}{s^2 + 1}$,

$$\mathcal{L}\{\sin 3t\} = \frac{1}{3} \cdot \frac{1}{(s/3)^2 + 1} = \frac{3}{s^2 + 9}$$

Laplace Transform of Derivatives

(a) If $\mathcal{L}\{f(t)\} \equiv F(s)$, then

$$\mathcal{L}\{f'(t)\} = sF(s) - f(0)$$

provided, $f(t)$ is continuous in $0 \leq t \leq N$ and exponential order for $t > N$, while $f'(t)$ is sectionally continuous for $0 \leq t \leq N$.

(b) If $\mathcal{L}\{f(t)\} \equiv F(s)$, then

$$\mathcal{L}\{f''(t)\} = s^2F(s) - sf(0) - f'(0)$$

(c) If $\mathcal{L}\{f(t)\} \equiv F(s)$, then

$$\mathcal{L}\{f^{(n)}(t)\} = s^n F(s) - s^{n-1}f(0) - s^{n-2}f'(0) - \dots - sf^{(n-2)}(0) - f^{(n-1)}(0)$$

Example C.5

If $f(t) = \cos 3t$, $\mathcal{L}\{f(t)\} \equiv F(s) = \frac{s}{s^2 + 9}$

Therefore,

$$\mathcal{L}\{f'(t)\} = \mathcal{L}\{-3 \sin 3t\} = \frac{s^2}{s^2 + 9} - 1 = -\frac{9}{s^2 + 9}$$

Laplace Transform of Integrals

If $\mathcal{L}\{f(t)\} \equiv F(s)$, then

$$\mathcal{L}\left\{\int_0^t f(u)du\right\} = \frac{F(s)}{s}$$

Example C.6

Since $\mathcal{L}\{\sin 2t\} = \frac{2}{s^2 + 4}$,

$$\mathcal{L}\left\{\int_0^t \sin 2u du\right\} = \frac{2/(s^2 + 4)}{s} = \frac{2}{s(s^2 + 4)}$$

Laplace Transform of Dirac δ -Function

The Dirac δ -function is defined as

$$\delta(t) = \begin{cases} \frac{1}{\varepsilon} & 0 \leq t \leq \varepsilon \\ 0 & t > \varepsilon \end{cases}$$

where $\varepsilon \rightarrow 0$. Thus, the Laplace transform of $\delta(t)$ is

$$\begin{aligned}\mathcal{L}\{\delta(t)\} &= \lim_{\varepsilon \rightarrow 0} \int_0^{\infty} \exp(-st)\delta(t)dt \\ &= \lim_{\varepsilon \rightarrow 0} \left\{ \int_0^{\varepsilon} \frac{\exp(-st)}{\varepsilon} + \int_{\varepsilon}^{\infty} \exp(-st)(0)dt \right\} \\ &= \lim_{\varepsilon \rightarrow 0} \left\{ \frac{1}{\varepsilon} \left[\frac{\exp(-st)}{-s} \right]_0^{\varepsilon} \right\} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1 - \exp(-s\varepsilon)}{s\varepsilon}\end{aligned}$$

Now,

$$\begin{aligned}\lim_{\varepsilon \rightarrow 0} \frac{1 - \exp(-s\varepsilon)}{s\varepsilon} &= \lim_{\varepsilon \rightarrow 0} \frac{1 - (1 - s\varepsilon + s^2\varepsilon^2/2! - \dots)}{s\varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0} \left(1 - \frac{s\varepsilon}{2!} + \dots + \text{terms containing powers of } s\varepsilon \right) \\ &= 1\end{aligned}$$

Hence,

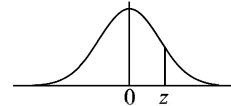
$$\mathcal{L}\{\delta(t)\} = 1$$

APPENDIX D

Statistical Tables

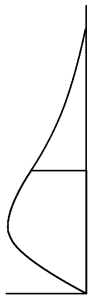
Table D.1 Areas for a standard normal distribution

An entry in the table is the area under the curve, between $z = 0$ and $+z$ as shown in the adjacent figure. Areas for $-z$ are obtained from symmetry.



| z | Second decimal place of z | | | | | | | | | |
|-----|-----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
| 0.0 | .0000 | .0040 | .0080 | .0120 | .0160 | .0199 | .0239 | .0279 | .0319 | .0359 |
| 0.1 | .0398 | .0438 | .0478 | .0517 | .0557 | .0596 | .0636 | .0675 | .0714 | .0754 |
| 0.2 | .0793 | .0832 | .0871 | .0910 | .0948 | .0987 | .1026 | .1064 | .1103 | .1141 |
| 0.3 | .1179 | .1217 | .1255 | .1293 | .1331 | .1368 | .1406 | .1443 | .1480 | .1517 |
| 0.4 | .1554 | .1591 | .1628 | .1664 | .1700 | .1736 | .1772 | .1808 | .1844 | .1879 |
| 0.5 | .1915 | .1950 | .1985 | .2019 | .2054 | .2088 | .2123 | .2157 | .2190 | .2224 |
| 0.6 | .2258 | .2291 | .2324 | .2357 | .2389 | .2422 | .2454 | .2486 | .2518 | .2549 |
| 0.7 | .2580 | .2612 | .2642 | .2673 | .2704 | .2734 | .2764 | .2794 | .2823 | .2852 |
| 0.8 | .2881 | .2910 | .2939 | .2967 | .2996 | .3023 | .3051 | .3078 | .3106 | .3133 |
| 0.9 | .3159 | .3186 | .3212 | .3238 | .3264 | .3289 | .3315 | .3340 | .3365 | .3389 |
| 1.0 | .3413 | .3438 | .3461 | .3485 | .3508 | .3531 | .3554 | .3577 | .3599 | .3621 |
| 1.1 | .3643 | .3665 | .3686 | .3708 | .3729 | .3749 | .3770 | .3790 | .3810 | .3830 |
| 1.2 | .3849 | .3869 | .3888 | .3907 | .3925 | .3944 | .3962 | .3980 | .3997 | .4015 |
| 1.3 | .4032 | .4049 | .4066 | .4082 | .4099 | .4115 | .4131 | .4147 | .4162 | .4177 |
| 1.4 | .4192 | .4207 | .4222 | .4236 | .4251 | .4265 | .4279 | .4292 | .4306 | .4319 |
| 1.5 | .4332 | .4345 | .4357 | .4370 | .4382 | .4394 | .4406 | .4418 | .4429 | .4441 |
| 1.6 | .4452 | .4463 | .4474 | .4484 | .4495 | .4505 | .4515 | .4525 | .4535 | .4545 |
| 1.7 | .4554 | .4564 | .4573 | .4582 | .4591 | .4599 | .4608 | .4616 | .4625 | .4633 |
| 1.8 | .4641 | .4649 | .4656 | .4664 | .4671 | .4678 | .4686 | .4693 | .4699 | .4706 |
| 1.9 | .4713 | .4719 | .4726 | .4732 | .4738 | .4744 | .4750 | .4756 | .4761 | .4767 |
| 2.0 | .4772 | .4778 | .4783 | .4788 | .4793 | .4798 | .4803 | .4808 | .4812 | .4817 |
| 2.1 | .4821 | .4826 | .4830 | .4834 | .4838 | .4842 | .4846 | .4850 | .4854 | .4857 |
| 2.2 | .4861 | .4864 | .4868 | .4871 | .4875 | .4878 | .4881 | .4884 | .4887 | .4890 |
| 2.3 | .4893 | .4896 | .4898 | .4901 | .4904 | .4906 | .4909 | .4911 | .4913 | .4916 |
| 2.4 | .4918 | .4920 | .4922 | .4925 | .4927 | .4929 | .4931 | .4932 | .4934 | .4936 |
| 2.5 | .4938 | .4940 | .4941 | .4943 | .4945 | .4946 | .4948 | .4949 | .4951 | .4952 |
| 2.6 | .4953 | .4955 | .4956 | .4957 | .4959 | .4960 | .4961 | .4962 | .4963 | .4964 |
| 2.7 | .4965 | .4966 | .4967 | .4968 | .4969 | .4970 | .4971 | .4972 | .4973 | .4974 |
| 2.8 | .4974 | .4975 | .4976 | .4977 | .4977 | .4978 | .4979 | .4979 | .4980 | .4981 |
| 2.9 | .4981 | .4982 | .4982 | .4983 | .4984 | .4984 | .4985 | .4985 | .4986 | .4986 |
| 3.0 | .4987 | .4987 | .4987 | .4988 | .4988 | .4989 | .4989 | .4989 | .4990 | .4990 |
| 3.1 | .4990 | .4991 | .4991 | .4991 | .4992 | .4992 | .4992 | .4992 | .4993 | .4993 |
| 3.2 | .4993 | .4993 | .4994 | .4994 | .4994 | .4994 | .4994 | .4995 | .4995 | .4995 |
| 3.3 | .4995 | .4995 | .4995 | .4996 | .4996 | .4996 | .4996 | .4996 | .4996 | .4997 |
| 3.4 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4998 |

Table D.2 Percentile values (χ^2_p) for the chi-square distribution



The shaded area in the adjacent figure = p and ν is the degree of freedom.

| ν | $\chi^2_{.995}$ | $\chi^2_{.99}$ | $\chi^2_{.975}$ | $\chi^2_{.95}$ | $\chi^2_{.90}$ | $\chi^2_{.75}$ | $\chi^2_{.50}$ | $\chi^2_{.25}$ | $\chi^2_{.10}$ | $\chi^2_{.05}$ | $\chi^2_{.025}$ | $\chi^2_{.01}$ | $\chi^2_{.005}$ |
|-------|-----------------|----------------|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|----------------|-----------------|
| 1 | 7.88 | 6.63 | 5.02 | 3.84 | 2.71 | 1.32 | .455 | .102 | .0158 | .0039 | .0010 | .0002 | .0000 |
| 2 | 10.6 | 9.21 | 7.38 | 5.99 | 4.61 | 2.77 | 1.39 | .575 | .211 | .103 | .0506 | .0201 | .0100 |
| 3 | 12.8 | 11.3 | 9.35 | 7.81 | 6.25 | 4.11 | 2.37 | 1.21 | .584 | .352 | .216 | .115 | .072 |
| 4 | 14.9 | 13.3 | 11.1 | 9.49 | 7.78 | 5.39 | 3.36 | 1.92 | 1.06 | .711 | .484 | .297 | .207 |
| 5 | 16.7 | 15.1 | 12.8 | 11.1 | 9.24 | 6.63 | 4.35 | 2.67 | 1.61 | 1.15 | .831 | .554 | .412 |
| 6 | 18.5 | 16.8 | 14.4 | 12.6 | 10.6 | 7.84 | 5.35 | 3.45 | 2.20 | 1.64 | 1.24 | .872 | .676 |
| 7 | 20.3 | 18.5 | 16.0 | 14.1 | 12.0 | 9.04 | 6.35 | 4.25 | 2.83 | 2.17 | 1.69 | 1.24 | .989 |
| 8 | 22.0 | 20.1 | 17.5 | 15.5 | 13.4 | 10.2 | 7.34 | 5.07 | 3.49 | 2.73 | 2.18 | 1.65 | 1.34 |
| 9 | 23.6 | 21.7 | 19.0 | 16.9 | 14.7 | 11.4 | 8.34 | 5.90 | 4.17 | 3.33 | 2.70 | 2.09 | 1.73 |
| 10 | 25.2 | 23.2 | 20.5 | 18.3 | 16.0 | 12.5 | 9.34 | 6.74 | 4.87 | 3.94 | 3.25 | 2.56 | 2.16 |
| 11 | 26.8 | 24.7 | 21.9 | 19.7 | 17.3 | 13.7 | 10.3 | 7.58 | 5.58 | 4.57 | 3.82 | 3.05 | 2.60 |
| 12 | 28.3 | 26.2 | 23.3 | 21.0 | 18.5 | 14.8 | 11.3 | 8.44 | 6.30 | 5.23 | 4.40 | 3.57 | 3.07 |
| 13 | 29.8 | 27.7 | 24.7 | 22.4 | 19.8 | 16.0 | 12.3 | 9.30 | 7.04 | 5.89 | 5.01 | 4.11 | 3.57 |
| 14 | 31.3 | 29.1 | 26.1 | 23.7 | 21.1 | 17.1 | 13.3 | 10.2 | 7.79 | 6.57 | 5.63 | 4.66 | 4.07 |
| 15 | 32.8 | 30.6 | 27.5 | 25.0 | 22.3 | 18.2 | 14.3 | 11.0 | 8.55 | 7.26 | 6.26 | 5.23 | 4.60 |
| 16 | 34.3 | 32.0 | 28.8 | 26.3 | 23.5 | 19.4 | 15.3 | 11.9 | 9.31 | 7.96 | 6.91 | 5.81 | 5.14 |
| 18 | 37.2 | 34.8 | 31.5 | 28.9 | 26.0 | 21.6 | 17.3 | 13.7 | 10.9 | 9.39 | 8.23 | 7.01 | 6.26 |
| 20 | 40.0 | 37.6 | 34.2 | 31.4 | 28.4 | 23.8 | 19.3 | 15.5 | 12.4 | 10.9 | 9.59 | 8.26 | 7.43 |
| 24 | 45.6 | 43.0 | 39.4 | 36.4 | 33.2 | 28.2 | 23.3 | 19.0 | 15.7 | 13.8 | 12.4 | 10.9 | 9.89 |
| 30 | 53.7 | 50.9 | 47.0 | 43.8 | 40.3 | 34.8 | 29.3 | 24.5 | 20.6 | 18.5 | 16.8 | 15.0 | 13.8 |
| 40 | 66.8 | 63.7 | 59.3 | 55.8 | 51.8 | 45.6 | 39.3 | 33.7 | 29.1 | 26.5 | 24.4 | 22.2 | 20.7 |
| 60 | 92.0 | 88.4 | 83.3 | 79.1 | 74.4 | 67.0 | 59.3 | 52.3 | 46.5 | 43.2 | 40.5 | 37.5 | 35.5 |
| 80 | 116.3 | 112.3 | 106.6 | 101.9 | 96.6 | 88.1 | 79.3 | 71.1 | 64.3 | 60.4 | 57.2 | 53.5 | 51.2 |
| 100 | 140.2 | 135.8 | 129.6 | 124.3 | 118.5 | 109.1 | 99.3 | 90.1 | 82.4 | 77.9 | 74.2 | 70.1 | 67.3 |

Psychrometric Table

t_d and t_w indicate dry-bulb and wet-bulb temperatures respectively. It is computed for a pressure of 74.27 cm Hg. Errors resulting from the use of this table for air temperatures above 10 °C and between 77.5 and 71 cm Hg will usually be within the errors of observation.

| $^{\circ}\text{C} (t_d - t_w)$ | |
|--------------------------------|---|
| t_d | t_w |
| 16 | 95 90 85 81 76 71 67 63 58 54 50 46 42 38 34 30 26 23 19 15 12 8 5 |
| 17 | 95 90 86 81 76 72 68 64 60 55 51 47 43 40 36 32 28 25 21 18 14 11 8 12.0 |
| 18 | 95 91 86 82 77 73 69 65 61 57 53 49 45 41 38 34 30 27 23 20 17 14 10 7 12.5 |
| 19 | 95 91 87 82 78 74 70 65 62 58 54 50 46 43 39 36 32 29 26 22 19 16 13 10 7 13.0 |
| 20 | 96 91 87 83 78 74 70 66 63 59 55 51 48 44 41 37 34 31 28 24 21 18 15 12 9 6 13.5 |
| 21 | 96 91 87 83 79 75 71 67 64 60 56 53 49 46 42 39 36 32 29 26 23 20 17 14 12 9 6 14.0 |
| 22 | 96 92 87 83 80 76 72 68 64 61 57 54 50 47 44 40 37 34 31 28 25 22 19 17 14 11 8 6 14.5 |
| 23 | 96 92 88 84 80 76 72 69 65 62 58 55 52 48 45 42 39 36 33 30 27 24 21 19 16 13 11 8 6 15.0 |
| 24 | 96 92 88 84 80 77 73 69 66 62 59 56 53 49 46 43 40 37 34 31 29 26 23 20 18 15 13 10 8 5 |
| 25 | 96 92 88 84 81 77 74 70 67 63 60 57 54 50 47 44 41 39 36 33 30 28 25 22 20 17 15 12 10 8 16.0 |
| 26 | 96 92 88 85 81 78 74 71 67 64 61 58 54 51 49 46 43 40 37 34 32 29 26 24 21 19 17 14 12 10 5 |
| 27 | 96 92 89 85 82 78 75 71 68 65 62 58 56 52 50 47 44 41 38 36 33 31 28 26 23 21 18 16 14 12 7 17.0 |
| 28 | 96 93 89 85 82 78 75 72 69 65 62 59 56 53 51 48 45 42 40 37 34 32 29 27 25 22 20 18 16 13 9 5 |
| 29 | 96 93 89 86 82 79 76 72 69 66 63 60 57 54 52 49 46 43 41 38 36 33 31 28 26 24 22 19 17 15 11 7 18.0 |
| 30 | 96 93 89 86 83 79 76 73 70 67 64 61 58 55 52 50 47 44 42 39 37 35 32 30 28 25 23 21 19 17 13 9 5 |
| 31 | 96 93 90 86 83 80 77 73 70 67 64 61 59 56 53 51 48 45 43 40 38 36 33 31 29 27 25 22 20 18 4 11 7 19.0 |
| 32 | 96 93 90 86 83 80 77 74 71 68 65 62 60 57 54 51 49 46 44 41 39 37 35 32 30 28 26 24 22 20 16 12 9 5 |
| 33 | 97 93 90 87 83 80 77 74 71 68 66 63 60 57 55 52 50 47 45 42 40 38 36 33 31 29 27 25 23 21 17 14 10 7 20.0 |
| 34 | 97 93 90 87 84 81 78 75 72 69 66 63 61 58 56 53 51 48 46 43 41 39 37 35 32 30 28 26 24 23 19 15 12 8 5 |
| 35 | 97 94 90 87 84 81 78 75 72 69 67 64 61 59 56 54 51 49 47 44 42 40 38 36 34 32 30 28 26 24 20 17 13 10 7 |
| 36 | 97 94 90 87 84 81 78 75 73 70 67 64 62 59 57 54 52 50 48 45 43 41 39 37 35 33 31 29 27 25 21 18 15 11 8 |
| 37 | 97 94 91 87 84 82 79 76 73 70 68 65 63 60 58 55 53 51 48 46 44 42 40 38 36 34 32 30 28 26 23 19 16 13 10 |
| 38 | 97 94 91 88 84 82 79 76 74 71 68 66 63 61 58 56 54 51 49 47 45 43 41 39 37 35 33 31 29 27 24 20 17 14 11 |
| 39 | 97 94 91 88 85 82 79 77 74 71 69 66 64 61 59 57 54 52 50 48 46 43 42 39 38 36 34 32 30 28 25 22 18 15 12 |
| 40 | 97 94 91 88 85 82 80 77 74 72 69 67 64 62 59 57 54 53 51 48 46 44 42 40 38 36 35 33 31 29 26 23 20 16 14 |

Condensed from Bulletin of the U.S. Weather Bureau No. 1071 (National Weather Service Forecast Office in Sacramento, California
www.wrth.noaa.gov/Sacramento/hmi/rhmbi.html)

APPENDIX F

Miscellaneous Data

Table F.1 SI base units

| <i>Base quantity</i> | <i>Name</i> | <i>Symbol</i> | <i>CGS value</i> |
|---------------------------|-------------|---------------|--------------------------------------|
| Amount of substance | mole | mol | |
| Electric current | ampere | A | 3×10^9 statamp [†] |
| Length | metre | m | 10^2 cm |
| Luminous intensity | candela | cd | |
| Mass | kilogram | kg | 10^3 g |
| Thermodynamic temperature | kelvin | K | |
| Time | second | s | 1 s |

[†] Assuming velocity of light $c = 3 \times 10^8$ m/s, which is an approximate value.

