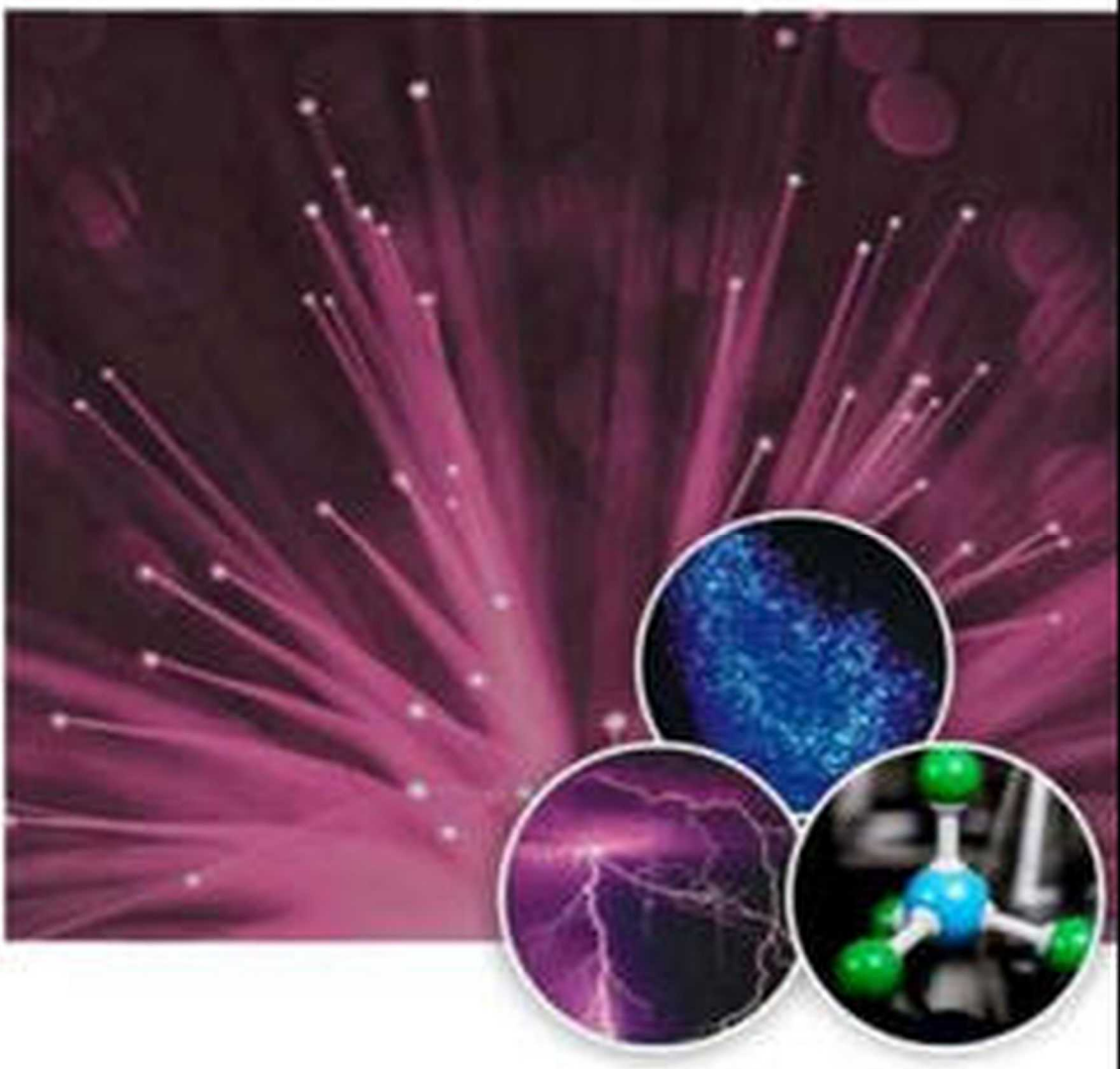


ENGINEERING PHYSICS

V Rajendran

The McGraw·Hill Companies

Engineering Physics

About The Author

V Rajendran, FUSI, FASI, FInstP**,** is the Director, Research and Developm ent, KSR Group of Institutions and Centre for Nano Science and Technology (CNST), K S Rangasamy College of Technology, Tamil Nadu, India. He holds an MTech in Nano Science and Technology from Anna University. He obtained a PhD in Ultrasonics from Annamalai University, and has been pivotal in developing the state-of-the-art role Centre for Nano Science and Technology Research, KSRCT at par with international standards, in coordination with the science and technology faculty. He has made enormous and innovative contributions to materials characterisation using online ultrasonic techniques and other conventional techniques.

Dr Rajendran has visited many countries including USA, West Indies, Germany, Italy, France, Sweden, Finland, Singapore, Hong Kong and Taiwan as visiting professor, guest scientist to various research laboratories, universities and conferences; and has received many laurels and awards such as National NDT Man of the Year 2004, TANSA Award, Outstanding Organiser Award, DAAD from Germany, INSA, TNSCST Young Scientist, and DAE /BRNS Visiting Scientist. Under his able guidance, five scholars have completed their doctoral degrees and 12 scholars are pursuing their PhD degrees. An enthusiastic academician, he has organised three national-level conferences on Ultrasonics, Acoustics and Instrumentation, one international conference on Nanomaterials and several workshops towards fulfilling his vision of imparting knowledge to budding research scholars. He has also attended several national and international conferences both in India and abroad. He has to his credit more than 100 research papers in reputed international journals, 65 papers in conference proceedings, 16 refereed books, 11 edited proceedings, 2 research and development books and 11 patents.

Dr Rajendran is also involved in several collaborative research projects with many leading research laboratories, both in India and abroad. His passion towards research led him to be the Principal Investigator of four completed, sponsored research projects with a fund involvement to the tune of INR 17 million. Presently, sponsored research projects to the tune of INR 15 million are under progress with sponsorship from Nanomission DST, DRDO, IGCAR, etc. Presently, he is involved in research including nano metal oxides for different industrial applications, nano silica from rice husk for applications in textiles and biomedicine, nano bioactive glasses and developing nanoparticles from herbs for biomedical applications.

He is the chairperson of the forthcoming international conference on nanomaterials and nanotechnology (NANO 2010) which will be held during December 13–16, 2010. His memberships include fellowships and life memberships of many scientific societies both in India and abroad. He is a member in the Board of Studies and Doctoral Committee in several universities (including Anna University). Also, he is in the peer review group for many international and national refereed journals apart form being the President of Acoustical Society of India.

The McGraw-Hill Companies

Engineering Physics

V Rajendran
Director—Research and Development
Centre for Nano Science and Technology
K S R College of Technology
and
K S R Group of Institutions
Tiruchengode, Tamil Nadu

Tata McGraw-Hill Publishing Company Limited
NEW DELHI

McGraw-Hill Offices
New Delhi New York St Louis San Francisco Auckland Bogotá Caracas
Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal
San Juan Santiago Singapore Sydney Tokyo Toronto

Tata McGraw-Hill

Published by Tata McGraw-Hill Publishing Company Limited,
7 West Patel Nagar, New Delhi 110 008.

Copyright © 2011, by Tata McGraw-Hill Publishing Company Limited.
No part of this publication may be reproduced or distributed in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise or stored in a database or retrieval system without the prior written permission of the publishers. The program listings (if any) may be entered, stored and executed in a computer system, but they may not be reproduced for publication.

This edition can be exported from India only by the publishers,
Tata McGraw-Hill Publishing Company Limited.

ISBN (13 digits): 978-0-07-107014-0
ISBN (10 digits): 0-07-107014-1

Vice President and Managing Director—MHE, Asia-Pacific Region: *Ajay Shukla*

Head—Higher Education Publishing and Marketing: *Vibha Mahajan*
Manager—Sponsoring (SEM & Tech Ed.): *Shalini Jha*
Assoc. Sponsoring Editor: *Suman Sen*
Executive—Editorial Services: *Sohini Mukherjee*
Sr Production Manager: *Satinder S Baveja*

Deputy Marketing Manager (SEM & Tech. Ed.): *Biju Ganesan*

General Manager—Production: *Rajender P Ghansela*
Assistant General Manager—Production: *B L Dogra*

Information contained in this work has been obtained by Tata McGraw-Hill, from sources believed to be reliable. However, neither Tata McGraw-Hill nor its authors guarantee the accuracy or completeness of any information published herein, and neither Tata McGraw-Hill nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that Tata McGraw-Hill and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

Typeset at Durga Print Media, K-II/553A, Sangam Vihar, New Delhi 110 062, text and printed at Avon Printers, Plot No. 16, Main Loni Road, Jawahar Nagar Industrial Area, Shahdara, Delhi-94.