Table F.2 Examples of derived units

| <i>Derived quantity</i> | <i>Name</i> | <i>Symbol</i> |
|-----------------------------------|----------------------------|------------------------|
| Acceleration | metre per second squared | m/s^2 |
| Amount-of-substance concentration | mole per cubic metre | mol/m^3 |
| Area | square metre | m^2 |
| Current density | ampere per square metre | A/m^2 |
| Luminance | candela per square metre | cd/m^2 |
| Magnetic field strength | ampere per metre | A/m |
| Mass density | kilogramme per cubic metre | kg/m^3 |
| Specific volume | cubic metre per kilogramme | m^3/kg |
| Speed, Velocity | metre per second | m/s |
| Volume | cubic metre | m^3 |
| Wave number | reciprocal metre | m^{-1} |

Table F.3 SI derived units with special names and symbols

| <i>Derived quantity</i> | <i>Name</i> | <i>Symbol</i> | <i>Expression in terms of other SI units</i> | <i>Expression in terms of SI base units</i> | <i>CGS value</i> |
|--|----------------|---------------|--|---|--|
| Absorbed dose, Specific energy (imparted), Kerma | gray | Gy | J/kg | $\text{m}^2 \cdot \text{s}^{-2}$ | |
| Activity (of a radionuclide) | becquerel | Bq | – | s^{-1} | |
| Capacitance | farad | F | C/V | $\text{m}^{-2} \cdot \text{kg}^{-1} \cdot \text{s}^4 \cdot \text{A}^2$ | 9×10^{11} cm |
| Celsius temperature | degree celsius | °C | – | K | |
| Dose equivalent | sievert | Sv | J/kg | $\text{m}^2 \cdot \text{s}^{-2}$ | |
| Electric charge, Quantity of electricity | coulomb | C | – | $\text{s} \cdot \text{A}$ | 3×10^9 statcoul |
| Electric conductance | siemens | S | A/V | $\text{m}^{-2} \cdot \text{kg}^{-1} \cdot \text{s}^3 \cdot \text{A}^2$ | |
| Electric potential difference, Electromotive force | volt | V | W/A | $\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-3} \cdot \text{A}^{-1}$ | $\frac{1}{300}$ statvolt |
| Electric resistance | ohm | Ω | V/A | $\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-3} \cdot \text{A}^{-2}$ | $\frac{1}{9} \times 10^{-11}$ s/cm |
| Energy, Work, Quantity of heat | joule | J | N·m | $\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-2}$ | 10^7 erg |
| Force | newton | N | – | $\text{m} \cdot \text{kg} \cdot \text{s}^{-2}$ | 10^5 dyne |
| Frequency | hertz | Hz | – | s^{-1} | |
| Illuminance | lux | lx | lm/m^2 | $\text{m}^2 \cdot \text{m}^{-4} \cdot \text{cd}$ $= \text{m}^{-2} \cdot \text{cd}$ | |
| Inductance | henry | H | Wb/A | $\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-2} \cdot \text{A}^{-2}$ | $\frac{1}{9} \times 10^{-11}$ s ² /cm |
| Luminous flux | lumen | lm | cd·sr | $\text{m}^2 \cdot \text{m}^{-2} \cdot \text{cd} = \text{cd}$ | |
| Magnetic flux | weber | Wb | V·s | $\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-2} \cdot \text{A}^{-1}$ | 10^8 maxwell |
| Magnetic flux density | tesla | T | Wb/m ² | $\text{kg} \cdot \text{s}^{-2} \cdot \text{A}^{-1}$ | 10^4 gauss |
| Plane angle | radian | rad | – | $\text{m} \cdot \text{m}^{-1}$ | |
| Power, Radiant flux | watt | W | J/s | $\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-3}$ | 10^7 erg/s |
| Pressure, Stress | pascal | Pa | N/m ² | $\text{m}^{-1} \cdot \text{kg} \cdot \text{s}^{-2}$ | 10 dyne/cm ² |
| Solid angle | steradian | sr | – | $\text{m}^2 \cdot \text{m}^{-2}$ | |

Table F.4 SI derived units for some physical quantities

| <i>Derived quantity</i> | <i>Name</i> | <i>Symbol</i> |
|---|---------------------------------|------------------------|
| Absorbed dose rate | gray per second | Gy/s |
| Angular acceleration | radian per second squared | rad/s ² |
| Angular velocity | radian per second | rad/s |
| Dynamic viscosity | pascal second | Pa·s |
| Electric charge density | coulomb per cubic metre | C/m ³ |
| Electric field strength | volt per metre | V/m |
| Electric flux density | coulomb per square metre | C/m ² |
| Energy density | joule per cubic metre | J/m ³ |
| Exposure (X- and γ - rays) | coulomb per kilogramme | C/kg |
| Heat capacity, entropy | joule per kelvin | J/K |
| Heat flux density, irradiance | watt per square metre | W/m ² |
| Molar energy | joule per mole | J/mol |
| Molar entropy, Molar heat capacity | joule per mole kelvin | J/(mol·K) |
| Moment of force | newton metre | N·m |
| Permeability | henry per metre | H/m |
| Permittivity | farad per metre | F/m |
| Radiance | watt per square metre steradian | W/(m ² ·sr) |
| Radiant intensity | watt per steradian | W/sr |
| Specific energy | joule per kilogramme | J/kg |
| Specific heat capacity, Specific entropy | joule per kilogramme kelvin | J/(kg ² ·K) |
| Surface tension | newton per metre | N/m |
| Thermal conductivity | watt per metre kelvin | W/(m·K) |

Table F.5 SI prefixes^a

| <i>Factor</i> | <i>Name</i> | <i>Symbol</i> | <i>Factor</i> | <i>Name</i> | <i>Symbol</i> |
|------------------|-------------|---------------|-------------------|-------------|---------------|
| 10 ²⁴ | yotta | Y | 10 ⁻¹ | deci | d |
| 10 ²¹ | zetta | Z | 10 ⁻² | centi | c |
| 10 ¹⁸ | exa | E | 10 ⁻³ | milli | m |
| 10 ¹⁵ | peta | P | 10 ⁻⁶ | micro | μ |
| 10 ¹² | tera | T | 10 ⁻⁹ | nano | n |
| 10 ⁹ | giga | G | 10 ⁻¹² | pico | p |
| 10 ⁶ | mega | M | 10 ⁻¹⁵ | femto | f |
| 10 ³ | kilo | k | 10 ⁻¹⁸ | atto | a |
| 10 ² | hecto | h | 10 ⁻²¹ | zepto | z |
| 10 ¹ | deka | da | 10 ⁻²⁴ | yocto | y |

^a Because the SI prefixes strictly represent powers of 10, they should not be used to represent powers of 2. Thus, one kilobit, or 1 kb, is 1000 bit and **not** 2¹⁰ bit = 1024 bit. To alleviate this ambiguity, prefixes for binary multiples have been adopted by the International Electrotechnical Commission (IEC) for use in information technology (See <http://physics.nist.gov/cuu/Units/binary.html>).

Table F.6 A few common conversions

| <i>Quantity</i> | <i>Unit</i> | <i>Value in specified units</i> |
|-------------------------|--|---------------------------------------|
| Angle | 1 rad | 57.30 degrees |
| Charge | 1 faraday | 9.652×10^4 C (coulomb) |
| | 1 A-h | 3600 C |
| Density* | 1 g/cm ³ | 62.43 lb/ft ³ |
| | 1 lb/ft ³ | 16.02 kg/m ³ |
| Energy/Work/Heat | 1 J | 9.481×10^{-4} BTU |
| | 1 ft-lb | 1.356 J |
| | 1 cal | 4.186 J |
| | 1 kWh | 3.6×10^6 J |
| | 1 eV | 1.609×10^{-19} J |
| Force | 1 N | 7.233 poundal |
| Length | 1 Å | 10^{-10} m |
| | 1 micron | 10^{-6} m |
| | 1 mil | 10^{-3} in |
| | 1 in | 2.54 cm |
| | 1 m | 39.37 in |
| | 1 ft | 30.48 cm |
| Magnetic field strength | 1 oersted | 79.58 A-turn/m |
| | 1 A-turn/in | 39.37 A-turn/m |
| | 1 esu | 2.655×10^{-9} A-turn/m |
| Magnetic flux | 1 esu | 299.8 Wb |
| | 1 maxwell (= 1 line or emu) | 10^{-8} Wb |
| Magnetic induction | 1 gauss (= 1 line/cm ²) | 10^{-4} T |
| Mass | 1 kg | 2.205 lb |
| | 1 lb | 453.6 g |
| | 1 ton | 2000 lb |
| Power | 1 hp | 745.7 W |
| | 1 BTU/h | 0.2930 W |
| | 1 ft-lb/s | 1.356 W |
| Pressure** | 1 atm | 1.01325×10^5 Pa |
| | 1 atm | 14.70 lb/in ² |
| | 1 cm Hg (0°C) | 1333 Pa |
| Volume | 1 m ³ | 35.31 ft ³ |
| | 1 ft ³ | 2.832×10^{-2} m ³ |

*For more density units and their conversions, see Table 13.2 at page 536.

**For more pressure units and their conversions, see Table 8.2 at page 283.

Table F.7 Units outside the SI that are accepted for use with the SI

| <i>Name</i> | <i>Symbol</i> | <i>Value in SI units</i> |
|---------------------------------------|---------------|--|
| astronomical unit ^a | ua | 1 ua = 1.49598×10^{11} m, approximately |
| bel ^b | B | 1 B = $(1/2) \ln 10$ Np ^c |
| day | d | 1 d = 24 h = 86,400 s |
| degree (angle) | ° | 1° = $(\pi/180)$ rad |
| electronvolt ^d | eV | 1 eV = 1.60218×10^{-19} J, approximately |
| hour | h | 1 h = 60 min = 3600 s |
| litre | L | 1 L = 1 dm ³ = 10^{-3} m ³ |
| metric ton ^e | t | 1 t = 10 ³ kg |
| minute (angle) | ' | 1' = $(1/60)^\circ = (\pi/10,800)$ rad |
| minute (time) | min | 1 min = 60 s |
| neper | Np | 1 Np = 1 |
| second (angle) | " | 1" = $(1/60)'$ = $(\pi/648,000)$ rad |
| unified atomic mass unit ^f | u | 1 u = 1.66054×10^{-27} kg, approximately |

^a The astronomical unit is a unit of length. Its value is such that, when used to describe the motion of bodies in the solar system, the heliocentric gravitation constant is $(0.017\ 202\ 098\ 95)^2$ ua³·d⁻². The value must be obtained by experiment, and is therefore not known exactly.

^b The bel is most commonly used with the SI prefix deci: 1 dB = 0.1 B.

^c Although the neper is coherent with SI units and is accepted by the CIPM, it has not been adopted by the General Conference on Weights and Measures (CGPM, *Conférence Générale des Poids et Mesures*) and is thus not an SI unit.

^d The electronvolt is the kinetic energy acquired by an electron passing through a potential difference of 1 V in vacuum. The value must be obtained by experiment, and is therefore not known exactly.

^e In many countries, this unit is called *tonne*.

^f The unified atomic mass unit is equal to 1/12 of the mass of an unbound atom of the nuclide ¹²C, at rest and in its ground state. The value must be obtained by experiment, and is therefore not known exactly.

Table F.8 Other units outside the SI that are currently accepted for use with the SI subject to further review

| <i>Name</i> | <i>Symbol</i> | <i>Value in SI units</i> |
|------------------|---------------|---|
| ångström | Å | 1 Å = 0.1 nm = 10^{-10} m |
| are | a | 1 a = 1 dam ² = 10^2 m ² |
| bar | bar | 1 bar = 0.1 MPa = 100 kPa = 1000 hPa = 10^5 Pa |
| barn | b | 1 b = 100 fm ² = 10^{-28} m ² |
| curie | Ci | 1 Ci = 3.7×10^{10} Bq |
| hectare | ha | 1 ha = 1 hm ² = 10^4 m ² |
| knot | | 1 nautical mile per hour = (1852/3600) m/s |
| nautical mile | | 1 nautical mile = 1852 m |
| rad ^a | rad | 1 rad = 1 cGy = 10^{-2} Gy |
| rem ^b | rem | 1 rem = 1 cSv = 10^{-2} Sv |
| roentgen | R | 1 R = 2.58×10^{-4} C/kg |
| torr | Torr | 1 Torr = (101,325/760) Pa |

^a The rad is a special unit employed to express absorbed dose of ionising radiation. When there is risk of confusion with the symbol for radian, rd may be used as the symbol for rad.

^b The rem is a special unit used in radioprotection to express dose equivalent. Note that this non-SI unit is exactly equivalent to an SI unit with an appropriate submultiple prefix.

Table F.9 A few physical constants

| <i>Name</i> | <i>Symbol</i> | <i>Approximate value</i> |
|-----------------------------|---------------|--|
| Acceleration due to gravity | g | 9.81 m/s ² |
| Avogadro number | N | 6.02×10^{23} /mole |
| Boltzmann constant | k | 1.38×10^{-23} J/K |
| Electron rest mass | m_e | 9.11×10^{-31} kg |
| Elementary charge | e | 1.60×10^{-19} C |
| Faraday constant | F | 9.65×10^4 C/mole |
| Gravitational constant | G | 6.67×10^{-11} N·m ² /kg ² |
| Neutron rest mass | m_n | 1.67×10^{-27} kg |
| Permeability constant | μ_0 | 1.26×10^{-6} H/m |
| Permittivity constant | ϵ_0 | 8.85×10^{-12} farad/m |
| Planck's constant | h | 6.63×10^{-34} J·s |
| Proton rest mass | m_p | 1.67×10^{-27} kg |
| Speed of light | c | 3.00×10^8 m/s |
| Stefan-Boltzmann constant | σ | 5.67×10^{-8} W/m ² K ⁴ |
| Universal gas constant | R | 8.31 J/K mole |

Solutions to Numerical Problems

Chapter 2

2.6 Average value = 104 V. True value = 100 V.

∴ Accuracy = 4%. Measured value = 104 ± 1 V.

∴ Precision = $\frac{1}{104} \times 100 = 0.96\%$

2.7 Each division = $\frac{5}{100}$ V = 0.05 V. $\frac{1}{5}$ th of each division = 0.01 V.

∴ Resolution = 0.01 V.

2.14 (a) (iv) (b) (i) (c) (i) (d) (ii) (e) (ii) (f) (i) (g) (iv)

(h) Current I in the circuit = $\frac{V_s}{60 \times 10^3}$ A. ∴ true output voltage, $V_o = \frac{10 \times 10^3 V_s}{60 \times 10^3} = \frac{V_s}{6}$.

Given, the loaded voltage $V_L = 5$ V and load resistance $R_L = 10 \times 1000 \Omega = 10 \times 10^3 \Omega$.

Now, Thevenin output resistance $R_o = 50 \text{ k}\Omega \parallel 10 \text{ k}\Omega = 50/6 \text{ k}\Omega$. ∴ $(R_o/R_L) = 5/6$.

From Eq (2.10), $V_o = V_L[1 + (R_o/R_L)] = 5[1 + (5/6)] = 55/6$ V. Substituting the value of $V_o = V_s/6$, we get $V_s = 55$ V. ∴ answer is (iii)

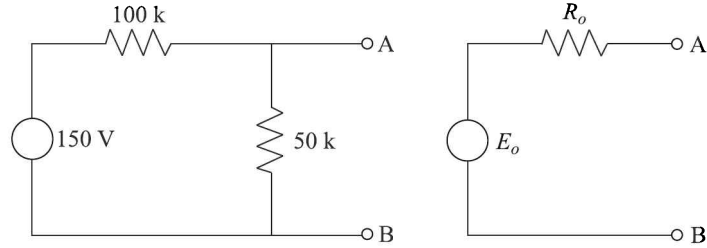
(i) (i) (j) (ii) (k) (iv) (l) (iii)

2.16 Using the symbols of Example 2.4, we get

$$\Delta l = \frac{100D}{\text{FSD}} = \frac{10}{3}D$$

| q_i | q_o | $ D $ | Δl |
|-------------------|-------|-------|------------|
| 0 | 1 | 1 | 3.33 |
| 5 | 4 | 1 | 3.33 |
| 10 | 12 | 2 | 6.67 |
| 15 | 14 | 1 | 3.33 |
| 20 | 22 | 2 | 6.67 |
| 25 | 28 | 3 | 10.0 |
| Av. $\Delta l =$ | | | 5.56% |
| Max. $\Delta l =$ | | | 10.0% |

2.18 The circuit arrangement is



Thevenin equivalent resistance and voltage of the circuit are

$$R_o = \frac{(100)(50)}{100 + 50} = 33.3 \text{ k}\Omega, \quad E_o = 50 \text{ V}$$

True value of the voltage across 50 kΩ resistor = $\frac{(150)(50)}{150 + 50} = 50 \text{ V}$

For Voltmeter 1, the input impedance at 50 V is $50 \text{ V} \times 1 \text{ k}\Omega/\text{V} = 50 \text{ k}\Omega$

$$\therefore E_L = \frac{50}{1 + (33.3/50)} = 30 \text{ V}$$

$$\% \text{error} = \frac{50 - 30}{50} \times 100 = 40$$

For Voltmeter 2, the input impedance = $50 \text{ V} \times 20 \text{ k}\Omega/\text{V} = 1000 \text{ k}\Omega$

$$\therefore E_L = \frac{50}{1 + (33.3/100)} = 48.4 \text{ V}$$

$$\% \text{error} = \frac{50 - 48.4}{50} \times 100 = 3.2$$

$$2.19 \quad V_{AB} = \frac{300 \times 40}{500} = 24 \text{ V.}$$

The Thevenin output resistance $R_o = 200 \Omega \parallel 300 \Omega = 120 \Omega$

Owing to the loading by the voltmeter of $R_L = 1200 \Omega$, the voltmeter will read

$$V_L = \frac{V_o}{1 + (R_o/R_L)} = \frac{24}{1 + (120/1200)} = 21.8 \text{ V}$$

\therefore The answer is $(24 - 21.8) = 2.2 \text{ V}$.