Cover: SDR Printers

RBLCRRBFRQRBQ

Preface

Engineering Physics, as a discipline, has always been regarded as a fascinating subject by aspiring engineering and technology students. Due to the shortage of good textbooks on the subject, a need was felt for one that fulfills the requirements of students in a way they can relate to, as well as remains within the scope of their prescribed syllabi. This textbook covers the complete syllabus of Engineering Physics for different branches of engineering and technology (BE/BTech), of almost all technical universities of India. The content of the book is both simple and coherent.

All along the text, an attempt has been made to emphasise on engineering applications in every chapter. Solved problems, short questions, descriptive questions, objective questions, objectives and real-time applications are included to expedite the learning process of students of all categories.

Key Features

- Complete coverage of all major topics of Engineering Physics
- Lucid style of presentation
- Rich pedagogy:
  - ♣ Chapter Objectives
  - ♣ Chapter Summary
  - ♣ Key Points to Remember
  - ♣ 190 Solved Problems
  - ♣ 710 Short Questions
  - ♣ 187 Descriptive Questions
  - ♣ 80 Exercises
  - ♣ 585 Objective Type Questions

Chapter Organisation

The book covers the 27 topics such as Elasticity, Viscosity, Vacuum Technique, Acoustics, Ultrasonics, Non-Destructive Testing, Interference, Polarisation, Photoelasticity, Optical and Other Instruments, Laser, Fiber Optics, Crystal Structure, Waves and Particles, Classification of Solids, Electron Theory of Solids, Statistics and Band Theory of Solids, Transport properties of Semiconductors, Superconducting Materials, Magnetic Materials, Dielectric Materials, Shape Memory Alloys, Nonlinear Materials, Metallic Glasses, Biomaterials, Nanomaterials Synthesis and Nanodevices, all of which will be useful to students and faculty for enhancing their knowledge and skills.

Elasticity, viscosity and vacuum techniques are thoroughly explained in **chapters 1, 2 and 3** respectively. **Chapters 4 and 5** deal with the principles, production, properties and applications of acoustics and ultrasonics respectively. Material testing like non-destructive testing has been discussed in **Chapter 6. Chapters 7, 8 and 9** cover respectively interference, polarisation and photoelasticity. **Chapter 10** deals with optical and other instruments.

Laser, fiber optics and crystal structure are presented in **chapters 11 to 13** respectively. **Chapter 14** discusses waves and particles. Semiconductors, electron theory of solids, statistics and band theory of solids are covered respectively in **chapters 15 to 17**, while **Chapter 18** deals with transport properties of semiconductors.

Superconducting materials, magnetic materials, dielectric materials have been discussed in **chapters 19 to 21. Chapter 22** discusses shape memory alloys and their properties. The newer materials like nonlinear materials, metallic glasses, biomaterials, nanomaterials synthesis and nanodevices are discussed in **chapters 23 to 27.**

The McGraw·Hill Companies

Acknowledgements

My heartfelt thanks are due to Dr W Selvamurthy, CC&RD(LS), DRDO for his foreword message. I am thankful to Dr K S Rangasamy (Chairman) and Shri R Srinivasan (Secretary) and all my colleagues, particularly Dr K Thyagarajah, Principal, K S Rangasamy College of Technology for facilitating a wonderful environment and for the encouragement showered on me in bringing out this book. I would also like to thank Mr K Sakthipandi and Mr R Yuvakkumar for their contribution in typesetting and proof corrections.

A special note of appreciation goes to all those reviewers who devoted time in going through the manuscript and giving constructive suggestions. Their names are given below:

T Mahalingam	*Maharaja Engineering College, Coimbatore*
K Sakthivel	*Coimbatore Institute of Technology, Coimbatore*
R G Sethuraman	*Kumaraguru College of Technology, Coimbatore*
S Gopal	*Kumaraguru College of Technology, Coimbatore*
K Thangaraj	*Kongu Engineering College, Coimbatore*
R Thiruneelakandan	*Nandha Engineering College, Coimbatore*
Suganthi Devadason	*Karunya University, Coimbatore*
S Rajesh	*Karunya University, Coimbatore*
T Chitravel	*Bannari Amman Institute of Technology, Coimbatore*
D Caleb	*Sri Sairam Engineering College, Chennai*
M Anbuchezhiyan	*Valliammai Engineering College, Chennai*
Sundareswari	*Sathyabama University, Chennai*
S N Zafarullah	*St Peter's Engineering College, Chennai*
I B Shameem Banu	*B S Abdur Rahman Crescent Engineering College, Chennai*
Rama Devi	*Anand Institute of Higher Education, Chennai*
R Kannan	*Pondicherry Engineering College, Pondicherry*

I am grateful to Vibha Mahajan, Ebi John, Shalini Jha, Suman Sen, Heragu Nagaraja, Sohini Mukherjee and Satinder S Baveja of Tata McGraw-Hill for their keen interest in bringing out this book. I am indebted to many of my friends, both in India and abroad, for their continuous support and feedback during the development of the manuscript. The patience and care shown by my family members also deserves a special mention.

I hope that this book will be well received by the students and teachers alike. I invite criticism and suggestions from the users of this book for making further improvements.

V Rajendran

Feedback

Constructive suggestions and criticism always go a long way in enhancing any endeavour. We request all readers to email us their valuable comments/views/feedback for the betterment of the book at tmh.corefeedback@gmail.com, mentioning the title and author name in the subject line. Also, please feel free to report any piracy of the book spotted by you.

Foreword

The text book *Engineering Physics* written by Dr V Rajendran is a comprehensive book to cater to the multi-disciplinary engineering and technology students and is a tailormade to suit the entire syllabus of major technological universities in India. I am happy to note that the overview of this book reflects the vast teaching and research experience of the author. I appreciate the care and efforts taken by the author in explaining the basic principles of physics and its engineering applications. This book is one of the best study materials having clear description of the concepts with appropriate illustrations.

It is very essential that the students should understand the concept, principles and working of the many phenomena of physics in a lucid manner. In order to explain the contents of each chapter, the author has taken great care to provide proper subheadings which describe the topics seriously. It is most important to note that the students must get a clear picture about the engineering materials based on the fundamental concepts of physics. The illustrations of various principles and exercises at the end of the topics will be helpful to make learners improve their problem-solving skills. The advanced topics like biomaterials, shape memory alloys, non-linear materials, metallic glasses, nanomaterial synthesis and nano devices have also been included.

I am sure that this textbook will meet all the needs of aspiring students who want to learn and update their knowledge in technological developments. This book will not only be a ready information storehouse for students entering the field of engineering and technology, but will also be useful as reference material for faculties working in different colleges in India.

I congratulate Dr Rajendran for his interactive and cognitive ways in bringing out this much-useful book.

W Selvamurthy

Distinguished Scientist and Chief Controller R&D (LS)
Government of India
Ministry of Defence
Defence Research and Develpoment Organisation (DRDO)
New Delhi

The McGraw·Hill Companies

Contents

Preface *v*

Foreword *vii*

1. Elasticity **1.1–1.40**

1.1 Introduction *1.1*
1.2 Classification of Elastic Materials *1.1*
1.3 Stress *1.2*
1.4 Strain *1.2*
1.5 Hooke's Law *1.2*
1.6 Elastic Behaviour of a Material *1.3*
1.7 Factors Affecting Elasticity *1.4*
1.8 Classification of Elastic Modulus *1.4*
1.9 Poisson's Ratio *1.7*
1.10 Relation Between Elastic Moduli *1.8*
1.11 Twisting Couple on a Wire *1.15*
1.12 Twisting Couple on a Solid Shaft *1.17*
1.13 Torsional Pendulum *1.18*
1.14 Bending of Beam *1.22*
1.15 Cantilever *1.25*

Keypoints to Remember *1.31*
Solved Problems *1.32*
Objectives Questions *1.41*
Exercises *1.43*
Short Questions *1.44*
Descriptive Questions *1.45*

2. Viscosity **2.1–2.16**

2.1 Introduction *2.1*
2.2 Coefficient of Viscosity *2.2*
2.3 Streamline and Turbulent Flow *2.2*
2.4 Reynold's Number *2.3*
2.5 Poiseuille's Equation for the Flow of a liquid Through a Tube *2.3*
2.6 Motion of a Rigid Body in a Viscous Medium *2.5*
2.7 Experimental Determination of Viscosity of a Liquid *2.7*

Keypoints to Remember *2.11*
Solved Problems *2.11*

The McGraw·Hill Companies

Objective Questions 2.13
Exercises 2.14
Short Questions 2.15
Descriptive Questions 2.15

3. Vacuum Technique **3.1–3.14**

3.1 Introduction *3.1*
3.2 Types of Pumps *3.1*
3.3 Vacuum Systems *3.2*
3.4 Production of Vacuum *3.3*
3.5 Measurements of Vacuum *3.7*
3.6 Importance of Vacuum in Industries *3.10*
3.7 Vacuum Techniques Applications *3.10*

Keypoints to Remember 3.11
Objective Questions 3.11
Short Answer Questions 3.12
Descriptive-Type Questions 3.13

4. Acoustics **4.1–4.44**

4.1 Introduction *4.1*
4.2 Classification of Sound *4.1*
4.3 Characteristics of Musical Sounds *4.3*
4.4 Transmission of Sound *4.8*
4.5 Transmission Loss *4.10*
4.6 Acoustics of Buildings *4.10*
4.7 Sabine's Formula for Reverberation *4.15*
4.8 Measurement of Absorption Coefficient *4.22*
4.9 Sound Absorbing Materials *4.24*
4.10 Factors Affecting Acoustics of Buildings and Their Remedies *4.25*
4.11 Principles to be Observed in the Acoustical Design of an Auditorium *4.28*
4.12 Acoustical Analysis and its Correction *4.30*
4.13 Sound Insulation *4.30*
4.14 Noise Pollution *4.31*
4.15 Noise Control in Machines *4.32*

Keypoints to Remember 4.33
Solved Problems 4.34
Exercises 4.42
Objectives Questions 4.42
Short Answer Questions 4.43
Descriptive Questions 4.44

The McGraw·Hill Companies

5. Ultrasonics **5.1–5.34**

5.1 Introduction *5.1*
5.2 Classification of Ultrasonic Waves *5.1*
5.3 Properties of Ultrasonic Waves *5.5*
5.4 Generation of Ultrasonic Waves *5.5*
5.5 Ultrasonic Velocity Measurements *5.9*
5.6 Absorption and Dispersion of Ultrasonic Waves *5.12*
5.7 Source of Sound Absorption and Dispersion *5.13*
5.8 Applications—Industry *5.13*
5.9 Applications—Medicine *5.16*
5.10 Acoustic Holography *5.20*
5.11 General Applications *5.23*

Keypoints to Remember *5.24*
Solved Problems *5.25*
Objectives Questions *5.29*
Exercises *5.32*
Short Answer Questions *5.32*
Descriptive Questions *5.33*

6. Non-Destructive Testing **6.1–6.45**

6.1 Introduction *6.1*
6.2 Classification of Testing Methods *6.2*
6.3 Visual Inspection *6.2*
6.4 Liquid Penetrant Testing *6.3*
6.5 Radiography *6.8*
6.6 Ultrasonic Testing *6.21*
6.7 Thermography *6.31*

Keypoints to Remember *6.36*
Solved Problems *6.38*
Objectives Questions *6.41*
Short Answer Questions *6.43*
Descriptive Questions *6.45*

7. Interference **7.1–7.32**

7.1 Introduction *7.1*
7.2 Interference Produced in Thin Films due to Reflected Light *7.1*
7.3 Air Wedge *7.4*
7.4 Anti-Reflection Coating *7.8*
7.5 Multi Layer Periodic System *7.10*
7.6 High-Pass and Low-Pass Filters *7.11*

The McGraw·Hill Companies

7.7 Nterference Filter (Fabry–Perot Interference Filter) *7.12*
7.8 Michelson's Interferometer *7.13*

Keypoints to Remember *7.22*
Solved Problems *7.23*
Objectives Questions *7.28*
Exercises *7.29*
Short−Answer Questions *7.30*
Descriptive Questions *7.31*

8. Polarisation **8.1–22**

8.1 Introduction *8.1*
8.2 Fundamentals of Polarised Light *8.2*
8.3 Nicol Prism *8.5*
8.4 Quarter Wave Plate *8.8*
8.5 Half Wave Plate *8.8*
8.6 Theory of Polarised Light *8.9*
8.7 Plane Polarised Light by Nicol Prism *8.12*
8.8 Production of Plane, Circularly and Elliptically Polarised Light *8.13*
8.9 Detection of Plane, Circularly and Elliptically Polarised Light *8.14*

Keypoints to Remember *8.16*
Solved Problems *8.17*
Objectives Questions *8.19*
Short Questions *8.20*
Exercises *8.21*
Descriptive Questions *8.21*

9. Photoelasticity **9.1–9.14**

9.1 Introduction *9.1*
9.2 Definitions *9.2*
9.3 Theory of Photoelasticity *9.4*
9.4 Arrangement of Optical Elements in a Polariscope *9.6*
9.5 Interpretation of Isoclinic and Isochromatic Fringes *9.7*
9.6 Photoelastic Bench *9.9*

Keypoints to Remember *9.10*
Solved Problems *9.10*
Objectives Questions *9.11*
Exercises *9.12*
Short Questions *9.13*
Descriptive Questions *9.13*

The McGraw·Hill Companies

10. Optical and Other Instruments **10.1–10.16**

10.1 Introduction *10.1*
10.2 Sextant *10.1*
10.3 Metallurgical Microscope *10.4*
10.4 Focusing of Electron Beams *10.5*
10.5 Electron Microscope *10.6*
10.6 Scanning Electron Microscope *10.7*
10.7 Transmission Electron Microscope *10.8*
10.8 Scanning Transmission Electron Microscope *10.10*

Keypoints to Remember *10.12*
Objectives Questions *10.13*
Short Questions *10.14*
Descriptive Questions *10.14*

11. Laser **11.1–11.52**

11.1 Introduction *11.1*
11.2 Principle of Laser *11.2*
11.3 Einstein's Theory of Stimulated Emission *11.3*
11.4 Population Inversion *11.5*
11.5 Methods of Achieving Population Inversion *11.6*
11.6 Threshold Condition (Schawlow and Townes Condition) *11.7*
11.7 Types of Lasers *11.11*
11.8 Determination of Wavelength of Laser Using Grating *11.23*
11.9 Particle Size Determination by Laser *11.25*
11.10 Applications of Lasers *11.26*

Keypoints to Remember *11.40*
Solved Problems *11.41*
Objectives Questions *11.47*
Exercises *11.50*
Short Questions *11.50*
Descriptive Questions *11.51*

12. Fiber Optics **12.1–12.48**

12.1 Introduction *12.1*
12.2 Optical Fibers–Principle *12.1*
12.3 Structure of Optical Fibers *12.3*
12.4 Acceptance Angle and Cone *12.4*
12.5 Numerical Aperture and Acceptance Angle *12.4*
12.6 Types of Optical Fibers *12.6*
12.7 Fabrication of Optical Fibers *12.13*
12.8 Loss in Optical Fibers *12.16*

The McGraw·Hill Companies

12.9 Fiber Optical Communication *12.19*
12.10 Splicing *12.20*
12.11 Light Sources for Fiber Optics *12.22*
12.12 Photodetectors *12.28*
12.13 Fiber Optical Sensor *12.32*
12.14 Classification of Optical Sensors *12.32*
12.15 Fiber Endoscope *12.35*
12.16 Engineering Applications of Optical Fibers *12.36*

Keypoints to Remember *12.36*
Solved Problems *12.38*
Objectives Questions *12.45*
Exercises *12.46*
Short Questions *12.47*
Descriptive Questions *12.48*

13. Crystal Structure **13.1-13.36**

13.1 Introduction *13.1*
13.2 Fundamental Terms of Crystallography *13.2*
13.3 Types of Crystals *13.5*
13.4 Relation Between The Interplanar Distance and The Interatomic Distance *13.9*
13.5 Crystal Structures of Materials *13.11*
13.6 Simple Cubic Crystal Structure *13.11*
13.7 Body Centred Cubic Structure *13.12*
13.8 Face Centred Cubic Structure or Cubic Close Packed Structure *13.15*
13.9 Hexagonal Closed Packed Structure *13.16*

Keypoints to Remember *13.20*
Solved Problems *13.20*
Objectives Questions *13.34*
Exercises *35*
Short Questions *13.35*
Descriptive Questions *13.36*

14. Waves and Particles **14.1–14.28**

14.1 Introduction *14.1*
14.2 De Broglie Wave *14.2*
14.3 De Broglie Wavelength *14.2*
14.4 Properties of Matter Waves. *14.4*
14.5 Matter Waves—Experimental Verification *14.4*
14.6 Schrödinger Wave Equation *14.8*
14.7 Application of Schrödinger's Equation to a Particle in a Box *14.12*

The McGraw·Hill Companies

14.8 Heisenberg Uncertainty Principle *14.15*

Keypoints to Remember *14.16*
Solved Problems. *14.17*
Objectives Questions14. *24*
Exercises *14.25*
Short Questions *14.26*
Descriptive Questions *14.27*

15. Semiconductors **15.1–15.36**

15.1 Introduction *15.1*
15.2 Classification of Solids on the Basis of Band Theory *15.2*
15.3 Classification of Semiconductors *15.3*
15.4 Solar Cells *15.7*
15.5 Display Devices *15.11*
15.6 Active Display Devices *15.11*
15.7 Passive Display Devices *15.18*
15.9 Different Modes of LCD *15.22*
15.10 Liquid Crystal Display System *15.23*
15.11 Comparison Between LEDs and LCDs *15.27*
15.12 Applications *15.28*

Keypoints to Remember *15.29*
Solved Problems *15.31*
Objectives Questions *15.32*
Short Answer Questions *15.34*
Descriptive Questions *15.35*