2.20 (a)→(g), (b)→(i), (c)→(e), (d)→(f)

Chapter 3

$$3.2 \text{ Given: } P_f = \cos \phi = \frac{P}{VI}$$

$$(i) \quad \frac{\delta P_f}{P_f} = \frac{\delta P}{P} + \frac{\delta V}{V} + \frac{\delta I}{I} = \pm(0.5 + 1 + 1) = \pm 2.5 \%$$

$$(ii) r_{Pf} = \sqrt{\left(\frac{\partial P_f}{\partial P}\right)^2 r_P^2 + \left(\frac{\partial P_f}{\partial V}\right)^2 r_V^2 + \left(\frac{\partial P_f}{\partial I}\right)^2 r_I^2}$$

To calculate uncertainties, we need to know values of P , V and I . Let us assume,
 $V = 100 \text{ V}$, $I = 100 \text{ A}$

$$\therefore r_V = 100 \times 1\% = 1 \text{ V}, r_I = 100 \times 1\% = 1 \text{ A}$$

$$\text{So, } P = VI = 10,000 \text{ W}$$

$$\therefore r_P = 10,000 \times 0.5\% = 50 \text{ W. Now,}$$

$$\frac{\partial P_f}{\partial P} = \frac{\partial}{\partial P} \left(\frac{P}{VI} \right) = \frac{1}{VI} = \frac{1}{10000}, \quad \frac{\partial P_f}{\partial V} = \frac{\partial}{\partial V} \left(\frac{P}{VI} \right) = -\frac{P}{V^2 I} = -\frac{1}{100}$$

$$\text{and } \frac{\partial P_f}{\partial I} = \frac{\partial}{\partial I} \left(\frac{P}{VI} \right) = -\frac{P}{VI^2} = -\frac{1}{100}$$

$$\therefore r_{Pf} = \sqrt{\left(\frac{1}{10000}\right)^2 (50)^2 + \left(\frac{1}{100}\right)^2 (1)^2 + \left(\frac{1}{100}\right)^2 (1)^2} = 0.015$$

$$\therefore \text{Uncertainty} = \frac{r_{Pf}}{P_f} \times 100 = \frac{0.015}{1} \times 100 = 1.5\%$$

$$3.3 \text{ Relative error} = \frac{205 - 200.4}{200.4} \times 100 = 2.3\%$$

$$3.4 R = R_1 + R_2 + R_3$$

$$\begin{aligned} \frac{\delta R}{R} &= \frac{R_1}{R} \cdot \frac{\delta R_1}{R_1} + \frac{R_2}{R} \cdot \frac{\delta R_2}{R_2} + \frac{R_3}{R} \cdot \frac{\delta R_3}{R_3} \\ &= \left(\frac{47}{167}\right) (4\%) + \left(\frac{65}{167}\right) (4\%) + \left(\frac{55}{167}\right) (4\%) = 4\% \end{aligned}$$

$$\text{Magnitude} = 167 \pm 6.7 \Omega$$

$$3.5 \text{ Voltmeter accuracy} = \pm(150 \times 1\%) \text{ V} = \pm 1.5 \text{ V}$$

$$\text{Ammeter accuracy} = \pm(100 \times 1\%) \text{ mA} = \pm 1 \text{ mA}$$

$$\text{Power } P = VI$$

$$\therefore \frac{\delta P}{P} = \frac{\delta V}{V} + \frac{\delta I}{I} = \pm \left(\frac{1.5}{80} + \frac{1}{70} \right) \times 100 = \pm 3.3\%$$

$$3.6 \text{ (a) } R_x = \frac{R_2 R_3}{R_1} = \frac{(1000)(842)}{100} = 8420 \Omega$$

(b) The limiting error in per cent and that in ohms are calculated as follows

$$\begin{aligned} \frac{\delta R_x}{R_x} &= \frac{\delta R_2}{R_2} + \frac{\delta R_3}{R_3} + \frac{\delta R_1}{R_1} = \pm(0.5 + 0.5 + 0.5)\% \\ &= \pm 1.5\% \\ &= 8420 \times \frac{1.5}{100} = 126.3 \Omega \end{aligned}$$

3.7 The calculation is presented in tabular form as

| x | $ d = \mu - x $ | $ d ^2$ |
|-------|-------------------|---------|
| 41.7 | 0.275 | 0.0756 |
| 42.0 | 0.025 | 0.0006 |
| 41.8 | 0.175 | 0.0306 |
| 42.0 | 0.025 | 0.0006 |
| 42.1 | 0.125 | 0.0156 |
| 41.9 | 0.075 | 0.0056 |
| 42.5 | 0.525 | 0.2756 |
| 41.8 | 0.175 | 0.0306 |
| 335.8 | | 0.4348 |

(a) $\mu = 41.975$ (b) $\sigma_{n-1} = \pm 0.2492$ (c) $r = \pm 0.1681$ (d) $r_m = \pm 0.0594$

3.8 (a) ± 0.334 (b) ± 0.225 (c) ± 0.0795

3.9 (a) 100.66Ω (b) $\sigma_{n-1} = \pm 0.877$ (c) $\nu = 0.769$

3.10 (a) $P = I^2 R = 4^2 \times 100 = 1600 \text{ W}$ $\frac{\partial P}{\partial I} = 2IR = 800$ $r_I = 0.02 \text{ A}$

$$\frac{\partial P}{\partial R} = I^2 = 16 \quad r_R = 0.2 \Omega$$

$$r_P = \sqrt{\left(\frac{\partial P}{\partial I}\right)^2 r_I^2 + \left(\frac{\partial P}{\partial R}\right)^2 r_R^2} = \sqrt{(800)^2(0.02)^2 + (16)^2(0.2)^2} = 16.3$$

$$\therefore \frac{r_P}{P} \times 100 = \frac{16.3}{1600} \times 100 = 1.02\%. \text{ Answer: (iv)}$$

(b) Accuracy is $\pm 1\%$ of the FSD (though not mentioned) = $\pm 1 \text{ V}$. This corresponds to 50% of the value. \therefore Ans. (iv)

(c) Maximum error = $\pm \left(\frac{200 \times 0.25}{100}\right) = \pm 0.5^\circ \text{C}$. Ans. (i)

(d) Given: $r = 40.0 \pm 0.5 \text{ mm}$

Now,

$$\text{Mass } m = \frac{4}{3}\pi r^3 \rho$$

where ρ is the density.

Taking logarithm of both sides,

$$\ln m = \ln 4 - \ln 3 + 3 \ln r + \ln \rho$$

Since ρ is constant, taking differentials of both sides, we get

$$\frac{\delta m}{m} = 3 \frac{\delta r}{r} = \frac{3 \times 0.5}{40}$$

$$\therefore \text{The estimated error in mass} = \frac{1.5}{40} \times 100 = 3.75$$

Ans. (i)

- (e) Given: $R_1 = 120 \Omega \pm 4.0 \Omega$, $R_2 = 110 \Omega \pm 3.0 \Omega$ and $R = R_1 + R_2 = 230 \Omega$
 Thus, $\sigma_{R1} = 4.0 \Omega$ and $\sigma_{R2} = 3.0 \Omega$. The standard deviation of the formed resistor is, therefore

$$\sigma_R = \sqrt{\left(\frac{\partial R}{\partial R_1} \sigma_{R1}\right)^2 + \left(\frac{\partial R}{\partial R_2} \sigma_{R2}\right)^2} = \sqrt{\sigma_{R1}^2 + \sigma_{R2}^2} = \sqrt{4^2 + 3^2} = 5 \Omega$$

Ans. (ii)

- (f) The calculation is given in the following table

| x | y | x^2 | xy |
|-----|------|-------|-------|
| 0 | 9.5 | 0 | 0.0 |
| 1 | 8.4 | 1 | 8.4 |
| 2 | 7.8 | 4 | 15.6 |
| 3 | 7.4 | 9 | 22.2 |
| 4 | 6.1 | 16 | 24.4 |
| 5 | 5.4 | 25 | 27.0 |
| 6 | 5.2 | 36 | 31.2 |
| 7 | 4.6 | 49 | 32.2 |
| 8 | 3.2 | 64 | 25.6 |
| 9 | 1.9 | 81 | 17.1 |
| 10 | 1.1 | 100 | 11.0 |
| 55 | 60.6 | 385 | 214.7 |

Therefore, from Eq. (3.12) the slope is

$$a_1 = \frac{11 \times 214.7 - 55 \times 60.6}{11 \times 385 - (55)^2} = -0.8027$$

Ans. (ii)

- (g) $\mu = \frac{5.9 + 5.7 + 6.1}{3} = 5.9$ V. Therefore, deviations are 0, -0.2 , 0.2 . So, the small

sample standard deviation is $s = \sqrt{\frac{0 + 0.04 + 0.04}{3 - 1}} = \pm 0.2$ Ans. (iv)

- (h) (ii)

- (i) (i)

3.11 The calculation is shown in a tabular form below

| x_i | (b) d_i | $ d_i ^2$ |
|-------|------------------------|-----------|
| 49.7 | -0.16 | 0.0256 |
| 50.1 | 0.24 | 0.0576 |
| 50.2 | 0.34 | 0.1156 |
| 49.6 | -0.26 | 0.0676 |
| 49.7 | -0.16 | 0.0256 |
| 249.3 | (c) $\Sigma d_i = 0.0$ | 0.292 |

- (a) $\mu = 49.86$ (d) $D = 0.232$ (e) $\sigma_{n-1} = 0.2702$

3.12 Voltmeter error = $\pm (100 \times 1.5\%) = \pm 1.5$ V

Ammeter error = $\pm (150 \times 1.5\%) = \pm 2.25$ A $P = VI$

$$\frac{\delta P}{P} = \frac{\delta V}{V} + \frac{\delta I}{I} = \pm \left(\frac{1.5}{70} + \frac{2.25}{80} \right) = \pm 0.0496 = 4.96\%$$

3.13 (a) Power dissipated, $P = \frac{V^2}{R} = \frac{(200)^2}{42} = 952.38$ W

$$(b) \frac{\partial P}{\partial V} = \frac{2V}{R} = \frac{2P}{V} \quad \frac{\partial P}{\partial R} = -\frac{V^2}{R^2} = -\frac{P}{R}$$

$$\begin{aligned} r_P &= \sqrt{\left(\frac{\partial P}{\partial V}\right)^2 r_V^2 + \left(\frac{\partial P}{\partial R}\right)^2 r_R^2} = \sqrt{4\left(\frac{P}{V}\right)^2 r_V^2 + \left(\frac{P}{R}\right)^2 r_R^2} \\ &= P\sqrt{4\left(\frac{r_V}{V}\right)^2 + \left(\frac{r_R}{R}\right)^2} = P\sqrt{4(2)^2 + (1.5)^2}\% \end{aligned}$$

$$\therefore \frac{r_P}{P} = 4.3\%$$

3.14 (a) $y = (5 + 0.75 - 4.2) \pm \sqrt{(0.03^2 + 0.001^2 + 0.001)^2}$
 $= 1.55 \pm 0.03 = 2 \pm 0.03$ after rounding off.

(b) Given, $y = 67.1(\pm 0.25) \times 1.05(\pm 0.02) \times 10^{-17}$

Let, $y = xz$, say. Then, $\frac{\partial y}{\partial x} = z$ and $\frac{\partial y}{\partial z} = x$

$$\begin{aligned} r_y &= \sqrt{\left(\frac{\partial y}{\partial x}\right)^2 r_x^2 + \left(\frac{\partial y}{\partial z}\right)^2 r_z^2} = \sqrt{z^2 r_x^2 + x^2 r_z^2} \\ &= \pm \sqrt{(1.05)^2 (0.25)^2 + (67.1)^2 (0.02)^2} = \pm 1.367 \end{aligned}$$

Thus, $y = (67.1 \times 1.05)(\pm 1.367) \times 10^{-17} = (70.5 \pm 1.4) \times 10^{-17}$, after rounding off.

(c) Given, $y = \frac{2.9(\pm 0.001)}{179(\pm 2)}$

Let $y = x/z$, say. Then $\frac{\partial y}{\partial x} = \frac{1}{z}$ and $\frac{\partial y}{\partial z} = -\frac{x}{z^2}$

$$\begin{aligned} r_y &= \sqrt{\left(\frac{\partial y}{\partial x}\right)^2 r_x^2 + \left(\frac{\partial y}{\partial z}\right)^2 r_z^2} = \sqrt{\left(\frac{r_x}{z}\right)^2 + \left(\frac{r_z x}{z^2}\right)^2} \\ &= \sqrt{\left(\frac{0.001}{179}\right)^2 + \left(\frac{2 \times 2.9}{179^2}\right)^2} = \pm 1.81 \times 10^{-4} \end{aligned}$$

Thus, $y = \frac{2.9}{179}(\pm 0.0002) = 0.0162(\pm 0.0002) = (1.6 \pm 0.02) \times 10^{-3}$, after rounding off.

3.15 See Appendix A.

3.16 When connected in series, $R = R_1 + R_2$ $\frac{\partial R}{\partial R_1} = \frac{\partial R}{\partial R_2} = 1$

$$r_{R_1} = 0.1 \quad r_{R_2} = 0.06\% \text{ of } 50 = 0.03$$

$$r_R = \sqrt{\left(\frac{\partial R}{\partial R_1}\right)^2 r_{R_1}^2 + \left(\frac{\partial R}{\partial R_2}\right)^2 r_{R_2}^2} = \sqrt{0.1^2 + 0.03^2} = \pm 0.1044$$

$\therefore R = 150 \pm 0.1 \Omega$ to the significant figure

When connected in parallel, $R = \frac{R_1 R_2}{R_1 + R_2}$

$$\frac{\partial R}{\partial R_1} = \frac{R_2}{R_1 + R_2} - \frac{R_1 R_2}{(R_1 + R_2)^2} = \frac{50}{150} - \frac{5000}{150^2} = 0.111$$

$$\frac{\partial R}{\partial R_2} = \frac{R_1}{R_1 + R_2} - \frac{R_1 R_2}{(R_1 + R_2)^2} = \frac{100}{150} - \frac{5000}{150^2} = 0.444$$

$$r_R = \sqrt{(0.111)^2(0.1)^2 + (0.444)^2(0.03)^2} = \pm 0.017$$

$\therefore R = 33.33 \pm 0.02 \Omega$

3.17 The relevant calculation is presented in a tabular form as

| x_i | d_i | $ d_i ^2$ |
|-------|------------------------|-----------|
| 99.7 | -0.16 | 0.0256 |
| 100.1 | 0.24 | 0.0576 |
| 100.2 | 0.34 | 0.1156 |
| 99.6 | -0.26 | 0.0676 |
| 99.7 | -0.16 | 0.0256 |
| 499.3 | (c) $\Sigma d_i = 0.0$ | 0.292 |

(a) $\mu = 99.86$ (b) $D = 0.232$ (d) $\sigma_{n-1} = 0.2702$ (e) $\nu = 0.073$

3.18 The given equation is not dimensionally correct. The LHS indicates a voltage whereas the RHS, a resistance. Anyway, the maximum error

$$= \frac{\delta R_1}{R_1} + \frac{\delta R_2}{R_2} + \frac{\delta R_3}{R_3} = \pm(0.1 + 0.1 + 0.1)\% = \pm 0.3\%$$

3.19 If h is small,

$$R = \frac{D^2 + 4h^2}{8h} \cong \frac{D^2}{8h} = \frac{40^2}{8(0.4)} = 500 \text{ mm}$$

Now,
$$\frac{\partial R}{\partial D} = \frac{2D}{8h} = \frac{40}{4(0.4)} = 25$$

and
$$\frac{\partial R}{\partial h} = -\frac{D^2}{8h^2} = -\frac{40^2}{8(0.4)^2} = -1250$$

Uncertainty of D , $r_D = 0.6745 \times 0.05 = 0.0337$ mm and that of h , $r_h = 0.6745 \times 0.001 = 6.745 \times 10^{-4}$ mm. Combining them, we get for the uncertainty in the measurement of R as

$$\begin{aligned} r_R &= \sqrt{\left(\frac{\partial R}{\partial D} r_D\right)^2 + \left(\frac{\partial R}{\partial h} r_h\right)^2} \\ &= \sqrt{(25 \times 0.0337)^2 + (-1250 \times 6.745 \times 10^{-4})^2} = 1.192 \end{aligned}$$

$$\therefore R = 500 \text{ mm} \pm 1.2 \text{ mm}$$

$$3.20 \quad R = \frac{C^2}{8(d-h)} - \frac{h}{2} = \frac{125^2}{8(25-1.25)} - \frac{1.25}{2} = 81.61 \text{ mm}$$

$$\frac{\partial R}{\partial C} = \frac{2C}{8(d-h)} = \frac{2(125)}{8(23.75)} = 1.3158$$

$$\frac{\partial R}{\partial h} = \frac{C^2}{8(d-h)^2} - \frac{1}{2} = 2.9626$$

$$\frac{\partial R}{\partial d} = -\frac{C^2}{8(d-h)^2} = -3.4626$$

$$r_R = \sqrt{1.3158^2 \times 0.05^2 + 3.4626^2 \times 0.001^2 + 2.9626^2 \times 0.0005^2} = 0.0659$$

$$\therefore R = 81.61 \pm 0.066 \text{ mm}$$

Chapter 4

4.1 Here, $\tau = 2$ s, and the temperature variation is sinusoidal. Let it be represented by,

$$y = A_i \sin \omega t$$

where A_i = amplitude of oscillation = $(350 - 300)/2 = 25^\circ\text{C}$

$\omega = 2\pi/T = 2\pi/20 = \pi/10$, T being the time period of oscillation = 20 s (given)

Assuming it to be a first-order system which the temperature measurement devices generally are, we get from Eq. (4.16),

$$A_o = \frac{25}{\sqrt{1 + (2\pi/10)^2}} = 21.17^\circ\text{C}$$

Average temperature was $\frac{350 + 300}{2} = 325^\circ\text{C}$

\therefore Maximum and minimum temperature indications are $(325 + 21.17) = 346.17^\circ\text{C}$ and $(325 - 21.17) = 303.83^\circ\text{C}$

Phase lag $\phi = \tan^{-1}(\omega\tau) = \tan^{-1}(2\pi/10) = 32.14^\circ$

For 360° phase change, it takes 20 s. So, the time lag = $(20/360) \times 32.14 = 1.79$ s.