16. Electron Theory of Solids **16.1–16.26**

16.1 Introduction *16.1*
16.2 Electrical Conduction *16.2*
16.3 Classification of Conducting Materials *16.2*
16.4 Classical Free Electron Or Drude–Lorentz Theory of Metals *16.3*
16.5 Expression For Electrical Conductivity and Drift Velocity *16.4*
16.6 Thermal Conductivity *16.8*
16.7 Expression For Thermal Conductivity *16.9*
16.8 Wiedemann–Franz Law *16.11*
16.9 Verification of Ohm's Law *16.12*
16.10 Classical Free Electron Theory: Advantages and Drawbacks *16.13*

The McGraw·Hill Companies

Keypoints to Remember 16.14
Solved Problems 16.15
Objectives Questions 16.24
Short Questions 16.25
Descriptive Questions 16.26

17. Statistics and Band Theory of Solids **17.1–17.36**

17.1 Introduction *17.1*
17.2 Fermi-Dirac Statistics *17.1*
17.3 Carrier Concentration (Free Electron Density) in Metals *17.7*
17.4 Effect of Temperature on Fermi Energy Function *17.7*
17.5 Significance of Fermi Energy *17.10*
17.6 Effective Mass of an Electron *17.11*
17.7 Concept of Hole *17.15*
17.8 Band theory of Solids – Origin of Energy Gap *17.16*
17.9 Conductivity of Copper and Aluminum *17.22*
17.10 Effect of Temperature and Impurity on Electrical Resistivity of Metals (Matthiessen's Rule) *17.23*

Keypoints to Remember 17.25
Solved Problems 17.25
Objectives Type Questions 17.33
Exercises 17.34
Short Questions 17.35
Descriptive Questions 17.35

18. Transport Properties of Semiconductorss **18.1–18.46**

18.1 Introduction *18.1*
18.2 Carrier Concentration in an Intrinsic Semiconductor *18.2*
18.3 Conductivity of Semiconductors *18.7*
18.4 Extrinsic Semiconductor *18.9*
18.5 ***n***-type Semiconductor *18.9*
18.6 ***p***-type Semiconductor *18.12*
18.7 Hall Effect *18.14*
18.8 Variation of Electrical Conductivity with Temperature *18.18*
18.9 Variation of Fermi Level with Temperature in Extrinsic Semiconductor *18.23*

Keypoints to Remember 18.25
Solved Problems 18.26

The McGraw·Hill Companies

Objectives Questions 18.43
Exercises 18.45
Short Questions 18.45
Descriptive Questions 18.46

19. Superconducting Material **19.1–19.16**

19.1 Introduction *19.1*
19.2 General Propertiess of Superconducting Materials *19.2*
19.3 Types of Superconductors *19.4*
19.4 Bardeen, Cooper and Schrieffer (BCS) Theory *19.5*
19.5 Electron–Phonon Interaction *19.6*
19.6 High Temperature Superconductors *19.7*
19.7 Applications *19.8*

Keypoints to Remember 19.11
Solved Problems 19.12
Objectives Questions 19.14
Short Questions 19.15
Descriptive Questions 19.16

20. Magnetic Materials **20.1–20.82**

20.1 Introduction *20.1*
20.2 Magnetic Parameters *20.2*
20.3 Bohr Magneton *20.3*
20.4 Classification of Magnetic Materials *20.4*
20.5 Origin of Permanent Magnetic Moment *20.9*
20.7 Paramagnetism *20.14*
20.8 Ferromagnetic Materials *20.26*
20.9 Antiferromagnetic Materials *20.43*
20.10 Ferrimagnetic Materials *20.44*
20.11 Hard and Soft Magnetic Materials *20.47*
20.12 Energy Product of Magnetic Matersials *20.49*
20.13 Ferrite Core Memory *20.52*
20.14 Magnetic Recording Materials *20.53*
20.15 Magnetic Principle of Analog Recording and Recording *20.54*
20.16 Magnetic Bubble Memory *20.55*
20.17 Magnetic Principle in Computer Data Storage *20.58*
20.18 Magnetic Tape *20.58*
20.19 Floppy Disk *20.60*
20.20 Magnetic Hard Disk *20.61*
20.21 Computer Aided Tomography *20.63*

Keypoints to Remember 20.65
Solved Problems 20.66

The McGraw·Hill Companies

Objectives Questions 20.74
Exercises 20.77
Short−Answer Questions 20.78
Descriptive Questions 20.80

21. Dielectric Materials **21.1–21.46**

21.1 Introduction *21.1*
21.2 Definitions *21.1*
21.3 Different Types of Polarisations *21.4*
21.4 Local or Internal Field *21.10*
21.5 Types of Dielectric Materials *21.12*
21.6 Classification of Electrical Insulating Materials *21.14*
21.7 Claussius-Mosotti Equation *21.15*
21.8 Experimental Determination of Dielectric Constant *21.17*
21.9 Dielectric Loss *21.19*
21.10 Dielectric Breakdown *21.20*
21.11 Ferroelectric Material *21.22*
21.12 Dielectric Properties *21.25*
21.13 Active and Passive Dielectrics *21.26*
21.14 Frequency and Temperature Dependence of Dielectric Properties *21.26*
21.15 Uses of Dielectric MateRials *18.29*
21.16 Application *21.32*

Keypoints to Remember *21.32*
Solved Problems *21.33*
One Mark Question *21.41*
Exercises *21.44*
Short Questions *21.45*
Descriptive Questions *21.45*

22. Shape Memory Alloys **22.1–22.18**

22.1 Introduction *22.1*
22.2 Origin of Shape Memory Alloys *22.2*
22.3 Principle of Phase Transformation in Shape Memory Alloys *22.2*
22.4 Shape Memory Alloys—Properties *22.5*
22.5 Processing Techniques *22.8*
22.6 Characterisation Techniques *22.81*
22.7 Commercial Shape Memory Alloys *22.10*
22.8 Shape Memory Alloys-Applications *22.14*

22.9 Shape Memory Alloys-Disadvantages *22.15*

Keypoints to Remember *22.15*
Objectives Questions *22.16*
Short Questions *22.17*
Descriptive Questions *22.17*

23. Nonlinear Materials **23.1–23.16.**

23.1 Introduction *23.1*
23.2 Basic Principle *23.2*
23.3 Classification of Nonlinear Materials *23.3*
23.4 Nonlinear Properties *23.4*
23.5 Nonlinear Materials *23.10*
23.6 Applications *23.12*

Keypoints to Remember *23.12*
Objectives Questions *23.13*
Short Questions *23.14*
Descriptive Questions *23.15*

24. Metallic Glasses **24.1–24.14**

24.1 Introduction *24.1*
24.2 Origin of Metallic Glasses *24.2*
24.3 Principle *24.2*
24.4 Preparation *24.3*
24.5 Properties *24.5*
24.6 Applications *24.10*

Keypoints to Remember *24.12*
Objectives Questions *24.12*
Short−Answer Questions *24.14*
Descriptive Questions *24.14*

25. Biomaterials **25.1–25.16**

25.1 Introduction *25.1*
25.2 Biomechanism *25.1*
25.3 Development of Biomaterials *25.3*
25.4 Classification of Biomaterials *25.8*
25.5 Processing and Properties *25.11*
25.6 Applications *25.11*

Keypoints to Remember *25.13*
Objectives Questions *25.13*
Short Questions *25.14*
Descriptive Questions *25.14*

The McGraw·Hill Companies

26. Nanomaterial Synthesis **26.1–26.38**

26.1 Introduction *26.1*
26.2 Synthesis of Nanostructured Materials *26.2*
26.3 Top-Down Approach-Nanomaterials Synthesis *26.4*
26.4 Bottom Up Process—Synthesis of Nanoparticles *26.11*
26.5 Vapour Phase Deposition *26.11*
26.6 Epitaxial Techniques—Synthesis of Nanomaterials *26.15*
26.7 Chemical Methods—Nanomaterial Synthesis *26.20*
26.8 Hybrid Methods—Synthesis of Nanomaterials *26.23*
26.9 Nanotechnology and Environment *26.28*
26.10 Properties and Possible Applications *26.29*
26.11 Storage *26.33*

Keypoints to Remember *26.34*
One Mark Question *26.35*
Short Questions *26.37*
Descriptive Questions *26.38*

27. Nanodevices **27.1–27.34**

27.1 Introduction *27.1*
27.2 Nanomagnets *27.2*
27.3 Classifications of Nanomagnetic Materials *27.3*
27.4 Magneto Resistances *27.6*
27.5 Probing Nanomagnetic Materials *27.9*
27.6 Nanomagnetism in Technology *27.9*
27.7 Applications of Semiconductor Nanostructures and Devices *27.11*
27.8 Applications of Semiconductor Nanostructure *27.12*
27.9 Organic Semiconductor Materials Devices *27.18*
27.10 Carbon Nanotubes *27.22*
27.11 Types of Carbon Nanotubes *27.22*
27.12 Synthesis of Carbon Nanotubes *27.24*
27.13 Properties of CNT *27.26*
27.14 Applications *27.28*

Keypoints to Remember *27.30*
Objectives Questions *27.31*
Short Questions *27.33*
Descriptive Questions *27.34*

Appendix 1 **A-1**

Appendix 2 **A-2**

Appendix 3 **A-3**

Appendix 4 **A-4**

Chapter

Elasticity

1

OBJECTIVES

- To understand the mechanical properties of engineering materials
- To explore the elastic properties of materials
- To study the different classes of modulus and their relationship
- To explore the possible methods to determine the modulus of materials with different shapes
- To study the restoring couple acting on a wire and shaft
- To underst.and the principle behind a cantilever and its possible applications

1.1 INTRODUCTION

The knowledge of physical, chemical, mechanical and technological properties of materials are more essentially required for selection of right materials for industrial applications. Properties such as dielectric constant, electrical conduction, thermal condition, etc., are used to explore newer materials for electronic applications. Similarly, properties such as stress, strain, elastic moduli, etc., are used to explore materials for industrial applications such as construction of bridges and railway wagons.