4.2 Assuming the $Y-t$ recorder to be an underdamped second order system,

$$\text{Maximum overshoot } M_p = \exp\left(-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}\right) = (6.2 - 6.0) = 0.2 \text{ cm}$$

Therefore,

$$\zeta = \frac{1}{\sqrt{1 + (\pi/\ln 0.2)^2}} = 0.456$$

4.7 It is a case of step input, the step being $(30 - 5) = 25$ bar

$$\text{Given: } p_{30} = 25[1 - \exp(-30/\tau)] + 5 = 20$$

$$\text{Therefore, } \tau = -30/\ln[1 - (20 - 5)/25] = 32.74 \text{ s}$$

$$95\% \text{ of the final value} = 28.5 \text{ bar}$$

$$\text{Then, } 25[1 - \exp(-t/32.74)] + 5 = 28.5$$

$$\therefore t = (-32.74) \ln[1 - (28.5 - 5)/25] \cong 92.11 \text{ s}$$

4.8 Given: $m = 8 \times 10^{-3}$ kg, $k = 1000$ N/m

$$\therefore \text{ natural frequency, } \omega_n = \sqrt{k/m} = 353.553 \text{ rad/s}$$

This gives,

$$f_n = \omega_n/2\pi = 56.27 \text{ Hz}$$

For a critically damped system,

$$\zeta = 1 = D/2\sqrt{km}$$

$$\therefore D = 5.66 \text{ kg/s.}$$

4.9 Given: $\tau = 1$ s and step $A = 100^\circ\text{C}$

$$\text{Therefore, from } q_o = 100[1 - \exp(-t/1)]$$

| | | | | |
|----------------------------|------|------|------|------|
| t (s) | 0.5 | 1.0 | 1.5 | 2.0 |
| q_o ($^\circ\text{C}$) | 39.3 | 63.2 | 77.7 | 86.5 |

4.10 For a transient input,

$$(q_o/kA) = (1/\tau) \exp(-t/\tau)$$

$$\text{Given: } (q_o/kA) = 0.33 \quad t = 0.12 \text{ s}$$

$$\text{So, } 0.33 = (1/\tau) \exp(-0.12/\tau) \cong (1/\tau)[1 - (0.12/\tau)]$$

Solving the equation, we get $\tau = 2.905$ s or 0.125 s. The second solution is trivial as can be verified by calculating the corresponding output.

$$\therefore \tau = 2.905 \text{ s}$$

For a cycling input, $T = 1.8$ s

$$\therefore \omega = 2\pi/T = 3.491 \text{ rad/s}$$

$$\text{Then, } (A_o/A_i) = [1/\sqrt{1 + (\omega\tau)^2}] = 0.098 = 9.8\%$$

$$\text{Phase shift, } \phi = \tan^{-1}(-\omega\tau) = -1.472 \text{ rad} = 84.4^\circ$$

If two such systems are cascaded, the transfer function becomes

$$G(s) = \frac{1}{(\tau s + 1)^2} = \frac{1}{\tau^2 s^2 + 2\tau s + 1}$$

Comparing it with the standard 2nd order transfer function $\frac{1}{(s/\omega_n)^2 + 2\zeta(s/\omega_n) + 1}$, we get

$$\tau^2 = 1/\omega_n^2 \quad (\text{i})$$

and

$$\tau = (\zeta/\omega_n) \quad (\text{ii})$$

From Eqs. (i) and (ii), it is apparent that $\zeta = 1$

Thus, $\omega_n = 1/\tau = 1/2.509 \text{ rad/s}$ which yields

$$f_n = \omega_n/2\pi = 1/[(2\pi)(2.905)] = 0.055 \text{ Hz}$$

4.11 Volume change corresponding to 1.5 mm length of the capillary is $\Delta V = \frac{\pi(0.25)^2(1.5)}{4} \text{ mm}^3$. If the required volume of the bulb is V_0 , then from the relation $V = V_0(1 + \gamma\Delta T)$, we have for $\Delta T = 1^\circ$,

$$V_0 = \frac{\Delta V}{\gamma} = \frac{\pi(0.25)^2(1.5)}{(4)(1.8 \times 10^{-4})} \cong 409 \text{ mm}^3$$

4.12 (a) The dynamical equation for the heat exchange is given by

$$MS \frac{dT_0}{dt} = AC(T_i - T_0)$$

where M is the mass of the material of the thermometer taking part in the heat exchange. The value of M is not given. We assume, $M = 0.01 \text{ kg}$. Or else, the next part cannot be worked out.

(b) If we define, $\tau = \frac{MS}{AC}$, and take Laplace transform of the above equation, we get on rearranging

$$\frac{T_o(s)}{T_i(s)} = \frac{1}{s\tau + 1}$$

This is a first order instrument with a ramp input of $T_i = t$ and $A = 1^\circ\text{C/s}$. Now,

$$\tau = \frac{(0.01)(500)}{(100)(10^{-3})} = 50 \text{ s}$$

Since it takes $t = 800 \text{ s}$ to reach 800°C , we get from Eq. (4.14) [see page 90]

$$T_0 = T_i - A\tau\{1 - \exp(-t/\tau)\}$$

$$\Rightarrow T_i = T_0 + A\tau = 800 + 50 = 850^\circ\text{C}$$

4.13 This second-order system has $K = 750$ N/m and $M = 25$ kg.

(a) From Eq. (4.20) [see page 97],

$$\omega_n = \sqrt{K/M} = \sqrt{750/25} = 5.48 \text{ rad/s}$$

(b) Given $\zeta = 0.7$

$$\therefore \text{damped frequency} = \omega_n \sqrt{1 - \zeta^2} = 5.48 \sqrt{1 - 0.7^2} = 3.91 \text{ rad/s}$$

(c) $D = 2\zeta\sqrt{KM} = 2(0.7)\sqrt{(750)(25)} = 191.7$ kg/s

Therefore from Eq. (4.19) [see page 97], the required equation is given by

$$25 \frac{d^2x}{dt^2} + 191.7 \frac{dx}{dt} + 750x = mg$$

where m is the test mass and g , the acceleration due to gravity.

4.14 (a) L/R . 63% [At $\tau = 1$, $q_o/A = 1 - \exp(-1) = 0.63$]

(b) Overdamped.

[If τ_1 and τ_2 are the two time constants, the denominator of the resulting transfer function becomes

$$(1 + s\tau_1)(1 + s\tau_2) = s^2\tau_1\tau_2 + s(\tau_1 + \tau_2) + 1$$

Comparing this with the denominator of the standard second-order transfer function, we get

$$\omega_n = 1/\sqrt{\tau_1\tau_2}$$

and

$$2\zeta\omega_n = \frac{\tau_1 + \tau_2}{\tau_1\tau_2}$$

Substituting the value of ω_n and rearranging, we get

$$\zeta = (\tau_1 + \tau_2)/(2\sqrt{\tau_1\tau_2})$$

which is > 1 .

Proof: $(\sqrt{\tau_1} - \sqrt{\tau_2})^2 > 0$

$$\Rightarrow \tau_1 + \tau_2 - 2\sqrt{\tau_1\tau_2} > 0$$

$$\Rightarrow \tau_1 + \tau_2 > 2\sqrt{\tau_1\tau_2}$$

$$\Rightarrow (\tau_1 + \tau_2)/2\sqrt{\tau_1\tau_2} > 1]$$

(c) $1/(600.2s + 1)$

[Given: $q_o = 1.264$ when $t = 10$ min, and $q_o = 2$ when $t \rightarrow \infty$. From the step input relation for a first-order system $q_o = A\{1 - \exp(-t/\tau)\}$, the second condition gives $q_o = 2 = A$. Substituting this value of A , we get from the first condition $\tau = 600.2$ s. The transfer function for a first-order system is $1/(s\tau + 1)$. Substituting the values of τ here, we get the result. *Note:* The sum states ‘‘A unit step is applied’’. Then, how can it attain ‘‘2 units at steady state’’?]

(d) 3 s

[Here step = (170 – 70) = 100 °C. From the given data, the thermometer indication increases by (133.2 – 70) = 63.2 °C in 3 s. Thus, 63.2 = 100{1 – exp(–3/τ)} which gives τ = 3 s.]

4.15 (a) $q_o = -50[1 - \exp(-0.6/1)] = -22.6$

Temperature = 50 – 22.6 = 27.4 °C

Ans. (iii)

(b) $q_o = (120 - 20)[1 - \exp(-1/1)] = 63.2^\circ\text{C}$

Ans (iii)

(c) (i)

(d) (ii)

(e) (ii)

(f) $y = 1 - 10 \exp(-t) \Rightarrow Y(s) = \frac{1}{s} - \frac{10}{s+1} = \frac{1-9s}{s(s+1)}$

$$\therefore G(s) = \frac{1-9s}{s+1}$$

Ans. (iv)

(g) $2 \frac{d^2y}{dt^2} + 4 \frac{dy}{dt} + 8y = 8x$

$$\Rightarrow \frac{1}{4} \frac{d^2y}{dt^2} + \frac{1}{2} \frac{dy}{dt} + y = x \equiv \frac{1}{\omega_n^2} \frac{d^2y}{dt^2} + \frac{2\zeta}{\omega_n} \frac{dy}{dt} + y$$

By comparison, $\frac{2\zeta}{\omega_n} = \frac{1}{2} = 0.5$. Often, the time constant of a second-order system is defined as $\tau = 2\zeta/\omega_n$

Ans (i)

(h) (iv)

(i) $G(s) = \frac{8}{s^2 + 6s + 8} = \frac{4}{s+2} - \frac{4}{s+4}$

$$\Rightarrow g(t) = 4e^{-2t} - 4e^{-4t}$$

Ans. (i)

(j) $\mathcal{L}\{\sin 2t\} = 2/(s^2 + 4)$

By the first shifting property,

$$\mathcal{L}\{e^{-2t} \sin 2t\} = 2/[(s+2)^2 + 4] = 2/(s^2 + 4s + 8)$$

Ans. (iii)

(k) $\frac{d^2x}{dt^2} + 2 \frac{dx}{dt} + 2x = 1$

Here, $a_0 = 2$, $a_1 = 2$ and $a_2 = 1$

$$\therefore \zeta = \frac{a_1}{2\sqrt{a_0 a_2}} = 0.707$$

Ans. (ii)

- (l) Comparing the transfer function with a standard 2nd order one, we get

$$\omega_n^2 = 25$$

$$\Rightarrow \omega_n = 5 \text{ rad/s}$$

$$\zeta = 6/2\omega_n = 0.6$$

$$\omega_d = \omega_n \sqrt{1 - \zeta^2} = 5\sqrt{1 - 0.6^2} = 4$$

Ans. (i)

- (m) (ii)

- (n) The system is critically damped. Therefore, from Eq. (4.21) [see page 97]

$$\zeta = \frac{D}{2\sqrt{KM}} = 1$$

which yields $D = 2\sqrt{KM}$.

Ans. (iv)

- (o) This is a first-order system with a ramp input of $\tau = 0.5$ min and $A = 5^\circ\text{C}/\text{min}$. From Eq. (4.15) [see page 91], the steady-state error = $A\tau = (0.5)(5) = 2.5^\circ\text{C}$

Ans. (ii)

- (p) This is a first-order system with a step input. From Eq. (4.12) [see page 89],

$$\frac{q_o}{KA} = 1 - \exp(-t/\tau) = 0.98$$

Solving for t , we get $t = 3.91\tau \cong 4\tau$.

Ans. (iii)

- (q) Given: $\tau = 5$ s, $A_i = 10^\circ\text{C}$ and time period $T = 20$ s

$$\therefore \omega = 2\pi/T = \pi/10$$

From Eq. (4.16) [see page 93],

$$A_o = \frac{A_i}{\sqrt{1 + \omega^2\tau^2}} = \frac{10}{1 + \sqrt{(0.1\pi)^2(25)}} = 5.37^\circ\text{C}$$

Ans. (ii)

- (r) $1 - \exp(-t/\tau) = 0.95$

$$\Rightarrow -(t/\tau) = \ln 0.05 = -2.995$$

$$\Rightarrow t \cong 3\tau$$

Ans. (i)

- (s) $A = 100 - 30 = 70^\circ\text{C}$

From Eq. (4.11) [see page 89], $66.5 = 70[1 - \exp(-30/\tau)]$

On solving this equation, we get $\tau = 10$ s. Now, from the second condition, $68 = 70[1 - \exp(-t/10)]$

This yields, $t = 35.6$ s

Ans. (iii)

- (t) In this case of a ramp input to a first-order system, the ramp equation is given by $\theta_i(t) = kt$ and the time constant is τ . But the steady state error has been defined in the opposite way. Therefore, the steady state error will be $-k\tau$ instead of $k\tau$.

Ans. (iv)

Chapter 5

5.3 (a)→(v), (b)→(vi), (c)→(i), (d)→(iii)

5.4 (a)→(h), (b)→(e), (c)→(f), (d)→(g)

5.5 (a)→(h), (b)→(f), (c)→(e), (d)→(g)

5.6 (a)→(f), (b)→(g), (c)→(h), (d)→(e)

5.7 (a) (i), (b) (iv), (c) (iv)

Chapter 6

6.8 Given: $L = 200$ cm, its resistance = $1 \Omega/\text{cm}$, and $E = 1$ V.

- (a) $R_L = 200 \Omega$. From the balancing condition of two currents across AC, we have

$$\frac{4}{R_H + 200} = \frac{E}{100 + 50} = \frac{1}{150}$$

This yields, $R_H = 400 \Omega$. The potentiometer current = $(1/150) = 6.67$ mA.

- (b) If one cell is reversed, both of them will be in series and the circuit will have a resistance of $400 + 100$ (from B to C) + $50 = 550 \Omega$. So, the current through the galvanometer will be $\frac{4 + 1}{550} \cong 9$ mA.

6.9 (a) (i) and (ii)

- (b) From Eq. (6.34) [see page 208], we know

$$N_A = \sqrt{n_1^2 - n_2^2} = \sqrt{1.44^2 - 1.40^2} = 0.34$$

Ans. (iv)

- (c) (i)

(d) (ii) because it is the only code which changes one bit in successive steps.

- (e) If P is the power, I current and R resistance, from the given data $I = \sqrt{P/R} = \sqrt{4/10000} = 0.02$ A

So, voltage across R is $10000 \times 0.02 = 200$ V

\therefore Sensitivity = $(200/100) = 2.0$ V/mm

Ans. (ii)

(f) (iv)

(g) Given: pulses/s = 5500

$$\therefore \text{Pulses/min} = 5500 \times 60$$

Sensitivity is 500 pulses/revolution. So,

$$\text{rpm} = \frac{5500 \times 60}{500} = 660$$

Ans. (iv)

(h) (i) [see Fig. 6.16 at page 185]

(i) (i)

(j) Given: $R_p = 4 \text{ k}\Omega$, 0.02 W and $E_i = 15 \text{ V}$. So,

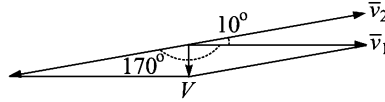
$$i_p = \sqrt{W/R_p} = \sqrt{0.02/(4 \times 10^3)} = 2.236 \text{ mA}$$

$$\text{Also, from } i_p = \frac{E_i}{R_s + R_p}$$

$$R_s = \frac{E_i}{i_p} - R_p = \frac{15}{2.236 \times 10^{-3}} - 4 \times 10^3 = 2.71 \text{ k}\Omega$$

Ans. (ii)

(k) (ii)

(l) Given: $\bar{v}_1 = 1.0 \angle 0^\circ$ and $\bar{v}_2 = 1.0 \angle 10^\circ$. In an LVDT, the output voltages are connected in series opposition. So, the situation is like that shown in the following figure.It is clear from the figure, $V = \sqrt{2 + 2 \cos 170^\circ} = 0.174$

Ans. (iii)

(m) Since, the electrostatic attraction between two plates of a charged capacitor is a Coulomb force, $F \propto 1/x^2$. Alternatively, the potential energy ϕ of a charged capacitor is given by

$$\phi = \frac{1}{2} CV^2 = \frac{\epsilon A}{2x} V^2$$

So,

$$F = -\frac{\partial \phi}{\partial x} \propto \frac{1}{x^2}$$

Ans. (ii)

(n) Given: voltmeter sensitivity = $500 \Omega/\text{V}$ and FS = 100 V

$$\therefore \text{Voltmeter resistance } R_v = 50 \text{ k}\Omega$$

Let, $R = R_x \parallel R_v$. So, the circuit current

$$I = \frac{V}{R_s + R}$$

The voltage across $R = \frac{VR}{R_s + R} = 20$ V (given)

This gives,

$$\frac{100R}{100 \times 10^3 + R} = 20$$

yielding $R = 25 \text{ k}\Omega$

$$\text{Now, } R = 25 = \frac{R_x R_v}{R_x + R_v} = \frac{50R_x}{50 + R_x}$$

giving $R_x = 50 \text{ k}\Omega$

Ans. (ii)

$$(o) \text{ Let, } R = R_C \parallel R_L = \frac{R_C R_T}{R_C + R_T} \quad [\because R_L = R_T \text{ (given)}]$$

Resistance of the circuit is

$$(R_T - R_C) + R = (R_T - R_C) + \frac{R_C R_T}{R_C + R_T} = \frac{R_C^2 - R_T^2 + R_C R_T}{R_C + R_T}$$

$$\text{Therefore, } \frac{V_o}{V_s} = \frac{1}{2} = \frac{R}{(R_T - R_C) + R} = \frac{R_C R_T}{R_C^2 - R_T^2 + R_C R_T}$$

$$\text{This yields } R_C^2 - R_C R_T - R_T^2 \equiv x^2 - x - 1 = 0 \quad [x = R_C/R_T]$$

This quadratic equation has the positive solution of $x = \frac{1 + \sqrt{5}}{2}$.

Ans. (ii)

(p) From Eq. (6.34) [see page 208],

$$N_A = n \sin \theta$$

Given: $n = 1.33$ and $N_A = 0.39$

So, $\theta = 17.05^\circ$

Ans. (iii)

(q) (i)

(r) (iv)

6.10 (a) magnetic

$$(b) \sqrt{n_1^2 - n_2^2}$$

(c) (iii), (ii)

Chapter 7

7.1 (c) Change in resistance of the gauge because of change in temperature,

$$\Delta R = R\alpha\Delta T = (120)(12 \times 10^{-6})(60) = 0.0864 \Omega$$

Strain due to differential expansion of gauge metal and steel

$$\varepsilon = (\Delta\alpha_T)\Delta T = (13 - 30) \times 10^{-6} \times 60 = -0.00102$$

[Note that the gauge gets compressed]

Assuming $G_f = 2$ for 'advance', corresponding change in resistance,

$$\Delta R' = G_f \varepsilon R = (2)(-0.00102)(120) = -0.2448 \Omega$$

Total change of gauge resistance = $-0.2448 + 0.0864 = -0.1584 \Omega$

Since the equivalent resistance of an equal-arm bridge is R , the input voltage

$$E_i = (120)(25 \times 10^{-3}) = 3 \text{ V}$$

$$\therefore \Delta E = \frac{\Delta R}{4R} E_i = -\frac{0.1584 \times 3}{4 \times 120} = -0.99 \text{ mV}$$

7.2 Here, $G_f = 2$, $\varepsilon = 1 \times 10^{-6}$, $R = 130 \Omega$

$$\therefore \Delta R = G_f \varepsilon R = 2.6 \times 10^{-4} \Omega$$

7.3 (c) Here, $R = 120 \Omega$, $G_f = 2$, $E_i = 4 \text{ V}$, $R_m = 100 \Omega$ and $\varepsilon = 1 \times 10^{-6}$

$$\therefore i_m = \frac{G_f \varepsilon E_i}{4(R + R_m)} = \frac{(2)(1 \times 10^{-6})(4)}{4(120 + 100)} = 9.09 \times 10^{-9} \text{ A} = 9.09 \text{ nA}$$

7.5 (c) $G_f = 2.5$, $\varepsilon = 50 \times 10^{-6}$, $R = 120 \Omega$

$$\therefore \Delta R = G_f \varepsilon R = 0.015 \Omega$$

7.6 (c) $R = 200 \Omega$, $E_i = 10 \text{ V}$. Force $F = 100 \text{ N}$, Area $A = 2 \text{ cm} \times 2 \text{ cm} = 0.0004 \text{ m}^2$

Young's modulus $Y = 2.1 \times 10^{11} \text{ N/m}^2$

$$\therefore \varepsilon = \frac{F}{AY} = \frac{100}{(0.0004)(2.1 \times 10^{11})} = 1.19048 \times 10^{-6}$$

$$\Delta E_o = \frac{G_f \varepsilon}{4} E_i = \frac{(2)(1.19048 \times 10^{-6})(10)}{4} = 5.95 \mu\text{V}$$

7.7 (c) $R = 350 \Omega$, $\varepsilon = 500 \times 10^{-6}$

Change in resistance $\Delta R = (2)(500 \times 10^{-6})(350) = 0.35 \Omega$

7.8 Given: $G_f = 2$, $\sigma = 1000 \text{ kg/cm}^2$ and $E = 2 \times 10^6 \text{ kg/cm}^2$

Therefore,

$$\varepsilon = \frac{\sigma}{E} = \frac{1000}{2 \times 10^6} = 5 \times 10^{-4}$$

The per cent change in resistance of the gauge is

$$100 \times \frac{\Delta R}{R} = 100 G_f \varepsilon = 100(2)(5 \times 10^{-4}) = 0.1$$

Poisson's ratio cannot be calculated with the available data.

7.9 Since this is a quarter bridge,

$$(E_o)_{\text{bridge}} = \frac{G_f \varepsilon E_i}{4} = \frac{(2)\varepsilon(12)}{4} = 6\varepsilon \text{ V}$$

Given: $A_D = 2500$, $\text{CMRR} = 20 \log \frac{A_D}{A_{\text{com}}} = 80 \text{ dB}$

This gives $\frac{A_D}{A_{\text{com}}} = 10000$

So, $A_{\text{com}} = \frac{2500}{10000} = 0.25$

Common mode voltage $V_{\text{com}} = \frac{E}{2} = \frac{12}{2} = 6 \text{ V}$

Therefore,

$$\text{Common mode output } (V_{\text{com}})_o = 6 \times 0.25 = 1.5 \text{ V}$$

It means that the actual strain output is $(4.5 - 1.5) = 3.0 \text{ V}$. Now, $A_D(6\varepsilon) = \text{output voltage}$. Thus,

$$\text{Actual strain} = \frac{\text{Actual output voltage}}{A_D \times 6} = \frac{3.0}{(2500)(6)} = 200 \mu\text{strain}$$

$$\text{Indicated strain} = \frac{\text{Indicated output voltage}}{A_D \times 6} = \frac{4.5}{(2500)(6)} = 300 \mu\text{strain}$$

7.10 (b) In this half-bridge,

$$\Delta E_o = \frac{G_f \varepsilon}{2} E_i$$

Given: $G_f = 2.0$, $E_i = 1.0 \text{ V}$ and $(\Delta E_o)_{\text{min}} = 1 \times 10^{-6} \text{ V}$

$$\therefore \varepsilon_{\text{min}} = \frac{2(\Delta E_o)_{\text{min}}}{G_f E_i} = \frac{2(10^{-6})}{(2)(1)} = 1 \mu\text{strain}$$

7.11 Given: $\nu = 0.3$

For a strain of ε

$$\Delta R_{G1} = (120)(2.0)\varepsilon = 240\varepsilon$$

$$\therefore R_{G1} = (120 + 240\varepsilon)\Omega$$

$$\Delta R_{G2} = (120)(2.0)\nu\varepsilon = (240)(0.3)\varepsilon = 72\varepsilon$$

$$\therefore R_{G2} = (120 + 72\varepsilon)\Omega \text{ So,}$$

$$V_A = \frac{1000}{2000} E_i = \frac{E_i}{2} \quad V_C = \frac{R_{G1}}{R_{G1} + R_{G2}} E_i = \frac{120 + 240\varepsilon}{340 + 312\varepsilon} E_i$$

$$\Delta E_o = V_C - V_A = \frac{120 + 240\varepsilon}{340 + 312\varepsilon} E_i - \frac{E_i}{2}$$

For $\varepsilon = 1 \mu\text{strain}$ and $E_i = 1 \text{ V}$,

$$\Delta E_o = \frac{120 + (240)(1 \times 10^{-6})}{240 + (312)(1 \times 10^{-6})} - \frac{1}{2} = 0.50000035 - 0.5 = 0.35 \mu\text{V}$$

7.12 (a) (i)

(b) For a quarter bridge, $E_o = G_f \varepsilon E_i / 4$ Given: $E_o = 1 \text{ mV}$, $\varepsilon = 500 \times 10^{-6}$ and $E_i = 4 \text{ V}$

$$\therefore G_f = \frac{4E_o}{\varepsilon E_i} = \frac{4(1 \times 10^{-3})}{(500 \times 10^{-6})(4)} = 2$$

Ans. (ii)

(c) (iv)

(d) (i)

(e) Given: $R_g = 100 \Omega$, $\Delta R_g = 0.35 \Omega$, $L = 20 \text{ cm}$, and $\Delta L = 0.2 \text{ mm}$

$$\therefore \varepsilon = \Delta L / L = (0.02) / (20) = 0.001 \text{ and } \Delta R / R = (0.35) / (100) = 0.0035$$

$$G_f = (\Delta R / R) / \varepsilon = (0.0035) / (0.001) = 3.5$$

Ans. (ii)

(f) (iii)

(g) The resistance of the combination is

$$\frac{(120)(200 \times 10^3)}{120 + 200 \times 10^3} = 119.928 \Omega$$

$$\therefore \Delta R = 119.928 - 120 \cong -0.072 \Omega$$

$$\text{Thus, } \varepsilon = \frac{\Delta R / R}{G_f} = -\frac{0.072}{(120)(2)} = -300 \mu\text{m/m}$$

Ans. (iv)

(h) (i)

(i) Given: $R_g = 600 \Omega$, $G_f = 2.5$, $\varepsilon = 100 \times 10^{-6}$, and $E_i = 4 \text{ V}$

This is a quarter bridge

$$\therefore E_o = \frac{G_f \varepsilon}{4} E_i = \frac{(2.5)(100 \times 10^{-6})}{4} = 250 \mu\text{V}$$

Ans. (ii)

Chapter 8

8.6 Charge sensitivity $d = Q/F$ where Q = charge and F = force applied. The unit of charge sensitivity is pC/N. Here, the sensitivity unit is pC/Torr, which means d has been defined as $d = Q/p$ where p = pressure applied. Therefore, we define voltage sensitivity as voltage per unit pressure. Then,

$$\text{Voltage generated } E_o = \frac{Q}{C} = \frac{pd}{C}$$

$$\text{Voltage sensitivity} = \frac{E_o}{p} = \frac{d}{C} = \frac{7.84 \text{ pC/Torr}}{200 \text{ pF}} = 0.0392 \text{ V/Torr}$$

The natural frequency of the crystal cannot be calculated because of insufficient data.

8.12 (b) Let α_1 and α_2 be the cross-sectional areas of the well and the capillary. Then from the given data,

$$\frac{\alpha_2}{\alpha_1} = \frac{d_2^2}{d_1^2} = \frac{1}{400}$$

If a pressure differential Δp acts between the limbs and if h_1 = depression of mercury column in the well, and h_2 = rise of the same in the capillary,

$$\Delta p = (h_1 + h_2)\rho g = 1 \text{ Pa}$$

Now, from the continuity of mercury thread,

$$\alpha_1 h_1 = \alpha_2 h_2 \quad \Rightarrow h_1 = \frac{\alpha_2}{\alpha_1} h_2$$

Therefore

$$\Delta p = \left(\frac{\alpha_2}{\alpha_1} + 1 \right) h_2 \rho g \cong h_2 \rho g \quad [\because (\alpha_2/\alpha_1) \ll 1]$$

This is why it is said, for a well-type manometer, it is sufficient to read the mercury column height in the capillary. Now,

$$\begin{aligned} h_2 &= \frac{\Delta p}{\rho g} = \frac{1 \text{ N/m}^2}{(13.6 \text{ g/cm}^3)(981 \text{ cm/s}^2)} \\ &= \frac{10 \text{ dyne/cm}^2}{(13.6 \text{ g/cm}^3)(981 \text{ cm/s}^2)} = 7.495 \times 10^{-4} \text{ cm} \end{aligned}$$

8.17 See Section 8.5.3 at page 299.

8.18 Given: $d = 2 \text{ pC/N} = 2 \times 10^{-12} \text{ C/N}$, $E = 8.6 \times 10^{10} \text{ N/m}^2$, $\varepsilon = 42 \times 10^{-12} \text{ F/m}$ and $|x| = 10^{-8} \text{ m}$. We know,

$$E = \frac{F/A}{x/t}$$

$$\therefore \frac{F}{A} t = Ex$$

The open circuit output voltage is

$$E_o = \frac{d}{\varepsilon} \cdot \frac{F}{A} \cdot t = \frac{d}{\varepsilon} \cdot Ex = \frac{2 \times 10^{-12}}{42 \times 10^{-12}} (8.6 \times 10^{10})(10^{-8}) = 40.9 \text{ V}$$

8.19 Given: $d = 2 \times 10^{-12} \text{ C/N}$, $\varepsilon = 4 \times 10^{-11} \text{ F/m}$, $E = 8.6 \times 10^{10} \text{ N/m}^2$, $t = 4 \text{ mm} = 4 \times 10^{-3} \text{ m}$, $x = 10^{-9} \text{ m}$, and $D = 8 \text{ mm} = 8 \times 10^{-3} \text{ m}$ which gives $A = \pi(8 \times 10^{-3})^2/4 = 5.0265 \times 10^{-5} \text{ m}^2$

$$(a) \text{ Force } F = Ex \frac{A}{t} = (8.6 \times 10^{10})(10^{-9}) \frac{(5.0265 \times 10^{-5})}{(4 \times 10^{-3})} = 1.0807 \text{ N}$$

$$(b) \text{ Capacitance } C = \frac{\varepsilon A}{t} = \frac{(4 \times 10^{-11})(5.0265 \times 10^{-5})}{(4 \times 10^{-3})} = 5.0265 \times 10^{-13} \text{ F} = 0.503 \text{ pF}$$

$$(c) \text{ Charge } Q = dF = (2 \times 10^{-12})(1.0807) = 2.1614 \text{ C} = 2.1614 \text{ pC}$$

$$(d) \text{ Developed voltage } E_o = \frac{2 \times 10^{-12}}{4 \times 10^{-11}} \cdot \frac{1.0807}{5.0265 \times 10^{-5}} \cdot (4 \times 10^{-3}) = 4.3 \text{ V}$$

8.20 Given: $d = 2 \times 10^{-12} \text{ C/N}$, $C_p = 1600 \times 10^{-12} \text{ F}$, $R_p = 10^{12} \Omega$, $R = 10^8 \Omega$, $C = 10^{-9} \text{ F}$, and $F = 0.1 \sin 10t$.

Therefore,

$$\text{Charge developed } Q = dF = (2 \times 10^{-12})(0.1) \sin 10t = 0.2 \times 10^{-12} \sin 10t$$

$$\text{Input voltage } e_i = \frac{Q}{C_p} = \frac{0.2}{1600} \sin 10t \text{ V}$$

$$\text{Input impedance } Z_i = R_p \parallel C_p = \frac{R_p}{\sqrt{1 + \omega^2 R_p^2 C_p^2}}$$

$$\text{Input current } i_i = \frac{e_i}{Z_i} = \frac{e_i}{R_p} \sqrt{1 + \omega^2 R_p^2 C_p^2}$$

$$\text{Output impedance } Z_o = R \parallel C = \frac{R}{\sqrt{1 + \omega^2 R^2 C^2}}$$

Output voltage $= e_o$

$$\text{Output current } i_o = \frac{e_o}{Z_o} = \frac{e_o}{R} \sqrt{1 + \omega^2 R^2 C^2}$$

Since, $|i_i| = |i_o|$, we have

$$\frac{e_i}{R_p} \sqrt{1 + \omega^2 R_p^2 C_p^2} = \frac{e_o}{R} \sqrt{1 + \omega^2 R^2 C^2}$$

Thus,

$$\begin{aligned} |e_o| &= \frac{R}{R_p} \sqrt{\frac{1 + \omega^2 R_p^2 C_p^2}{1 + \omega^2 R^2 C^2}} \cdot |e_i| \\ &= \frac{10^8}{10^{12}} \sqrt{\frac{1 + (10)^2 (10^{12})^2 (1600 \times 10^{-12})^2}{1 + (10)^2 (10^8)^2 (10^{-9})^2}} \cdot \frac{0.2}{1600} \\ &= 10^{-4} \cdot \frac{16 \times 10^3}{\sqrt{2}} \cdot \frac{0.2}{1600} \text{ V} \\ &\cong 0.141 \text{ mV} \end{aligned}$$

So, Ans. (a)

8.21 From Eq. (8.4), we have

$$h_d = \left[\sin \theta + \frac{\alpha_t}{\alpha_w} \right] l$$

This gives
$$\sin \theta = \left[\frac{h_d}{l} - \frac{\alpha_t}{\alpha_w} \right]$$

So,
$$\theta = \sin^{-1} \left[\frac{h_d}{l} - \frac{\alpha_t}{\alpha_w} \right]$$

Ans. (c)

8.22 (a) (iv) (b) (iv) (c) (ii)

(d) (i) [Capacitance introduced by the cable \propto length of the cable. So, if the cable length is doubled, so will be the cable capacitance.]

$$\text{Frequency} \propto \frac{1}{C}$$

$$\therefore \frac{f_{\text{new}}}{f_{\text{old}}} = \frac{C_{\text{old}}}{C_{\text{new}}} = \frac{1}{2}$$

$$\Rightarrow f_{\text{new}} = \frac{f_{\text{old}}}{2} = 500 \text{ Hz}$$

(e) (i) (f) (iii) [∵ for vacuum, the gauge pressure is negative.] (g) (iv) (h) (iii)

(i) (iv) (j) (i) (k) (i) [Zero drift voltage $(0.01)(20) = 0.2 \text{ V}$]

At 20°C, the sensitivity is

$$[10 + (0.01 \times 20)] = 10.2 \text{ V/MPa}$$

$$\therefore 0.2 + 10.2p = 7.4 \Rightarrow p = 0.71 \text{ MPa}$$

(l) (iii) [There are $(23 \times 24 \times 3600) = 1987200 \text{ s}$ in 23 days. But the clock generates $(1987200 - 30.32) = 1987169.68 \text{ s}$. This is produced by a crystal of frequency 32.768 kHz. To produce the exact number of seconds, the frequency of the crystal should be $\frac{32.768}{1987167.68} \times 1987200 = 32.7685 \text{ kHz}$]

(m) (i) [If α is the area of cross-section of the barometer tube, we have $h\alpha = h_{\text{true}}(1 + \beta T)\alpha \Rightarrow h_{\text{true}} = h/(1 + \beta T)$]

(n) (ii) (o) (iii) (p) (ii) (q) (iii)

Chapter 9

9.1 Since it is a full bridge, $\Delta E_o = \frac{\Delta R}{R} E_i = \frac{5.2}{300}(7.5) = 0.13 \text{ V}$.

$$\text{Sensitivity} = \frac{\Delta E_o}{\text{Force}} = \frac{0.13}{0.1} = 1.3 \text{ V/N}$$

9.2 (b) Given: $\varepsilon = 500 \times 10^{-6}$, $R_g = 120 \Omega$, $G_f = 2.1$ and $I_{\text{max}} = 50 \text{ mA}$. ∵ this will be a full-bridge measurement, for each arm one gauge will be compressed and the other stretched, keeping the resistance of each arm = $2R_g$. The total bridge resistance is R_g

$[2R_g \parallel 2R_g]$. For maximum 50 mA current in each arm, the total bridge current should be 100 mA. So, the input voltage is $(120)(0.1) = 12$ V

\therefore From Eq. (7.24) (see page 251),

$$\Delta E_o = (2.1)(500 \times 10^{-6})(12) = 12.6 \text{ mV}$$

9.3 Given: $\omega_n = 8$ Hz, $\omega = 16$ Hz, $\zeta = 0.8$, and $x_0 = 1.5$ mm

$$\therefore u = 16/8 = 2$$

From Eq. (9.8) (see page 332),

$$x_i = \frac{x_0 \sqrt{(1-u^2)^2 + (2\zeta u)^2}}{u^2} = \frac{1.5 \sqrt{(1-4)^2 + (3.2)^2}}{4} = 1.64 \text{ mm}$$

$$\therefore \text{Error} = (1.64 - 1.5) = 0.14 \text{ mm}$$

9.4 Given: $G = 8 \times 10^{10}$ N/m², $l = 150$ mm, $r_1 = 33$ mm, $r_2 = 25$ mm and $\varepsilon_{45} = 5.5 \times 10^{-6}$.
From Eq. (9.17) (see page 356),

$$\begin{aligned} T &= \pi G \varepsilon_{45} \left(\frac{r_1^4 - r_2^4}{r_1} \right) \\ &= \pi (8 \times 10^{10}) (5.5 \times 10^{-6}) \left[\frac{(33 \times 10^{-3})^4 - (25 \times 10^{-3})^4}{33 \times 10^{-3}} \right] \\ &= 33.3 \text{ N-m} \end{aligned}$$

From Eqs. (9.14) and (9.17) (see page 356),

$$\frac{\phi}{2l} = \frac{\varepsilon_{45}}{r_1} \text{ which gives}$$

$$\phi = \varepsilon_{45} \frac{2l}{r_1} = (5.5 \times 10^{-6}) \frac{2(150)}{33} = 5 \times 10^{-5} \text{ rad} = 0.0029^\circ$$

9.5 Because it is a linear acceleration, putting $u = 0$ in Eq. (9.9) (see page 333), we get

$$a_0 = \omega_n^2 x_0 \text{ or, } x_0 = \frac{a_0}{\omega_n^2}$$

$$\text{Given: } \omega_n = 300 \pm 3 \text{ Hz, } a_0 = 300 \pm 5\% = 300 \pm 15 \text{ m/s}^2$$

Thus, the allowable uncertainty in the displacement measurement is

$$\begin{aligned} r_{x_0} &= \sqrt{\left(\frac{\partial x_0}{\partial a_0} \Delta a_0 \right)^2 + \left(\frac{\partial x_0}{\partial \omega_n} \Delta \omega_n \right)^2} \quad \frac{\partial x_0}{\partial a_0} = \frac{1}{\omega_n^2} \quad \frac{\partial x_0}{\partial \omega_n} = -\frac{2a_0}{\omega_n^3} \\ \therefore r_{x_0} &= \sqrt{\left(\frac{\Delta a_0}{\omega_n^2} \right)^2 + \left(\frac{2a_0 \Delta \omega_n}{\omega_n^3} \right)^2} = \sqrt{\left(\frac{15}{300^2} \right)^2 + \left(\frac{300 \times 6}{300^3} \right)^2} \cong \pm 1.8 \times 10^{-4} \text{ m} \\ &= \pm 0.18 \text{ mm} \end{aligned}$$

9.6 Given: $a_0 = 5g$, $\zeta = 0.8$, $M = 0.005$ kg, $k = 20$ N/m and $g = 9.81$ m/s²

- (a) $\omega_n = \sqrt{20/0.005} = 63.25$ rad/s. \therefore linear acceleration, $a_0 = \omega_n^2 x_0$. So, linear displacement of the potentiometer is

$$x_0 = a_0/\omega_n^2 = (5 \times 9.81)/(63.25)^2 = 0.012 \text{ m} = 1.2 \text{ cm}$$