Elasticity is a branch of physics which deals with the elastic property of materials. Elasticity is the property by virtue of which bodies regain original shape and size after removal of the deforming force. When a deforming force acts on a body, there is a change in its length or shape or volume. Then, the body is said to be strained. In this chapter, elastic properties of materials and their applications for different uses are discussed briefly.

1.2 CLASSIFICATION OF ELASTIC MATERIALS

Generally, materials are classified as

(i) Perfectly elastic, and
(ii) Plastic

The McGraw·Hill Companies

The material which completely regains its original shape or size after removal of the deforming force is known as a *perfectly elastic material*. On the other hand, materials which do not regain their original shape or size even after removal of the deforming force are known as plastic materials. In general, there is no material which is perfectly elastic or plastic. Even a quartz fiber which nearly approaches the criteria for a perfectly elastic body does not regain its original state after the removal of a large deforming force. On the other hand, putty is a perfectly plastic body, which regains its original state upon removal of the applied deforming force. Thus, a relative measurement is used to represent a material, i.e., plastic or elastic in comparison with other materials.

1.3 STRESS

When a deforming force or load is acting on a body, a force of reaction comes into play internally inside the body due to relative displacement of molecules. The displaced molecules tend to regain their original position in the body due to their elastic nature. Therefore, the restoring force, or recovering force per unit area, is known as stress i.e.,

i.e., $$\text{Stress} = \frac{F}{A} \tag{1.1}$$

where F is the applied force in newtons and A, the area of cross section in square meters. The unit of stress is N m^{-2} or Pascal.

When stress is applied normal to the surface of a body and wire, the result is a change in volume and length respectively. The restoring force per unit area is known as *tangential* or *shearing stress* which is acting along the surface. On the other hand, if there is any change in the shape of a body then the stress is tangential stress. When the volume of a body increases then the stress is known as *expansive stress*, while it is *compressive stress* for decrease in volume.

1.4 STRAIN

Under equilibrium, the ratio of change in dimension produced by an external force to its original dimension is called *strain*. The observed change in dimension may be either in length or volume or shape. Strain is a dimensionless quantity and has no unit.

1.5 HOOKE'S LAW

In 1679, Robert Hooke proposed a relation between stress and strain. When a material is within the elastic limit, stress is directly proportional to strain. The ratio of stress to strain is a constant and is known as *coefficient of elasticity* or *modulus of elasticity*. Therefore,

$$\text{Stress} \propto \text{Strain}$$

or

$$\text{Stress} = E \times \text{Strain}$$

Therefore, modulus of elasticity

$$E = \frac{\text{Stress}}{\text{Strain}} \tag{1.2}$$

Equation (1.2) gives the value of modulus of elasticity.

The McGraw·Hill Companies

Elastic and Plastic Limits

Generally, for a solids, there is a limit to recover its original state upon removal of the deforming force. When the applied stress to a body is increased up to a maximum value, the body recovers its original shape and size upon removal of the deforming force. The maximum stress up to which a body exhibits the property of elasticity is known as *elastic limit*. When the applied stress to the body exceeds the maximum stress limit, the body does not regain its original shape or size completely upon withdrawal of the applied stress. Therefore, any body beyond its elastic limit is in its plastic limit.

1.6 ELASTIC BEHAVIOUR OF A MATERIAL

Consider that a body is subjected to a uniform stress. When the stress increases, it results in a change in its dimensions, i.e., there is strain on the body. The relationship between the applied stress and strain is studied by drawing a curve between strain and stress, as shown in Fig.1.1. The curve is called a *stress–strain curve*. Let a body be clamped at one end and loaded at the other end gradually from zero value until the wire breaks. Figure 1.1 shows the nature of stress–strain curve for a body.

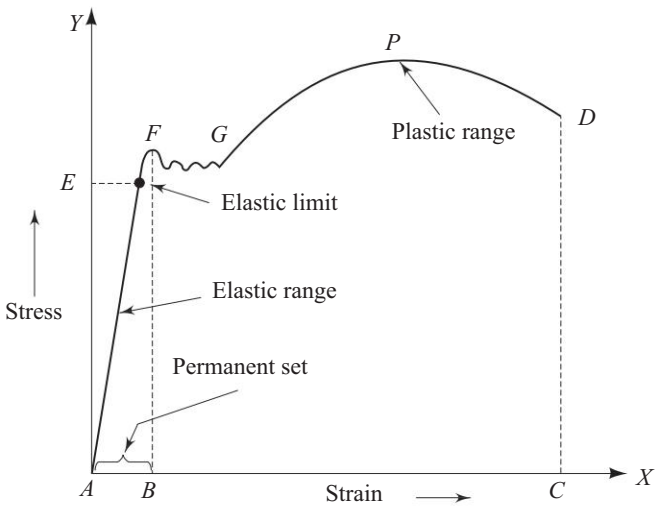


Fig. 1.1 *Stress–strain curve–elastic behaviour of a solid*

In the region *AE* of the curve, the applied stress is directly proposal to the strain. When the body is within this applied stress, it regains its original shape or size. Therefore, the region *AE* is perfectly in the elastic region and the point *E* is known as *elastic limit*. Further, the material in the region *AE* obeys Hooke's Law.

When stress is applied beyond the elastic limit *F*, an exponential growth in strain takes place and is represented by the region *EF* in the curve. In the region *EF*, the material is partly elastic and partly plastic. Under this condition, upon removal of deforming force, the material takes a new path *FB* instead of *FE* to return to its original state. On return to the original state, the body has a small strain *AB* which is known as *residual strain*. The residual strain acquires a permanent set *AB* as shown in Fig.1.1.

The McGraw·Hill Companies

Beyond the point *F* in the region *EF,* there is no increase in stress. However, a random variation in the strain is seen up to the point *G*. A large increase in strain starts at the point *F* and hence, it is called the *yield point*. The region *GP* is known as *plastic region*. In this region a steady increase in strain with increase in stress is noticed with a decrease in cross-sectional area of the solid. Beyond the point *P,* an increase in strain without any increase in stress is noticed. The corresponding region is represented as *PD*. The strain corresponding to the point *P* is known as *breaking stress* or *ultimate stress*. Finally, the solid breaks at the point *D*, which is called the *breaking point*. At this point, stress on the solid becomes zero. The solid is under fatigue when it is subjected into continuous stress and strain. When a body is continuously subjected to stress and strain, it gets fatigue.

1.7 FACTORS AFFECTING ELASTICITY

The elastic properties of solids depend on the following parameters:

(i) Effect of Stress Due to the application of large constant stress or repeated number of cycles of stresses on a body a gradual decrease in elasticity of a body results.

(ii) Effect of Temperature Generally, the elasticity decreases with increase in temperature. For example, at room temperature, carbon filament is in elastic state. When it is heated by passing a current, it is converted into plastic state. Most of the solids follow a similar behaviour with increase in temperature. However, elasticity remains unaffected with change in a temperature in a material like invar steel.

(iii) Effect of Hammering and Rolling The elastic properties of solids are increased while doing operations like hammering and rolling.

(iv) Effect of Annealing When solid materials are annealed (heating or cooling), it results in the formation of uniform crystalline structure along with the formation of larger crystal grains. Thus, the softening of the materials takes place. Therefore, the elastic **property** of **the materials** decreases due to annealing.

(v) Effect of Impurities When small impurities like carbon and potassium are added to molten iron and gold, it enhances their elastic property. The addition of impurity increases the compactness of materials and hence, elastic property.

1.8 CLASSIFICATION OF ELASTIC MODULUS

The linear volume and shear strains are the most important parameters used to explore elastic properties of materials. The elastic moduli of solids are classified into three different types based on different types of stress, namely,

(i) Young's modulus
(ii) Bulk modulus, and
(iii) Rigidity modulus

Young's, bulk, and rigidity modulus are obtained from the respective linear volume and shear strain of materials. The different types of modulus and their relationships are discussed in detail in the following heading.