Damping constant $D = 2\zeta\sqrt{kM} = 2(0.8)\sqrt{(20)(0.005)} = 0.51$ kg/s

- (b) Since, $x_0 \propto a_0$, for $2g$ acceleration, the displacement of the potentiometer, from $x/x_0 = 2/5$, is $x = 0.48$ cm. Given, the pot resistance = 1 k Ω and the recorder resistance = 10 k Ω . Had there been no recorder, the voltage developed across the wiper would be $E_i \times \frac{(0.4)(1000)}{1000} = 0.4E_i$ V. The recorder is parallel to the wiper of the pot. So, the resistance of the combination is

$$\frac{(0.4)(10)}{0.4 + 10} = 0.38 \text{ k}\Omega$$

Now, the total resistance of the circuit is $(0.6 + 0.38) = 0.98$ k Ω and the measured voltage across the wiper is $\frac{0.38}{0.98}E_i = 0.39E_i$ V. Therefore,

$$\text{Error} = \frac{0.4 - 0.39}{0.4} \times 100 = 2.5\%$$

9.7 (a) (ii) [$\because \omega/\omega_n < 0.4$]

(b) (iv)

(c) (iv) [Because accelerometers operate at $f \ll f_n$ and are independent of f_n]

(d) (i) [$F = D \frac{dx}{dt}$ where D is the damping constant. $\therefore D = \frac{F}{dx/dt} = \frac{20 \text{ N}}{10 \text{ mm/s}} = 2 \text{ Ns/mm}$]

(e) (iii) [k is the same, $\therefore \omega \propto 1/\sqrt{m}$. This gives $\omega_2 = \omega_1/\sqrt{m_2} = 0.1\omega_1$]

(f) (i) [The piezoelectric transducer does not respond to static pressure]

(g) (iii)

Chapter 10

10.5 (a) (iv) (b) can be all (c) (i) (d) (iii) (e) (ii) (f) (iii) (g) (ii) (h) (iii) (i) (iii) (j) (ii) (k) (ii) (l) (iii) (m) (iii) (n) (i) [Note: 'Fixed points' specified] (o) (iv) (p) (iii) [Note: A 'range' is specified] (q) (ii) (r) (i) (s) (i) [$(55 \times 10^{-6})(300 - 100) = 11.0$ mV]

(t) (ii) [Neglecting C which is small, A and B constants of Eq. (10.9) [see page 393] can be evaluated from the given data as $A = 6.568 \times 10^{-4}$ and $B = 2.9303 \times 10^{-4}$. Substituting these values in the inverse form $R_T = \exp[\frac{1}{B}(\frac{1}{T} - A)]$, we get $R_{423} = 339.073 \Omega$. So, the current I in the circuit is

$$\frac{5}{1339.073} = 3.7339 \times 10^{-3} \text{ A}$$

Thus, the power dissipation through the thermistor is $I^2 R_{423} = 4.7 \times 10^{-3}$ W]

(u) (iv)

10.14 Given: $R_T = 1 \text{ k}\Omega$, $\alpha_T = 4.5\%/^\circ\text{C}$, heat dissipation 0.2°C/mW .

(a) For 0.1°C temperature rise the power should be 0.5 mW . Therefore,

$$I_{\max}^2 R_T = \frac{V_{\max}^2}{(2 \times 10^3)^2} (10^3) = 0.5 \times 10^{-3}$$

This yields $V_{\max} = 1.414 \text{ V}$

$$(b) \alpha_T = \frac{\Delta R_T / \Delta T}{R_T} = \frac{4.5}{100}$$

$$\therefore \Delta R_T = (0.045)(10^3)(1) = 45 \Omega$$

$$V_o = V \left(\frac{R_T + \Delta R_T}{R_T + \Delta R_T + 1000} - \frac{1}{2} \right) = 1.414 \left(\frac{1045}{2045} - \frac{1}{2} \right) = 15.6 \text{ mV}.$$

10.15 From the given data,

$$R_B = R_0(1 + \alpha\Delta T) \quad \text{where} \quad \Delta T = (T_2 - T_1)$$

Let $R_2/R_0 \equiv R$

$$(a) V_o = V \left(\frac{R_A}{R_A + R_2} - \frac{R_B}{R_B + R_3} \right)$$

Substituting $R_A = R_0$, $R_3 = R_2$ and abovementioned quantities, we get

$$V_o = V \left(\frac{1}{1 + R} - \frac{1}{1 + \frac{R}{1 + \alpha\Delta T}} \right)$$

(b) If $T_1 = 100^\circ\text{C}$ and T_2 varies from 50°C to 150°C , $\Delta T = \pm 50$. So, $\alpha\Delta T \ll 1$. Under that condition,

$$V_o = \left(\frac{1}{1 + R} - \frac{1}{1 + R(1 - \alpha\Delta T)} \right) = -V \frac{R\alpha\Delta T}{(1 + R)[1 + R(1 - \alpha\Delta T)]}$$

Thus,

$$|V_o| \propto \Delta T$$

10.16 (a) At 0°C , when the resistance of the RTD will be minimum, the current through it will be maximum. Then the current I through the RTD is

$$I = \frac{E}{10 + 10 + 100} = \frac{E}{120} \text{ A}$$

To keep the dissipated power within 1 mW ,

$$I^2 R_{\text{RTD}} = 1 \times 10^{-3}$$

This yields,

$$E_{\max} = \sqrt{\frac{1 \times 10^{-3}}{R_{\text{RTD}}}} \times 120 = 10^{-2} \times 120 = 1.2 \text{ V}$$

- 10.17 Since the ratio arm of the bridge has equal resistances, the bridge will remain balanced if $R_3 = R_4$. Now, given $R_{20} = 500 \Omega$, we have $500 = R_4(1 + 20 \times 0.005)$. This yields, $R_4 = 454.55 \Omega$. So, to null the bridge $R_3 = 454.55 \Omega$
- 10.18 $E_L = 1.47 \text{ mV} = E_{\text{actual}} / \left(1 + \frac{100}{1000}\right)$. This yields, $E_{\text{actual}} = 1.617 \text{ mV}$. This voltage is developed by 20 thermocouples connected in series. So, the emf generated by each thermocouple is $1.617 \div 20 = 0.0809 \text{ mV}$. Temperature of the first point is 25°C , the corresponding emf being 0.990 mV . By linear interpolation between 25°C and 27°C we find that the difference in emf corresponds to $\frac{2}{1.071 - 0.990} \times 0.0809 = 1.996^\circ\text{C}$.
- 10.19 This is basically one thermocouple arrangement having two junctions and its reference junction has been kept at 2°C rather than at 0°C . From the table, we find that the output corresponding to 2°C is $37 \mu\text{V}$. So, this emf is to be added to the indicated reading of $48 \mu\text{V}$ to get the correct temperature. We find from the table that $(37 + 48) = 85 \mu\text{V}$ corresponds to 50°C . Ans. (d)
- 10.20 (a) \rightarrow (h), (b) \rightarrow (e), (c) \rightarrow (f), (d) \rightarrow (g)

Chapter 11

- 11.5 (b) $v = 200 \text{ km/h} = 55.56 \text{ m/s}$
From Eq. (11.8) [see page 454],

$$\text{Differential pressure } p_{\text{imp}} - p_{\text{stat}} = \frac{v^2 \rho}{2} = \frac{(55.56)^2 (0.9)}{2} = 1.389 \text{ kPa}$$

- 11.10 (c) Flow rate = $\frac{10 \times 30 + 30 \times 20}{50} = 18 \text{ kg/min}$
Total mass flow = 900 kg

- 11.11 (a) (i) (b) (ii) (c) (ii) (d) (ii) (e) (ii) (f) (iii) (g) (i) [Because $Q \propto \sqrt{\Delta p}$]
(h) (iv) (i) (i)
(j) (iii) [From Eq. 11.11 [see page 456],

$$\frac{Q_1}{Q_2} = \sqrt{\frac{\rho_2}{\rho_1}}$$

With ρ_1 calibration, the flow rate is $2.2 \text{ m}^3/\text{min}$. Thus, we are given, $Q_1 = 2.2$, $\rho_1 = 1.2$ and $\rho_2 = 2$

$$\therefore Q_2 = \sqrt{\frac{1.2}{2}} (2.2) = 1.70$$

- (k) (i) (l) (iii) (m) (iii) (n) (ii) (o) (iv) (p) (i) (q) (i) (r) (ii)
(s) (iv)
- 11.16 (c) Given: $\Delta f = 805 \text{ cps}$, $l = 0.12 \text{ m}$ and $\theta = 45^\circ$. Therefore from Eq. (11.22) [see page 465], we get

$$v = \frac{\Delta f l}{2 \cos \theta} = \frac{(805)(0.12)}{2 \cos 45^\circ} = 68.3 \text{ m/s}$$

11.17 (b) Given: $k = 30/\text{gallon}$ and $Q_v = 1000$ gallons/min

From Eq. (11.18) [see page 462],

$$\omega = \frac{Q_v}{k} = \frac{1000}{30} = 33.3 \text{ rpm}$$

11.18 Given: $d = 5 \times 10^{-2}$ m, $B = 0.1$ T, $\mathcal{E} = 0.1 \times 10^{-3}$ V

$$(a) v = \frac{\mathcal{E}}{Bl} = \frac{0.1 \times 10^{-3}}{(0.1)(5 \times 10^{-2})} = 0.02 \text{ m/s.}$$

(b) Given: $\mathcal{E}_{p-p} = 9.4$ V. So, $\mathcal{E} = 4.7$ V. $Z = 250 \times 10^3 \Omega$, Gain = 1000

$$\text{Therefore, } \mathcal{E}_L = \frac{4.7}{1000} = \frac{\mathcal{E}_o}{1 + \frac{250 \times 10^3}{2.5 \times 10^6}}$$

This yields $\mathcal{E}_o = 5.17$ mV. Consequently,

$$v = \frac{\mathcal{E}_o}{Bl} = \frac{5.17 \times 10^{-3}}{(0.1)(5 \times 10^{-2})} = 1.034 \text{ m/s}$$

11.19 Given: $\rho = 1.03 \times 10^3$ kg/m³, $v = 1$ m/s

From Eq. (11.7) [see page 454], we have

$$\frac{p_{\text{stag}}}{\rho} = \frac{p_{\text{stat}}}{\rho} + \frac{v^2}{2}$$

$$(a) \text{ Therefore, the pressure due to the flow velocity} = \frac{v^2 \rho}{2} = \frac{1(1.03 \times 10^3)}{2} = 515 \text{ N/m}^2$$

$$(b) p_{\text{stat}} = p_{\text{stag}} - \frac{v^2 \rho}{2} = 10000 - 515 = 9485 \text{ N/m}^2$$

Now,

$$1 \text{ N/m}^2 = \frac{10^3}{13.6 \times 10^3 \times 9.81} \text{ mm of Hg}$$

So, the required pressure is 71.09 mm of Hg

11.20 Given: $l = 0.05$ m, $B = 0.1$ T, $v = 0.02$ m/s, $Z = 200 \times 10^3 \Omega$, $Z_L = 10^6 \Omega$ and gain $A = 1000$

$$\mathcal{E}_o = Blv = (0.1)(0.05)(0.02) = 0.1 \text{ mV}$$

$$\mathcal{E} = \mathcal{E}_o \times A = 0.1 \text{ V}$$

$$\text{Therefore, } \mathcal{E}_L = \frac{0.1}{1 + [(200 \times 10^3)/10^6]} = 0.083 \text{ V}$$

Ans. (d)

11.21 We know from Eq. (11.5) [see page 448] that $Q \propto \sqrt{\Delta p}$

Given: $\Delta p_1 = 30 \times 10^3$ Pa, $\Delta p_2 = 20 \times 10^3$ Pa and $Q_1 = 50$ L/s

$$\therefore Q_2 = Q_1 \sqrt{\frac{\Delta p_2}{\Delta p_1}} = 50 \sqrt{\frac{20}{30}} = 40.8 \text{ L/s}$$

Ans. (c)

Chapter 12

12.3 (b) (i) Range: 0 to 2200 mmwc; Span: 2200 mmwc.

(ii) Zero is not true zero. It is 220 mmwc. The true upper limit of the range should be 2420 mmwc.

12.12 Given: $l = 8 \text{ m}$, $\frac{d_1}{d_2} = 2$, $h = 7 \text{ m}$, $\varepsilon_R = 2.4$, $\varepsilon_0 = 8.85 \times 10^{-12} \text{ F/m}$

When empty,

$$\text{Capacitor } C_2 = \frac{2\pi\varepsilon_0 l}{\ln(d_1/d_2)} = \frac{2\pi(8)(8.85 \text{ pF})}{\ln 2} \cong 642 \text{ pF}$$

When filled with liquid up to 7 m,

$$C'_2 = C_{\text{air}} + C_{\text{liq}} = \frac{2\pi\varepsilon_0(1)}{\ln 2} + \frac{2\pi\varepsilon_R\varepsilon_0(7)}{\ln 2} = \frac{C_2}{8} [1 + 7\varepsilon_R] \cong 1428 \text{ pF}$$

(a) For null,

$$\frac{(15)100}{100 + 10000} = \frac{(15)C_2}{C_1 + C_2} = \frac{(15)(642)}{C_1 + 642}$$

yielding $C_1 = 64.2 \text{ nF}$

(b) Here, $V_o = 15 \left(\frac{1428}{1428 + 64200} - \frac{100}{100 + 10000} \right) \cong 178 \text{ mV}$.

12.13 (a) (iv) (b) (ii) [Actually, it is an inverse exponential relation. See Eq. (12.16) at page 511]

(c) (iii) [But also (i), because the surface area of the plates gets a little altered]

(d) (i)

Chapter 13

13.7 (a) (i) (b) (iv) (c) (iii) (d) (i) (e) (iv) (f) (iii) [Because, from Stokes' formula viscosity \propto (terminal velocity) $^{-1}$]

13.8 From Eq. (13.36) [see page 579],

$$\mu = \frac{\pi \Delta p a^4}{8Vl}$$

Given: $a_2 = 2a_1$, $\frac{\Delta p_1}{l_1} = \frac{\Delta p_2}{l_2}$, $V_1 = V_2$

$$\therefore \mu_2 = \mu_1 \left(\frac{a_2}{a_1} \right)^4 = 16\mu_1$$

Ans. (a)

13.9 Given: sensitivity = 59 mV/pH, $R_{\text{elec}} = 10^9 \Omega$, range = 0 – 14 pH. So, the electrode output range = 0 to $(59 \times 14 =) 826 \text{ mV}$

- (a) If V is the input to the buffer amplifier, then for 100 mV input to the recorder, we have

$$\frac{V \times 100 \Omega}{100 \Omega + 100 \Omega} = 100 \text{ mV}$$

This yields

$$V = 200 \text{ mV}$$

If R is the input impedance of the amplifier, we have then

$$\frac{826 \times R}{R + 10^9} = 200$$

giving $R \cong 320 \text{ M}\Omega$. Sensitivity = $200 \div 14 = 14.3 \text{ mV/pH}$

- (b) For pH = 7, the electrode output is $59 \times 7 = 413 \text{ mV}$. With the change in the electrode resistance, now the input voltage to the amplifier is

$$\frac{(320 \times 10^6)(413 \times 10^{-3})}{(320 \times 10^6) + (2 \times 10^9)} \cong 57 \text{ mV}$$

For an input of 200 mV, the pH is 14. So, for 57 mV input, the pH reading is $\frac{14 \times 57}{200} \cong 4$. The error is $\frac{7-4}{7} \times 100 \cong 43\%$

13.10 All data are in SI units except $r = 0.5 \pm 0.01 \text{ mm} = (0.5 \pm 0.01) \times 10^{-3} \text{ m}$

- (a) $\mu = \frac{\pi(0.5 \times 10^{-3})^4(50 \times 10^3)}{8(4 \times 10^{-7})(3)} = 1.023 \times 10^{-3} \text{ kg/(m-s)} = 10.23 \text{ centipoise}$.

- (b) We observe that only r and pressures have measurement errors. Now, since errors only add up, error in measuring $(p_1 - p_2) = \Delta p$ is $\delta(\Delta p) = 3 + 2 = 5 \text{ kPa}$. Thus, we have

$$\ln \mu = 4 \ln r + \ln(\Delta p)$$

On differentiation, it yields

$$\frac{\delta \mu}{\mu} = 4 \frac{\delta r}{r} + \frac{\delta(\Delta p)}{\Delta p} = 4 \frac{0.01}{0.5} + \frac{5}{50} = 0.18$$

It means that the limiting or absolute error is 18%

- (c) Considering only the variables having errors, we have $\mu \propto \pi r^4 p$, where for the sake of writing ease, we have replaced Δp with p . Now,

$$\frac{\partial \mu}{\partial r} = \frac{4\pi r^3 p}{8QL} = 8.1812$$

and

$$\frac{\partial \mu}{\partial p} = \frac{\pi r^4}{8QL} = 2.0453 \times 10^{-8}$$

Therefore, the root sum square error is

$$\begin{aligned} r_\mu &= \sqrt{\left(\frac{\partial\mu}{\partial r}\right)^2 (\delta r)^2 + \left(\frac{\partial\mu}{\partial p}\right)^2 (\delta p)^2} \\ &= \sqrt{(8.1812)^2 (0.01 \times 10^{-3})^2 + (2.0453 \times 10^{-8})^2 (5 \times 10^3)^2} \\ &= 1.31 \times 10^{-4} \text{ kg/(m-s)} \end{aligned}$$

The error is, therefore, $\frac{1.31 \times 10^{-4}}{1.023 \times 10^{-3}} \times 100 = 12.8\%$

Chapter 14

14.3 (a) (iv) (b) (i) (c) (i) (d) (iii) [Resolution = $\frac{\lambda_1 + \lambda_2}{2(\lambda_1 - \lambda_2)} = \frac{1200}{2 \times 0.02} = 3000$]

(e) (i) (f) (i) [From Eq. (14.52) at page 658, $A = -\log T = -\log 0.5 = 0.3$]

(g) (iv) (h) (ii) (i) (iii) (j) (iii) [From Eq. (14.52), $T = k \exp(-c)$.
 $\therefore \frac{T_2}{T_1} = \exp(-c_2 + c_1) = \exp(-c_1) \because c_2 = 2c_1$. Therefore, $T_2 = T_1 \exp(-c_1) = \frac{T_1^2}{k}$]

(k) (ii) [From Eq. 14.67 at page 698, $\lambda = \frac{12400}{40000} = 0.31 \text{ \AA}$] (l) (ii) [see Eq. (14.37) at page 644] (m) (iii) (n) (i) [Differentiate Eq. (14.68) of page 701 w.r.t. θ to obtain the result] (o) (iii) [From Eq. (14.47) at page 657,

$a = \frac{A}{bc} = \frac{0.6}{(2)(10^{-4})} = 3000 \text{ litre/mol/cm}$] (p) (i) (q) (ii) (r) (iii) [From Eq. (14.36) of page 644, when other parameters are constant, we have $l \propto 1/\sqrt{m}$]

(s) (i) [In Eq. (14.47), given $A = 0.01$, $a = 10^4 \text{ L/mol/cm}$ and $b = 1 \text{ cm}$.