The McGraw·Hill Companies

1.8.1 Young's Modulus

Let us consider a wire of length L and area of cross-section A. The top end of the wire is fixed as shown in Fig. 1.2, while at the bottom end of the wire, a known load is applied. Consider that ΔL is the elongation in length of the wire due to the action of stress (E), which is acting along the length of the wire. Therefore, the ratio of the change in length to unit original length is known as *longitudinal* or *linear* or *elongation strain*. Longitudinal or linear stress is the force applied per unit area of cross – section. Therefore, Young's modulus is defined as the ratio of longitudinal stress to linear strain within elastic limit.

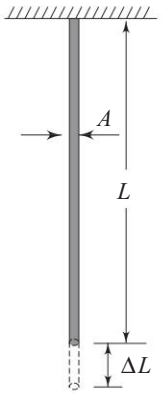


Fig. 1.2 *Young's modulus—wire*

i.e., $$\text{Young's modulus} = \frac{\text{Longitudinal stress}}{\text{Linear strain}} \tag{1.3}$$

or $$Y = \frac{F / A}{\Delta L / L} \tag{1.4}$$

where F is the force and A the area of cross-section.

Rearranging the above equation, we get,

$$Y = \frac{F\,L}{A\,\Delta L} \tag{1.5}$$

We know that $F = mg$ and $A = \pi r^2$, where r is the radius of the wire and m, the mass of the wire. Substituting the value of F and A in the above equation, we get,

$$Y = \frac{m\,g\,L}{\pi r^2\,\Delta L} \tag{1.6}$$

Equation (1.6) gives the Young's modulus of the wire. The unit for Young's modulus is N m^{-2} or Pascal.

1.8.2 Bulk Modulus

Let us consider a cube of volume V and area of cross-section A as shown in Fig.1.3. Consider that hydrostatic stress or bulk stress ΔP ($\sim F/A$) is applied under normal condition to the whole surface of the body. As a result, a change in volume takes place without any change in the shape of the body. Let the change in volume be dV. Therefore, the bulk strain developed in the body is $-\Delta V/V$. The increase in hydrostatic stress with decrease in volume is indicated by the negative sign. Therefore, the ratio of bulk stress to bulk strain within elastic limit is known as *bulk modulus*.

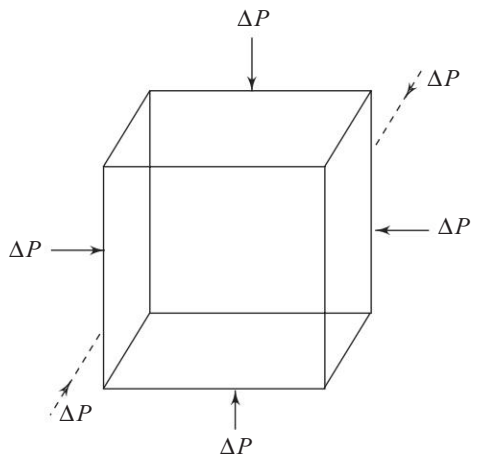


Fig. 1.3 *Bulk modulus—Cube*

$$\text{Bulk modulus} = \frac{\text{Bulk stress}}{\text{Bulk strain}} \tag{1.7}$$

or

$$K = \frac{-\Delta P V}{\Delta V} \qquad \text{N m}^{-2} \text{ or Pascal} \tag{1.8}$$

where P is the pressure.

Equation (1.8) gives the bulk modulus of a cube. The reciprocal of bulk modulus is called *compressibility* ($1/K$).

1.8.3. Rigidity Modulus

Let a solid cube be subjected to a force in the upper portion and hence, produce a couple. The couple moves the upper portion of the solid cube as shown in Fig.

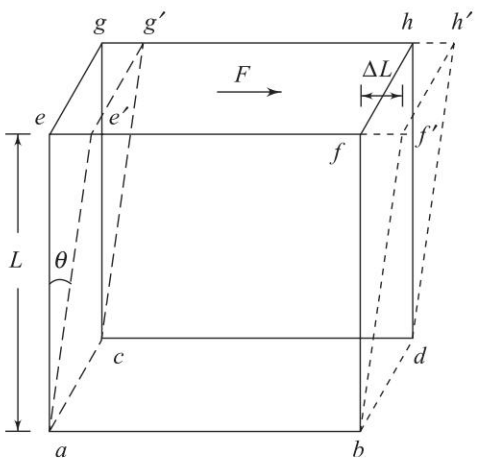


Fig. 1.4 *Solid cube—Rigidity modulus*

The McGraw·Hill Companies

1.4. The upper portions of the cube, A, B, C and D, are shifted respectively to point A', B', C' and D'. As a result, the face $ABCD$ is sheared through an angle θ. Thus, the vertical face is shifted through an angle θ. Therefore,

$$\text{Shearing strain } \theta = \frac{A\,A'}{E\,A} = \Delta L / L \tag{1.9}$$

The shearing or tangential stress is equal to the ratio of the force (F) acting on the face to the area A of the face $ABCD$. Therefore,

$$\text{Shearing stress} = \frac{F}{A} \tag{1.10}$$

When the material is within elastic limit, rigidity modulus is defined as the ratio of shearing stress to shearing strain. *Therefore,*

$$\text{Rigidity modulus} = \frac{\text{Shearing stress}}{\text{Shearing strain}} \tag{1.11}$$

Substituting stress and strain values in the above equation, we get,

$$\eta = \frac{F / A}{\Delta L / L}$$

or

$$\eta = \frac{F\ L}{A\ \Delta L} \tag{1.12}$$

Equation (1.12) gives the rigidity modulus of the cube. The equation for rigidity modulus is similar to that of Young's modulus. In rigidity modulus, *FIA* is the tangential stress, while in Young's modulus, the stress is a linear stress.

1.9 POISSON'S RATIO

Whenever a body is subjected to a stress in a particular direction, then there is a change in dimension of a body in the other two perpendicular directions. For example, when a wire is stretched along its length then there is an increase in length of the wire. As a result, the wire becomes thinner and hence, the cross-section of the wire decreases. The ratio of the change in the lateral strain to the linear strain within elastic limit gives Poisson's ratio,

$$\text{Linear stress} = \frac{\text{change in diameter of the wire}}{\text{original diameter of the wire}} = \frac{\Delta d}{d} \tag{1.13}$$

$$\text{Lateral strain} = \frac{\text{change in length of the wire}}{\text{original length of the wire}} = \frac{\Delta \ell}{\ell} \tag{1.14}$$

$$\text{Therefore, Poisson's ratio} = \frac{\text{Lateral strain}}{\text{Linear strain}} \tag{1.15}$$

The McGraw·Hill Companies

or $$\sigma = \frac{\text{Change in diameter / original diameter of the wire}}{\text{Change in length / original of the wire}}$$

Substituting the values of Eq. (1.13) and (1.14) in Eq. (1.15), we get,

$$\sigma = \frac{\Delta d / d}{\Delta l / l}$$

where $\beta = \Delta d/d$ and $\alpha = \Delta l/l$, we get,

$$\text{Poission's ratio } \sigma = \frac{\beta}{\alpha} \tag{1.16}$$

where α and β are respectively the longitudinal and lateral strain per unit stress. Within elastic limit, lateral strain is proportional to linear stress. The value of Poisson's ratio mainly depends upon the nature of the material. It has no units. In most materials, the value of Poisson's ratio lies between 0.2 and 0.4.

1.10 RELATION BETWEEN ELASTIC MODULI

The relationship between elastic moduli, namely, Young's, bulk and rigidity moduli, and Poisson's ratio are very useful to explore mechanical properties of materials. The relations between the three moduli and Poisson's ratio are discussed in the following section:

1.10.1 Relation between Young's Modulus and Elongation

Consider a cube subjected to a unit tension along one edge. Then α is the elongation along the direction of force per unit length per unit tension. Therefore,

$$\text{Linear Stress} = \frac{\text{Force}}{\text{Area}} = 1 \tag{1.17}$$

Similarly,

$$\text{Linear strain} = \frac{\text{Change in length}}{\text{Length}} = \frac{\alpha}{1} = \alpha \tag{1.18}$$

Substituting the linear stress and strain from Eq. (1.17) and Eq. (1.18) in Young's modulus equation, we get,

$$\text{Young's modulus } Y = \frac{1}{\alpha} \tag{1.19}$$

Equation (1.19) shows the relationship between Young's modulus and elongation of a cube.