$\therefore c = \frac{0.01}{(1)(10^4)} = 1 \mu\text{mol/L}$] (t) (iv) [From Eq. (14.47), $\frac{a_1 b_1 c_1}{a_2 b_2 c_2}$. Given: $c_1 = c$, $c_2 = 0.5c$, $b_1 = 4 \text{ cm}$, $b_2 = 1 \text{ cm}$, and $a_1 = a_2$, being the same analyte

$\therefore \frac{A_1}{A_2} = \frac{(4)(c)}{(1)(0.5c)} = 8$]

(u) (i) [From Eq. (12.16) at page 511, since I/I_0 is the same, we have $\mu_1 \rho_1 d_1 = \mu_2 \rho_2 d_2$. This gives,

$$d_2 = \frac{\mu_1 \rho_1 d_1}{\mu_2 \rho_2} = \frac{(0.02)(2699)(2.5)}{(0.075)(8960)} \cong 0.2 \text{ mm}]$$

14.4 (a) \rightarrow (e), (b) \rightarrow (c)

14.5 (a) \rightarrow (h), (b) \rightarrow (g), (c) \rightarrow (f), (d) \rightarrow (e)

14.6 (a) \rightarrow (f), (b) \rightarrow (h), (c) \rightarrow (e), (d) \rightarrow (g)

14.7 (a) \rightarrow (f), (b) \rightarrow (e), (c) \rightarrow (g), (d) \rightarrow (h)

14.8 (iii)

14.9 (iv)

14.10 (b) From Eq. (14.67) [see page 698],

$$\lambda_0 = \frac{12340}{100000} = 0.1234 \text{ \AA}$$

14.11 If 0.012 g of pure A produced a peak of area 15 cm², 0.12 g of pure A should have produced a peak of area $(0.12/0.012) \times 15 = 150 \text{ cm}^2$. But the sample has produced a peak of area 30 cm² only. So, the percentage of A in the sample is $(30/150) \times 100 = 20$

14.12 From the Duane-Hunt law [Eq. (14.67)], we know

$$\lambda_0 (\text{\AA}) = \frac{12340}{V(\text{volt})}$$

This yields
$$\frac{\lambda_1}{\lambda_2} = \frac{V_2}{V_1} = 1.5$$

Therefore,
$$\Delta\lambda = \lambda_1 - \lambda_2 = \lambda_1 \left(1 - \frac{1}{1.5}\right) = \frac{\lambda_1}{3} \quad (\text{i})$$

But,
$$\Delta\lambda = 26 \times 10^{-12} \text{ m} = 0.26 \text{ \AA}(\text{given}) \quad (\text{ii})$$

From Eqs. (i) and (ii), we get $\lambda_1 = 0.78 \text{ \AA}$. Substituting this value in Eq. (14.67), we get

$$V_1 = \frac{12340}{0.78} = 15820.5 \text{ V}$$

14.13 (b) (i) From Eqs. (14.47) and (14.52) [see Beer-Lambert law at page 657], we get

$$-\log T = abc$$

\therefore a and b are constant here, we have

Thus,
$$\frac{\log T_2}{\log T_1} = \frac{c_2}{c_1} = 2$$

$$\log T_2 = 2 \log T_1 = 2 \log(0.8) = -0.19382$$

$$\Rightarrow T_2 = 10^{-0.19382} = 0.64$$

So, the required transmittance is 64%

(ii) Here, T and a are constants and $c_2 = 2c_1$

So,

$$b_1 c_1 = b_2 c_2$$

$$\Rightarrow b_2 = \frac{b_1 c_1}{c_2} = \frac{b_1}{2} = 0.5 \text{ cm}$$

14.14 From Eq. (12.16),

$$I = I_0 \exp(-\mu \rho d)$$

Given: $\mu_{\text{Al}} = 3.48 \text{ cm}^2/\text{g}$, $\mu_{\text{Pb}} = 72 \text{ cm}^2/\text{g}$, $\rho_{\text{Al}} = 2.7 \text{ g/cm}^3$ and $\rho_{\text{Pb}} = 11.3 \text{ g/cm}^3$

(a) Here, $d = 2.6$ cm.

$$\therefore \frac{I}{I_0} = \exp[-(3.48)(2.7)(2.6)] = 2.457 \times 10^{-11}$$

(b) Here, $\frac{I}{I_0} = 2.457 \times 10^{-11}$.

$$\therefore \exp[-(72)(11.3)d] = 2.457 \times 10^{-11}$$

$$\text{Thus, } d = -\frac{\ln(2.457 \times 10^{-11})}{(72)(11.3)} = 0.03 \text{ cm}$$

Chapter 15

15.1 (a) (i) (b) (iii) (c) (ii)

Chapter 16

16.21 (a) (iv) (b) (i) (c) (iv) (d) (i) (e) (iv) [For 8-bit,

$$\text{Number of decision levels} = 2^8 - 1 = 255$$

$$\text{So, Quantum size } Q = \frac{1.275}{255} \cong 5 \text{ mV}$$

$$\therefore \text{Quantisation error} = \pm(Q/2) = \pm 2.5 \text{ mV}$$

(f) (iii)

$$(g) (ii) \quad [V_A = \frac{10X_C}{R + X_C} = \frac{10}{(R/X_C) + 1} = \frac{10}{2} = 5 \text{ V}]$$

$$V_B = \frac{10R}{R + R} = 5 \text{ V} \quad \therefore V_A - V_B = 0]$$

$$(h) (ii) \quad (i) (i) \quad (j) (i) \quad \left[\frac{1}{2^8 - 1} \times 100 = 0.39 \right]$$

$$(k) (iii) \quad \left[\frac{10}{2^n - 1} = 0.5 \times 10^{-3} \Rightarrow 2^n = 2001 \Rightarrow n = 11 \right]$$

(l) (iii) (m) (iii) (n) (i) (o) (iii) (p) (i) [Voltage V with wiper at half-way of the pot is 1 V. So, from $\frac{1}{10 \times 10^3} = -\frac{V_o}{15 \times 10^3}$ we have $V_o = -1.5 \text{ V}$] (q) (iv)

(r) (iv) (s) (i) (t) (i) (u) (i)

16.26 550 pulses in 100 ms

$$\therefore \text{Frequency} = 5500 \text{ Hz}$$

Input voltage: output frequency ratio = 0.2

$$\therefore \text{Input voltage} = 5500 \times 0.2 = 1100 \text{ V}$$

16.29 Assume it to be an 8-bit ADC.

| <i>Step</i> | <i>Setting</i> | <i>Calculation</i> | <i>Comparison with 2.875</i> | <i>Result</i> |
|-------------|----------------|--------------------------------------|------------------------------|---------------|
| I | $d_7 = 1$ | $5 \div 2 = 2.5$ | less | $d_7 = 1$ |
| II | $d_6 = 1$ | $2.5 + (5 \div 2^2) = 3.75$ | greater | $d_6 = 0$ |
| III | $d_5 = 1$ | $2.5 + (5 \div 2^3) = 3.125$ | greater | $d_5 = 0$ |
| IV | $d_4 = 1$ | $2.5 + (5 \div 2^4) = 2.8125$ | less | $d_4 = 1$ |
| V | $d_3 = 1$ | $2.8125 + (5 \div 2^5) = 2.96875$ | greater | $d_3 = 0$ |
| VI | $d_2 = 1$ | $2.8125 + (5 \div 2^6) = 2.890625$ | greater | $d_2 = 0$ |
| VII | $d_1 = 1$ | $2.8125 + (5 \div 2^7) = 2.8515625$ | less | $d_1 = 1$ |
| VIII | $d_0 = 1$ | $2.8515625 + (5 \div 2^8) = 2.87109$ | less | $d_0 = 1$ |

$$\text{Output} = 10010011 = \left(\frac{1}{2} + \frac{1}{16} + \frac{1}{128} + \frac{1}{256} \right) \times 5 = 2.87109375 \text{ V}$$

16.30 In the given circuit,

$$\frac{V_+}{100} + \frac{V_+ - 20}{10 + 0.05} = 0 \quad (\text{i})$$

$$\frac{V_- - V_o}{100} + \frac{V_- - 10}{10 + 0.5} = 0 \quad (\text{ii})$$

Eq. (i) yields $V_+ = 18.17 \text{ mV}$. \therefore for an ideal op-amp $V_- = V_+$, substituting this value in Eq. (ii), we get

$$\frac{18.17 - V_o}{100} + \frac{18.17 - 10}{10.5}$$

Solving this equation, we get $V_o = 95.98 \text{ mV}$

\therefore Ans. (b)

Chapter 17

17.4 Sensitivity = $10 \text{ mV}/10^4 = 1 \mu\text{V}$.

17.5 (a) $\pm 0.5\%$ of $5 \text{ V} = \pm 0.025 \text{ V}$

On a 3-digit display, resolution = 10^{-3} V

In 10 V range, the LSB value = 0.01 V

2 digits = $\pm 0.02 \text{ V}$

\therefore Total possible error = $\pm 0.045 \text{ V}$

(b) $\pm 0.5\%$ of $0.10 \text{ V} = \pm 0.0005 \text{ V}$

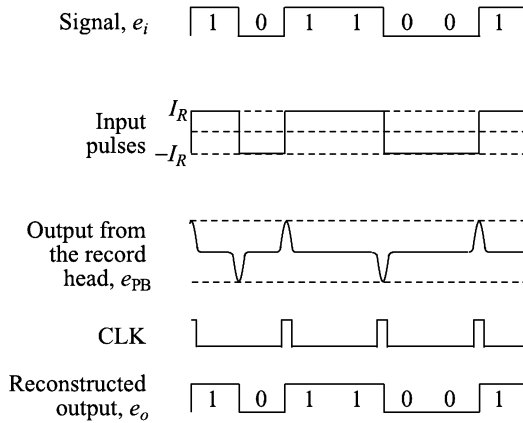
LSB value = 0.01 V

2 digits = $\pm 0.02 \text{ V}$

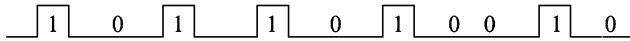
\therefore Total possible error = $\pm 0.0205 \text{ V}$

(c) % error = $(0.0205/0.10) \times 100 = 20.5$.

17.6



17.7



17.8 $\pm 0.5\%$ of 2 mA = ± 0.010 mA

5 counts in 20 mA range of $3\frac{1}{2}$ -digit display = $\pm 0.02 \times 5 = \pm 0.10$ mA

\therefore Total possible error = ± 0.11 mA

Worst case error = $\pm(0.11/2) \times 100 = \pm 5.5\%$

17.9 (a) (i) (b) (i)

17.10 Given: the number of bits $n = 12$, the conversion time $t_c = 5 \mu\text{s}$ and sequencing time $t_{\text{seq}} = 5 \mu\text{s}$.

(a) Quantum size $Q = \frac{1}{2^n} = \frac{1}{2^{12}} = 2.44 \times 10^{-4}$

So, Quantisation error $e_Q = \pm \frac{Q}{2} \times 100 = \pm 0.012\%$

Ans. (ii)

(b) Cut-off frequency $f_c = \frac{1}{t_c} = \frac{1}{5 \times 10^{-6}} = 200$ kHz

Ans. (iv)

(c) Sampling period $t_s = t_{\text{seq}} + t_c = 10 \mu\text{s}$

\therefore Sampling rate = $f_s = \frac{1}{t_s} = 100\text{k samples/s}$

Ans. (iii)

Index

- AAS, 677
- absolute pressure, 280
- absorbance, 658
- accelerometer
 - capacitive, 335
 - Hall effect, 337
 - inductive, 335
 - magnetoresistive, 337
 - MEMS, 337
 - pendulous integrating gyro, 338
 - piezoelectric, 336
 - piezoresistive, 337
 - resistive, 335
 - thermal, 337
- accuracy, 4, 821
- acquisition time, 797
- actuation depth, 507
- additivity, 858
- advance, 243
- AES, 707
- AFS, 681
- aliasing, 802
- alphanumeric display, 822
- alphatron, *see* ionisation gauge
- amplifier
 - charge mode, 144
 - voltage mode, 143
- anemometer
 - constant-current, 471
 - constant-temperature, 471
- angle of acceptance, 207
- anti-Stokes lines, 417
- aperture time, 797
- apodised grating, 271
- arc excitation, 671
- arithmetic mean, *see* mean
- asymmetry potential, 565
- atomic
 - absorption spectroscopy, 677
 - emission spectroscopy, 675
 - fluorescence spectroscopy, 681
- atomiser, 679
- Auger electron spectroscopy, 707
- availability, 70
 - five nines, 71
- backplane, 824
- balling, 536
- bandwidth, 106
- barkometer, 536
- bath tub curve, 71
- baume, 536
- Bayard-Alpert geometry, 316
- beaker wall effect, 548
- becquerel, 544
- Beer's law, 658
- Beer-Lambert law, 657
- bellows, 293
- belt weighing
 - continuous, 348
- Bernoulli distribution, 619
- Bernoulli's theorem, 447
- bi-colour level gauge, 492
- bimetals, 376
- bimorph, 147
- binomial distribution, 619
- biphase-mark, 844
- blackbody, 423
- bluff body, 467

- Bohr magneton, 691
 bolometer, 662
 bond strength, 656
 Bourdon gauge, 294
 Bragg
 and Pierce law, 706
 condition, 701
 Bremsstrahlung, 697
 bridge
 Blumlein, 747
 Callendar and Griffiths', 383
 current-sensitive, 251
 full, 251
 half, 250
 quarter, 250
 voltage-sensitive, 250
 Wheatstone, 248, 742
 brix, 536
 Bronsted acids, 633
 buffer solution, 563
 bus monitoring, 851
- Callendar and Griffiths' bridge, 383
 Callendar-Van Deusen relation, 385
 capacitive transducer, *see* transducer
 capacity factor, 623
 catalytic
 analyser, 615
 combustion analyser, 607
 cell constant, 545
 cermet, 175
 channeltron, 651
 charge
 coefficient, 135
 sensitivity, 135
 chemical
 ionisation, *see* ionisation
 seal, 321
 chi-square test, 49
 chirp, 271
 chromatogram, 617
 chromatography, 616
 closed-loop voltage gain, 757
 coherent detection technique, 273
 cold junction compensation, 409, 753
 combination electrode, 568
 common mode
 failure, 74
 rejection ratio, 764
 compliance voltage, 786
- Compton scattering, *see* scattering conductivity
 cells, 545
 measurement, 544
 confocal microscope, 204
 Connes advantage, 666
 consistency measurement, 586
 constantan, 243
 contact potential, 402
 convection gauge, *see* gauge
 Coriolis
 densitometer, *see* densitometer
 flowmeter, *see* flowmeter
 viscometer, *see* viscometer
 coupling factor, 137
 creep, 10
 critically-damped system, 100, 104
 cross-section ionisation detector, 631
 Curie
 law, 609
 temperature, 132
 temperature for PTC, *see* temperature
 curie, 544
 current loop, 4–20 mA, 785
 current measurement, 17
 curve fitting, 51
 cut-off frequency, 106
 cuvette, 682
 cyclotron frequency, *see* frequency
- Dall flow tube, 454
 Daly detector, 651
 damping ratio, 96
 Daniel cell, 553
 dB, 30
 dead zone, 9
 dead-weight gauge, 283
 decision level, 799
 degree of freedom, 50
 demodulation, 771
 densitometer
 bubbler, 537
 Coriolis, 538
 displacer and float type, 539
 DP cell, 538
 gamma-ray, 543
 ultrasonic, 543
 vibrating U-tube, 541
 weight-based U-tube, 542
 density
 compensation, 457
 measurement, 535

- derivative spectroscopy, 693
deuterium discharge lamp, 684
deviation, 38
 mean absolute, 39
 standard, 40
dew-point, 522
diaphragm, 291
 level indicator, 499
 protector, 322
diatomaceous earth, 628
differential mode gain, 764
diffusion current, 572
Dirac δ -function, 83, 863
direct current plasma, 672
discharge
 coefficient, 448
 ionisation detector, 631
dispersive infrared analysis, 660
dissolved oxygen analysis, 561
dot-matrix display, 825
double density recording, 847
double-focussing analyser, 640
DPI, 840
drag-cup tachometer, 362
drift, 9, 756
dropping mercury electrode, 571
dual scanning, 834
Duane-Hunt law, 698
dummy gauge, 258
Dunmore cell, 525
dust loading, 596, 597
dynamic
 characteristics, 80
 compensation, 779
dynamometer, 351
 absorption type, 352
 driving type, 352
 transmission type, 355

echo ranging, 211
ECN, 630
eddy-current, 223
effective carbon number, 630
EFPI
 dual wavelength method, 267
 QPS, 266
 white light interferometry, 267
electric susceptibility, 135
electrochemical analyser, 616
electrode
 Ag/AgCl, 564
 antimony, 566
 calomel, 564
 glass, 565
 quinhydrone, 566
electrodeless
 conductivity measurement, 549
 discharge lamp, 678
electromotive force, 551
electron
 capture detector, 630
 impact ionisation, 633
 multiplier tube, 651
 spectroscopy for chemical analysis, 713
 spin resonance spectroscopy, 690
electrospray ionisation, 637
electrothermal excitation, 670
elevation head, 447
elution time, 627
emissivity, 425
encoder, 216
 absolute, 217
 incremental, 217
 tachometer, 217
end cooling, 605
environmental error, *scc* error
equalisation, 841
error, 29
 environmental, 23
 gross, 12
 guarantee, 31
 human, 11
 instrumental, 12
 interference, 12
 limiting, 31
 misuse, 12
 multiple earths, 14
 observational, 12
 probable, 31, 45
 random, 24
 systematic, 12
 voltage, 796
ESCA, 713
excimer laser, 272
expansion ratio, 449

F2F, 845
Fabry-Pérot interferometer, *scc* interferometer

-
- Faraday
 - cup, 650
 - effect, 117
 - fast atom bombardment ionisation, 635
 - Fellgett advantage, 666
 - ferry, 243
 - fibre
 - Bragg grating sensor, 269
 - Brillouin scattering sensor, 272
 - fibre-optic
 - level indicator, *see* level indicator
 - pyrometer, 434
 - strain gauge, 264
 - Fick's law, 573
 - fidelity, 94
 - field lock, 688
 - fieldbus, 792
 - filled system thermometer, *see* thermometer
 - filter
 - band-pass, 774
 - band-reject, 774
 - high-pass, 774
 - low-pass, 773
 - notch, 779
 - twin-T, 779
 - finesse, 267
 - fingerprint
 - pattern, 705
 - region, 657
 - spectrum, 698
 - flame
 - excitation, 669
 - ionisation detector, 629
 - float viscometer, 585
 - floats, 493
 - flow
 - coefficient, 448
 - nozzle, 453
 - flowmeter
 - anemometer, 470
 - Coriolis, 473
 - Doppler frequency shift, 463
 - electromagnetic, 459
 - impact, 478
 - lobed impeller, 481
 - nutating disc, 479
 - open channel, 482
 - positive displacement, 479
 - sliding vane, 480
 - thermal, 477
 - transit time, 464
 - turbine, 461
 - ultrasonic, 463
 - vortex shedding, 466
 - flume, 483
 - fluorescence, 681
 - force-summing device, 280, 291
 - formazin turbidity unit, 591
 - fotonic sensor, 206
 - four-wire
 - connection of RTD, 384
 - transmitter, *see* transmitter
 - free induction decay, 690
 - frequency
 - cyclotron, 648
 - Larmor, 686
 - magnetron, 648
 - resonant, 106
 - to voltage conversion, 761
 - frequency domain analysis, 81
 - fringing, 190
 - front slope region, 209
 - FTIR, 662
 - advantages, 665
 - galvanic
 - analyser, 613
 - cells, 551
 - gauge
 - convectron, 314
 - glass, 491
 - ionisation, 314
 - Knudsen, 309
 - McLeod, 307
 - molecular momentum, 311
 - Pirani, 313
 - pressure, 281
 - thermocouple, 312
 - viscous friction, 311
 - Gaussian distribution, *see* normal distribution
 - Geiger-Müller counter, 714
 - gel-permeation chromatography, 618
 - geometric mean, *see* mean
 - globar, 660
 - Golay pneumatic cell, 661
 - Gray code, 218
 - grey body, 426
 - gross error, *see* error
 - guard ring, 190
 - gyromagnetic ratio, 685