1.10.2 Relation between Young's and Bulk Modulus

Consider a cube *ABCDEFGH* with unit volume. P_x, P_y and P_z are stresses acting along *x*, *y* and *z* directions as shown in Fig. 1.5. The stress P_x is perpendicular to faces *ACEG* and *BDFH*, while P_y is perpendicular to faces *ABCD* and *EFGH*. However, the stress P_z is perpendicular to faces *ACBD* and *EGFH*. Each tension acting on the cube produces an extension in its own direction and a lateral contraction in the other two directions. α is the

The McGraw·Hill Companies

elongation along the direction of force per unit length per unit tension. Then, $P_x\alpha$, $P_y\alpha$ and $P_z\alpha$ are the elongations produced respectively along the edges *AB, BF* and *BD*.

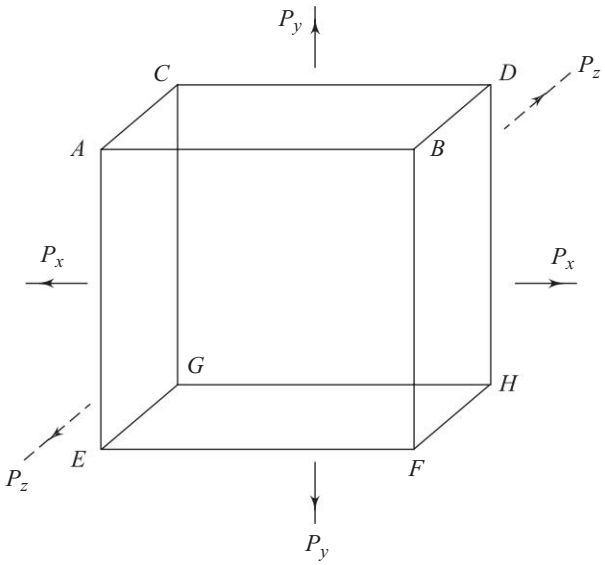


Fig. 1.5 *Cube—application of bulk stress*

Consider that β is the contraction produced perpendicular to the direction of force per unit length per unit stress. Therefore, the contraction produced along the edges *AB, BF* and *BD* are respectively $P_x\beta$, $P_y\beta$ and $P_z\beta$.
Therefore, length of the edges is

$$AB = 1 + P_x\alpha - P_y\beta - P_z\beta \tag{1.20}$$

$$BF = + P_y\alpha - P_x\beta - P_z\beta \tag{1.21}$$

$$BD = + P_z\alpha - P_x\beta - P_y\beta \tag{1.22}$$

Therefore, volume of the cube

$$V = AB \times BF \times BD \tag{1.23}$$

Substituting *AB, BF* and *BD* values from Eqs (1.20), (1.21) and (1.22) in Eq.(1.23), we get, volume of the cube,

$$V = \{1 + P_x\alpha - \beta(P_y + P_z)\}\ \{1 + P_y\alpha - \beta(P_x + P_z)\}\{(1 + P_z\alpha - \beta(P_x + P_y)\} \tag{1.24}$$

Simplifying the above equation by neglecting squares, and products of α and β, we get,

$$V = 1 + (\alpha - 2\beta)(P_x + P_y + P_z) \tag{1.25}$$

When the stresses are equal $P_x = P_y = P_z = P$
Therefore, Eq. (1.25) can be simplified as,

$$V = 1 + 3P\,(\alpha - 2\beta) \tag{1.26}$$

The McGraw·Hill Companies

It is clear from Eq. (1.26) that, volume of the cube is increased by 3P ($\alpha - 2\beta$) from its original value.

Therefore, volume strain $= \dfrac{3P(\alpha - 2\beta)}{1}$

$$= 3P(\alpha - 2\beta) \tag{1.27}$$

We know that, bulk modulus $= \dfrac{\text{Stress}}{\text{Volume strain}}$ (1.28)

Substituting the values of volume strain from Eq. (1.27) and stress (P) in the above equation (1.28), we get,

Bulk modulus $K = \dfrac{P}{3P(\alpha - 2\beta)}$

or $$K = \frac{1}{3(\alpha - 2\beta)} \tag{1.29}$$

Equation (1.29) can be written in terms of Young's modulus and Poisson's ratio,

$$K = \frac{1}{3\alpha(1 - 2\beta/\alpha)}$$

Substituting the value of $Y = 11a$ in above equation, we get,

$$K = \frac{Y}{3(1 - 2\sigma)} \tag{1.30}$$

Equation (1.30) gives the relationship between the bulk modulus, Young's modulus and Poission ratio. The compressibility is equal to 3 ($\alpha - 2\beta$).

1.10.3 Relation Between Young's Modulus and Poission's Ratio

ABCDEFGH is a solid cube with cube edge L as shown in Fig. 1.6. The bottom face of the cube *EFGH* is kept fixed while a tangential force F is applied to the upper face

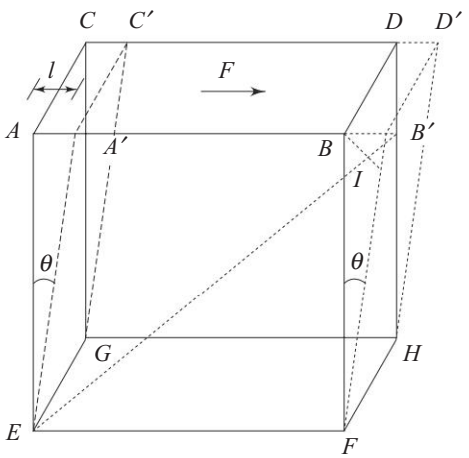


Fig. 1.6 *Cube-Shearing stress*

The McGraw·Hill Companies

ABCD of the cube. Due to the application of shearing force, the upper face is sheared through an angle θ.

The edge of the upper face of the cube *A, B, C* and *D* takes the new position as *A′*, *B′*, *C′* and *D′*. Therefore,

$$\text{Shearing stress} = \frac{F}{\text{area } ABCD}$$

$$= \frac{F}{L^2} = T \tag{1.31}$$

Consider that ℓ *is* the displacement on the edges *A* and *B*.

Therefore, $AA' = BB' = l$

(or) $$\text{Shear strain} = \frac{l}{L} = \theta \tag{1.32}$$

Substituting the values of stress and strain from Eq. (1.31) and (1.32) in Eq. (1.11), we get,

Rigidity modulus $$\eta = \frac{T}{\theta} \tag{1.33}$$

The shearing stress acting along the upper face *AB* is equivalent to a linear tensile stress along the diagonal *EB* and a compressive stress along the diagonal *AF* at right angles to each other. Thus, the tensile stress *T* along *EB* is equal to the compressive stress *T* along *AF*. We know that, α and β are longitudinal and lateral strain per unit stress. Therefore, the extension along the diagonals *EB* and *AF* are respectively *EBT* α and *EBT* β.

Therefore, total extension along the diagonal *EB* is

$$IB' = EB\ T(\alpha + \beta)$$

$$IB' = \sqrt{2}\ L\ T(\alpha + \beta) \tag{1.34}$$

where the diagonal *EB* of cube is equal to $\sqrt{2}\ L$.

Draw a perpendicular line from *B* to the line *EB*' to meet at *I*. *IB′* is the extension of diagonal *EB*.

$$IB' = BB' \cos BB'l$$

$$= l \cos 45°$$

where the angle *BB′ I* is nearly equal to 45°. Therefore,

$$IB' = \frac{\ell}{\sqrt{2}} \tag{1.35}$$

Substituting the value of *IB′* from Eq. (1.35) in Eq. (1.34), we get,

$$\frac{\ell}{\sqrt{2}} = \sqrt{2}\ L\ T(\alpha + \beta)$$

or $$\frac{T}{(l/L)} = \frac{1}{2(\alpha+\beta)} \tag{1.36}$$

We know that the strain $\theta = l/L$. Therefore, Eq. (1.36) becomes

$$\frac{T}{\theta} = \frac{1}{2(\alpha+\beta)} \tag{1.37}$$

Equation (1.37) can be written in terms of Young's modulus by Eq. (1.19) by taking rigidity modulus as $\eta = T/\theta$,

Rigidity modulus $$\eta = \frac{Y}{2(1+\beta/\alpha)} \tag{1.38}$$

In terms of Poisson's ratio, the above equation can be written as,

$$\eta = \frac{Y}{2(1+\sigma)} \tag{1.39}$$

The relationship between Young's modulus, rigidity modulus and Poisson's ratio are given by Eqs. (1.38) and (1.39).

1.10.4 Relation Between Elastic Constants

Similar to the relationship between Young's modulus, rigidity modulus and Poisson's ratio, the relationship between *Y, K* and *n* can be obtained from Eqs (1.29) and (1.37).