-
- Hagen-Poiseuille formula, 579
 - half-wave potential, 573
 - Hall effect, 121
 - harmonic mean, *see* mean
 - HART, 791
 - hazard rate, 71
 - heat of reaction method, 607
 - Henry's law, 526
 - Hersch cell, 614
 - HETP, 619
 - histogram, 35
 - hollow cathode lamp, 678
 - homogeneity, 859
 - Hooke's law, 137, 238
 - hopcalite, 615
 - HPLC, 631
 - human error, *see* error
 - humidity measurement, 522
 - hydrometer, 536
 - hydrostatic tank gauging, 500
 - hygrometer
 - capacitive, 530
 - electrolytic, 528
 - hair, 524
 - impedance, 531
 - infrared absorption, 532
 - microwave absorption, 533
 - piezoelectric, 531
 - wet and dry bulb, 523
 - hyperfine coupling, 695
 - hysteresis, 9, 324

 - IC temperature sensor, 411
 - IEEE 1451, 167
 - Ilkovic equation, 573
 - impulse input, *see* input
 - inductively coupled plasma, 673
 - inherent shortcomings, 12
 - input
 - impulse, 83
 - ramp, 83
 - step, 83
 - instrument
 - first order, 86
 - second order, 96
 - zero order, 85
 - instrumental error, *see* error
 - instrumentation amplifier, 765
 - interdigital transducer, 153
 - interference error, *see* error
 - interferogram, 664
 - interferometer
 - Fabry-Pérot, 264
 - Michelson, 663
 - micro laser, 205
 - Sagnac, 417
 - interlaced scan, 833
 - intrinsically safe barrier, 738
 - inverse
 - piezoelectric effect, 131
 - transducer, *see* transducer
 - ion cyclotron resonance, 648
 - ionisation
 - chamber, 714
 - chemical, 633
 - thermospray, 638
 - ionisation gauge
 - alphatron, 317
 - cold cathode, 316
 - hot cathode, 314
 - ISFET, 569
 - isoelastic, 256
 - ITS-90, 374, 375

 - Jackson turbidity unit, 591
 - Jaquinot advantage, 666
 - Josephson effect, 117
 - Joule effect, 118

 - karma, 243
 - katharometer, 628
 - kinematic viscosity, 578
 - King's law, 471
 - Kingdon's trap, 649
 - kisselguhr, 591
 - Knudsen gauge, *see* gauge

 - Lambert's law, 658
 - laminar flow, 446
 - Landé g -factor, 691
 - Laplace transform, 80, 861
 - Larmor frequency, *see* frequency
 - laser transducer, *see* transducer
 - laser-induced excitation, 674
 - law of
 - intermediate metals, 403
 - intermediate temperatures, 403
 - LCD, 823

- least squares method, *see* method
LED, 822
level indicator
 air bubbler, 502
 capacitive, 508
 differential pressure, 500
 displacer, 497
 fibre-optic, 505
 gamma-ray, 511
 inductive, 507
 magnetostrictive, 496
 microwave, 511
 optical, 503
 resistive, 505
 ultrasonic, 510
linearisation, 56
 by bridge, 751
 thermistor, 395
linearity, 6, 176
liquid junction, 565
Littrow mounting, 660
load cell, 339
 fibre-optic, 346
 hydraulic, 343
 magnetoelastic, 344
 piezoelectric, 345
 pneumatic, 343
 reluctance-based, 344
 resonant wire, 346
 strain gauge, 341
loading effect, 16
 potentiometer, 175
lock-in amplifier, 793
LVDT, 179
 open wiring, 182
 ratiometric wiring, 183

magnetic
 quantum number, 685
 sector-type mass analyser, 639
 wind, 609
magnetoelastic effects, 117
magnetoelasticity, 129
magnetoelastic transducer, *see* transducer
magnetostriction, 118
 positive, 119
 negative, 119
magnetostrictive level indicator, *see* level indicator
magnetron frequency, *see* frequency
manometer, 285
 double bell, 290
 inclined-limb, 286
 inverted bell, 289
 ring-balance, 288
Martin and Synge, 618
mass
 absorption coefficient, 706
 resolving power, 641
 spectrometer, 632
Mathieu equations, 646
matrix
 active, 834
 assisted laser desorption, 636
 passive, 834
Matthiessen-Herzog geometry, 641
Matteucci effect, 119
maximum power transfer theorem, 21
McLeod gauge, *see* gauge
MDT, 69
mean
 arithmetic, 36
 failure rate, 66
 geometric, 37
 harmonic, 37
median, 36
mercury vapour analyser, 616
method
 of equal effects, 34
 of extended differences, 53
 of least squares, 54
 of sequential differences, 52
MFM, 845
microchannel plate detector, 652
microphone detector, 661
microstrain, 240
Miller sweep, 761
minimum ignition energy, 726
misuse error, *see* error
mobile phase, 616
mode, 36
modulation
 amplitude, 767
 frequency, 770
 phase, 771
moisture analyser
 neutron backscatter, 533
 NMR, 535
molar absorptivity, 683
molecular momentum gauge, *see* gauge

- monochromator, 680, 701
Moseley's law, 705
MTBF, 69
MTTF, 66
multijunction cell, 163
multimode optical fibre, 207
multimorph, 148
multiple earths error, *see* error
- NDIR, 659
nebuliser
 crossflow, 667
 ultrasonic, 667
nephelometer, 594
nephelometric turbidity unit, 591
nephelometry, 597
Nernst
 equation, 554, 612
 filament, 660
 layer, 573
Newtonian fluids, 578
nichrome, 243
 wire, 661
non-dispersive infrared analysis, 659
nonlinearity, 850
normal
 distribution, 41
 properties, 43
 equation, 54
Norton theorem, 12
nozzle-flapper transducer, 170
NRZ, 844
NRZ-M, 844
nuclear magnetic resonance, 684
null offset, 129
numerical aperture, 207
Nyquist theorem, 802
- observational error, *see* error
OES, 675
OP-TDLAS, 598
opacity measurement, 595
open-loop dc gain, 757
optical
 tachometer, *see* tachometer
 transducers, *see* transducer
opto-isolator, 788
orbitrap, 649
- order of a system, 84
orifice plate, 449
ORP, 551
 measurement, 555
Ostwald viscometer, *see* viscometer
overdamped system, 99
overpressure protector, 318
oxidation potential, 552
oxygen analysis, 607
- paramagnetic oxygen analyser, 608
partition
 coefficient, 622
 ratio, 623
passive
 transducer, *see* transducer
peak
 overshoot, 102
 time, 101
Peltier
 coefficient, 400
 effect, 400, 527
permeation dryer, 716
Petry's chamber, 671
pH measurement, 567
phase-locked loop, 796
photoconductivity, 160
photodiode, 164
photoelectric emission, 707
photoelectricity, 155
photoemission, 155
photopic region, 596
phototransistor, 164
phototube, 158
photovoltaic cell, 162
piezoelectric
 actuator, 146
 coefficients, 134
 material, 131
piezoelectricity, 130
piezoresistivity, 150, 245
piezoresistor, 151
Pirani gauge, *see* gauge
Pitot tube, 454
pixel, 832
Planck's law, 424
plasma desorption ionisation, 634
plate theory, 618
platform scale, 347

- PLL, 796
pneumatic instrumentation, 731
poise, 578
poiseuille, 578
Poisson
 arrangement, 341
 ratio, 239
polarisation density, 135
polarographic analyser, 614
polarography, 571
poling, 132
polychromator, 675
Pope cell, 525
potentiometer, 85, 173
precision, 5
pressductor, 344
pressure
 head, 447
 regulators, 733
 units, 282
probable error, *see* error
 of combinations, 46
 of the mean, 45
proof
 mass, 333
 pressure, 324
proving ring, 340
proximity sensor, 220
 capacitive, 225
 Hall effect, 228
 inductive, 221
 optical, 229
 ultrasonic, 231
PRT
 thin film, 381
 wire-wound, 381
pulsation damper, 321
Purnell equation, 624
pyroelectricity, 134
pyrometer, 421
 broadband, 428
 fibre-optic, 434
 narrow-band, 429
 ratio, 433
PZT, 131
- quadratic spline, 61
quadrature
 detector, 689
 encoder, 368
 output, 368
- quadrupole
 ion trap, 646
 mass analyser, 642
quantisation, 798
 error, 800
quantum
 efficiency, 715
 yield, 158
quevenue, 536
quinhydrone electrode, *see* electrode
- radiation dose, 544
Raman scattering, *see* scattering
ramp input, *see* input
range, 30
 dynamic, 30
rangeability, 483
raster scan, 831
rate theory, 626
Rayleigh scattering, *see* scattering
RB, 844
reaction quotient, 554
reactivation point, 324
redox potential, 553
reduced mass, 655
reduction potential, 552
redundancy, 73
Redwood viscometer, *see* viscometer
reference reflection, 265
reflectron mass analyser, 645
regression analysis, 54
relative
 humidity, 522
 sensitivity factor, 713
relay amplifier, 732
reliability, 65
reluctance
 variation of, 221
repeatability, 9
reproducibility, 9
resolution, 10, 624, 799, 821, 850
resonant
 frequency, *see* frequency
 peak, 106
response
 dynamic, 98
 frequency, 93, 104, 198
 impulse, 93, 104
 ramp, 90, 103
 speed of, 95
 step, 89, 99

- responsivity, 159
retention
 time, 621
 volume, 622
Reynolds number, 446
richter, 536
Ringlemann cards, 597
rise time, 101
root mean square, 38
root-sum square formula, 46
rotameter, 455
rotational viscometer, *see* viscometer
RTD, 380
RVDT, 185
- saddle coil, 688
Sagnac interferometer, *see* interferometer
salt bridge, 564
Saybolt viscometer, *see* viscometer
scanning mass analyser, 638
scattering
 Compton, 696
 Raman, 416
 Rayleigh, 596
 Thomson, 696
scintillation counter, 715
Secchi disc, 592
Seebeck effect, 400
seismic mass, 333
selectivity factor, 624
self-heating, 382
sensing reflection, 265
sensitivity, 6, 176, 821
 cross or secondary, 129
 static, 96
servomechanism, 186
settling time, 102, 851
seven-segment display, 826
shaft encoder, 367
shedder bar, 467
shim coil, 688
shock, 330
side frequency, 767
siemens, 544
sight glass, 491
signal transmission, 785
signature curve, 469
significant figures, 5
sikes, 536
silistor, 391
single density recording, 847
siphon, 321
skin depth, 225
skin effect, 117
sling psychrometer, 523
smart
 sensor, 166
 transmitter, 790
smoke detector, 317, 715
snubber, 321
solar cell, 162
solid state detector, 662
solution-conductivity cell, 525
span, 30
spark excitation, 671
spectral response, 159
spectrum
 rotational, 654
 vibrational, 656
sphygmomanometer, 297
spline interpolation, 60
spring-balance displacer, 498
sputtering, 381
stagnation pressure, 454
standard
 electrode potential, 552
 hydrogen electrode, 563
standard deviation, *see* deviation
 of the mean, 45
standard electrode potential, 552
static error, 11
stationary phase, 616
statistic, 35
steelyard, 347
Stefan-Boltzmann law, 425
Steinhart-Hart equation, 393
step
 index fibre, 207
step input, *see* input
stiffness constant, 97
Stokes, 578
 formula, 579
 lines, 416
stopping potential, 157
strain
 lateral, 239
 longitudinal, 238
strain gauge
 calibration, 262
 foil type, 244
 semiconductor, 245
 wire-wound, 243
stroboscope, 366

- Strouhal number, 467
successive approximation, 386
supervisory control and data acquisition, 848
surface acoustic wave, 153
switch
 level, 514
 magnetic, 515
 thermal, 516
 vibrating, 518
 pressure, 322
 temperature, *see* temperature
synchro, 185
synchrotron radiation, 700
systematic error, *see* error
- tachogenerator, 360
tachometer, 359
 digital, 367
 eddy-current, 362
 fly-ball, 360
 Hall effect based, 365
 inductive, 364
 optical, 365
tantalum “getter”, 678
temperature
 compensation, 258
 Curie for PTC, 391
 scale, 372
 switch, 391
 transition, 391
TEOM, 599
Terfenol-D, 120
thermal detector, 661
thermistor, 390
 NTC, 392
 PTC, 391
thermo-magnetic analyser, 610
thermocouple, 88, 405
 gauge, *see* gauge
thermoelectricity, 399
thermometer
 fibre-optic, 413
 filled system, 377
 liquid-in-glass, 377
 mercury-in-glass, 87
 platinum resistance, 380
 quartz, 419
 SAW, 420
 ultrasonic, 421
thermospray ionisation, *see* ionisation
- thermowells, 435
Thevenin
 equivalent, 742
 theorem, 12
Thompson scattering, *see* scattering
Thomson effect, 117, 400, 401
three-wire
 connection of RTD, 383
 transmitter, *see* transmitter
threshold, 10
time constant, 86, 93
time domain analysis, 81
time-domain reflectometry, 535
 optical, 417
time-of-flight mass analyser, 644
tolerance, 49
toner, 840
toroidal conductivity measurement, *see* electrodeless
torque meter, 357
torque-tube
 displacer, 539
 level indicator, 498
Torr, 306
torsion meter, 356
total dissolved solids, 544, 592
tralles, 536
transducer
 active, 114
 capacitive, 187, 300
 digital, 115
 displacement, 216
 fibre-optic, 206
 Hall effect, 303
 inductive, 299
 inverse, 116
 laser, 200
 magnetoresistive, 215
 optical, 200
 passive, 114
 photoelectric, 301
 piezoelectric, 302
 primary, 116
 push-pull, 746
 resistive, 297
 secondary, 116, 296
 selection, 165
 surface acoustic wave, 305
 ultrasonic, 211
 vibrating element, 304
transfer function, 80
 properties, 81

- transformer isolated barrier, 739
transition
 energy, 685
 temperature, *scc* temperature
transmittance, 658
transmitter
 force-balance, 734
 four-wire, 787
 pneumatic, 733
 standards, 786
 three-wire, 788
 torque-balance, 734
 two-wire, 787
transportation lag, 735
trapped-ion mass analyser, 646
triboelectricity, 600
turbidity, 588
 attenuation coefficient, 593
 measurement, 588
 tube, 593
turbulent flow, 446
turndown, 483
twaddell, 536
two-wire
 connection of RTD, 383
 transmitter, *scc* transmitter
Tyndall effect, 589
- U-length, 436
ultrasonic
 thermometer, 420
 transducer, *scc* transducer
unavailability, 71
unbiased estimator, 40
underdamped system, 100, 104
unimorph, 147
unreliability, 65
UV-visible absorption spectroscopy, 682
- vacuum
 measurement, 306
 pressure, 282
valve manifold, 319
van Deemter equation, 626
variance, 39
vector scan, 833
velocity
 head, 447
 of approach factor, 448
vena contracta, 450
- Venturi tube, 451
video graphics array, 835
Villari effect, 118
viscometer
 capillary, 580
 Coriolis, 585
 Ostwald, 580
 Redwood, 582
 rotational, 584
 Saybolt, 582
viscosity measurement, 577
viscous
 drag, 577
 friction gauge, *scc* gauge
voltage
 constant, 136
 controlled oscillator, 771
 measurement, 16
 sensitivity, 136
Voltaic cell, 551, 553
volumetric phase ratio, 623
von Kármán
 trail, 436
 vortex street, 467
vortex shedding, 467
voting systems, 74
- wave number, 653
weigh-feeder, 349
weighing
 weigh-feeder, *scc* weigh-feeder
weir, 482
Weiss domain, 132
wet leg, 502
Wiedemann effect, 119
Wien's displacement law, 424
work function, 157
- X-ray
 fluorescence spectroscopy, 710
 limit, 315
 methods, 695
- Zener diode barrier, 738
zero
 elevation, 128, 502
 suppression, 501
zero point, 217
zirconia fuel cell, 611

FOURTH EDITION

INTRODUCTION TO MEASUREMENTS AND INSTRUMENTATION

ARUN K. GHOSH

The fourth edition of this highly readable and well-received book presents the subject of measurement and instrumentation systems as an integrated and coherent text suitable for a one-semester course for undergraduate students of Instrumentation Engineering, as well as for instrumentation course/paper for Electrical/Electronics disciplines.

Modern scientific world requires an increasing number of complex measurements and instruments. The subject matter of this well-planned text is designed to ensure that the students gain a thorough understanding of the concepts and principles of measurement of physical quantities and the related transducers and instruments. This edition retains all the features of its previous editions *viz.* plenty of worked-out examples, review questions culled from examination papers of various universities for practice and the solutions to numerical problems and other additional information in appendices.

NEW TO THIS EDITION

Besides the inclusion of a new chapter on *Hazardous Areas and Instrumentation* (Chapter 15), various new sections have been added and existing sections modified in the following chapters:

- Chapter 3 *Linearisation and Spline interpolation*
- Chapter 5 Classifications of transducers, Hall effect, Piezoresistivity, Surface acoustic waves, Optical effects (This chapter has been thoroughly modified)
- Chapter 6 *Proximity sensors*
- Chapter 8 *Hall effect and Saw transducers*
- Chapter 9 *Proving ring, Prony brake, Industrial weighing systems, Tachometers*
- Chapter 10 *ITS-90, SAW thermometer*
- Chapter 12 *Glass gauge, Level switches, Zero suppression and Zero elevation, Level switches*
- Chapter 13 *The section on ISFET has been modified substantially*

THE AUTHOR

ARUN K. GHOSH (PhD) is Visiting Professor, Sir J.C. Bose School of Engineering, Hooghly. He is former Professor of Applied Electronics and Instrumentation, Guru Nanak Institute of Technology, Kolkata. He has also served as Head, Instrumentation Centre, University of Kalyani and Principal, Murshidabad College of Engineering and Technology, Berhampore and Bengal College of Engineering and Technology, Durgapur. Dr. Ghosh received his doctorate in physics from Calcutta University and had his post-doctoral assignment at Rice University Houston (Texas). He has over 27 years of teaching experience and has published 35 research papers in journals of international repute.

You may also be interested in

Virtual Instrumentation Using LabVIEW, Jovitha Jerome

Microprocessor-Based Agri Instrumentation, Krishna Kant

Power Plant Instrumentation, K. Krishnaswamy and M. Ponni Bala

Sensors and Transducers, 2nd ed., D. Patranabis

Instrumentation and Control, D. Patranabis

ISBN: 978-81-203-4625-3



www.phindia.com