We know that from Eq. (1.29),

$$K = \frac{1}{3(\alpha-2\beta)}$$

Rearranging the above equation, we get,

$$(\alpha-2\beta) = \frac{1}{3K} \tag{1.40}$$

Similarly, from Eq. (1.37),

$$\eta = \frac{1}{2(\alpha+\beta)}$$

Rearranging the above equation, we get,

$$(\alpha+\beta) = \frac{1}{2\eta} \tag{1.41}$$

Multiplying Eq. (1.41) by 2 and adding with Eq. (1.40), we get,

$$3\alpha = \frac{1}{3K} + \frac{1}{\eta} \tag{1.42}$$

The McGraw·Hill Companies

Substituting value of Y (= $1/\alpha$) in the above equation, we get,

$$\frac{3}{Y}=\frac{1}{3K}+\frac{1}{\eta} \tag{1.43}$$

Multiplying both sides of the above equation by 3, we get,

$$\frac{9}{Y}=\frac{1}{K}+\frac{3}{\eta} \tag{1.44}$$

Young's modulus, bulk modulus and rigidity modulus are related using Eq. (1.44).

1.10.5 Relation Between Elastic Moduli and Poisson's Ratio

We know that Young's and bulk modulus are related (Eq. (1.30)) as,

$$K=\frac{Y}{3(1-2\sigma)}$$

Rearranging the above equation, we get, $Y = 3\,K\,(1 - 2\sigma)$ (1.45)

Similarly, rigidity and Young's modulus are related Eq. (1.39) as,

$$\eta=\frac{Y}{2(1+\sigma)}$$

Rearranging the above equation, we get,

$$Y = 2\,\eta\,(1 + \sigma) \tag{1.46}$$

The LHS of Eqs (1.45) and (1.46) are equal and hence,

$$2\,\eta\,(1 + \sigma) = 3K\,(1 - 2\sigma)$$

Therefore, Poisson's ratio

$$\sigma = \frac{3K-2\eta}{2\eta+6K} \tag{1.47}$$

Equation (1.47) gives Poisson's ratio in terms of η and K.

We know that σ in terms of K and Y from Eq. (1.30),

$$\sigma=\frac{3K-Y}{6K} \tag{1.48}$$

Similarly, we also know σ in terms of Y and η from Eq. (1.39),

$$\sigma=\frac{Y-2\eta}{2\eta} \tag{1.49}$$

LHS of Eqs (1.48) and (1.49) are equal and hence,

$$\frac{Y-2\eta}{2\eta}=\frac{3K-Y}{6K}$$

Rearranging the above equation, we get,

$$\frac{3}{2Y} = \frac{1}{6K} + \frac{1}{2\eta}$$

Multiplying both sides of the above equation by 6, we get,

$$\frac{9}{Y} = \frac{1}{K} + \frac{3}{\eta} \tag{1.50}$$

Equation (1.50) gives the relation between Young's, bulk and rigidity modulus. Eq. (1.50) is very similar as that of Eq. (1.44).

Limitation of Poisson's Ratio

Comparing the Eqs (1.45) and (1.46), we get,

$$3K(1 - 2\sigma) = 2\eta(1 + \sigma) \tag{1.51}$$

where K and η are positive quantities.

Case I $(\sigma > 0)$

When Poisson's ratio is positive, the RHS and LHS of Eq. (1.51), is positive.

Therefore, $(1 - 2\sigma) > 0$

or

$$2\sigma < 1$$

Therefore, Poisson's ratio σ is less than 0.5.

Case II $(\sigma < 0)$

When Poisson's ratio is negative, the RHS and LHS of Eq. (1.51), is positive,

Therefore, $(1 + \sigma) > 0$

or

$$\sigma > -1$$

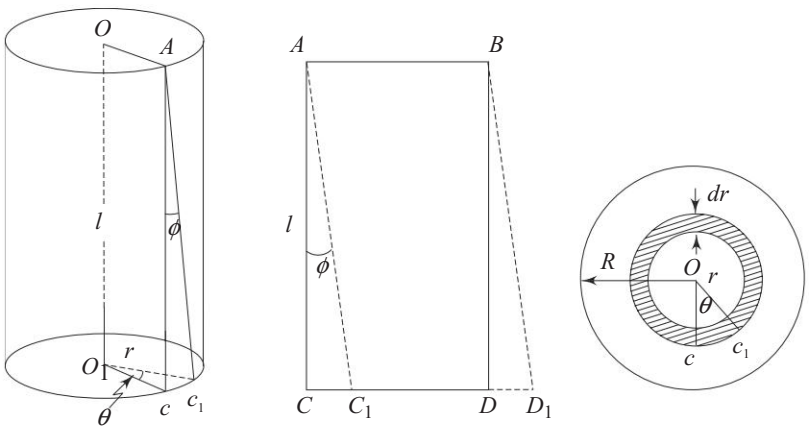
Therefore, Poisson's ratio σ is less than -1. It is clear from the above observation that the value of Poisson's ratio lies between -1 to 0.5.

However, in actual practice, the value of σ cannot be negative because the body does not expand when it expands longitudinally. Substituting $\sigma = 0.5$ in Eq. (1.45), we get the bulk modulus K as infinite and hence, the body is perfectly incompressible. However, there is no substance which is perfectly incompressible. Similarly, while taking $\sigma = -1$ in Eq. (1.51), we get rigidity modulus as infinite. Under these conditions, when a wire is extended linearly due to applied load, then the same is expanded laterally. This phenomenon is rare to exhibit. The value of Poisson's ratio ranges between 0.2 and 0.4 for most of the isotropic solid materials.

The McGraw·Hill Companies

1.11 TWISTING COUPLE ON A WIRE

Let *l* be the length of a cylindrical wire as shown in Fig. 1.7a. The upper end is clamped through a fixed support. The wire consists of a number of cylindrical tubes i.e., co-axial tubes. The radius of tubes ranges from zero to *a*. Consider one such cylindrical shell with radius *r* and thickness *dr* as shown in Fig. 1.7c.



(a) *Elementary tube* (b) *Elementary tube—twisted state* (c) *Cylinder shell*

Fig. 1.7 *Couple–wire*

In elementary tube, the line *AC* is parallel to axis of the tube. Consider that a couple is applied at the bottom of wire. As a results, the wire is twisted through an angle θ. Therefore, in this twisted state, the position of *AC* takes a new position as AC_1 which is shown in Fig. 1.7a.

We know that from Fig. 1.7a, the angle $COC_1 = \theta$

Therefore, $CC_1 = r\theta$ (1.52)

It is known that the displacement is maximum for the points lying on the rim while it decreases as one move towards the centre of wire *O*. This results a maximum shearing strain at the rim and a minimum shearing strain at the axis (centre).

Consider the line *AC* which is flattened out from the hollow cylinder (fig. 1.7b). Therefore, rectangle *ACDB* and AC_1D_1B are obtained before and after twisting. Hence, the shear and shearing strain are obtained,

i.e., $CAC_1 = \phi$

or

$$CC_1 = l\,\phi \tag{1.53}$$

Comparing Eqs (1.52) and (1.53), we get,

$$l\,\phi = r\theta$$

or Shearing strain $\phi = \frac{r\theta}{l}$ (1.54)

We know that, Rigidity modulus $\eta = \frac{\text{Shear stress}}{\text{Shear strain}}$

Therefore, Shearing stress = $\eta \times$ *Shearing strain*

Substituting the value ϕ from Eq. (1.54), we get,

$$\text{Shear stress} = \eta \frac{r}{l} \theta \quad (1.55)$$

The cross-sectional area of elementary tube is equal to = $2 \pi r\, dr$.

Therefore, total shearing force on the wire = Stress × area

Substituting the values, we get, Shearing force $= \frac{\eta r\theta}{l} 2\pi r\, d\, r$

Rearranging the above equation, we get,

$$= \frac{2\pi \eta\theta}{l} r^2\, d\, r \quad (1.56)$$

The moment of the force about the axis of the wire is

$$= \frac{2\pi\eta\theta}{l} r^2 dr r \quad (1.57)$$

Equation (1.57) can be integrated with the limit $r = 0$ and $r = a$, to obtain a twisting couple applied to the whole wire.

Therefore, the total twisting couple about the axis of the whole wire is

$$= \int_0^a \frac{2\pi\eta\theta}{l} r^3 dr \quad (1.58)$$

Integrating the above equation, we get,

$$= \frac{2\pi\eta\theta}{l} \frac{r^4}{4} \Big|_0^a$$

Simplifying the above equation, we get,

or $$= \frac{2\pi\eta\theta}{l} \frac{a^4}{4}$$

or $$= \frac{1}{2} \frac{\pi\eta a^4}{l} \theta \quad (1.59)$$

The McGraw·Hill Companies

When angular twist $\theta = 1$ radian,
The twisting couple per unit angular twist is

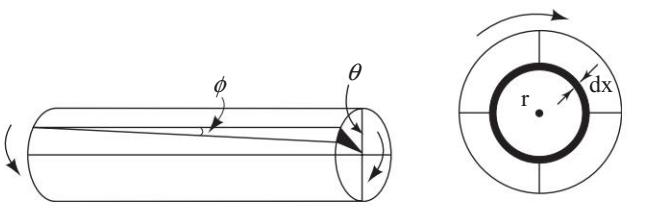
$$C = \frac{1}{2}\frac{\pi\eta a^4}{l} \tag{1.60}$$

Equation (1.60) gives the couple per unit twist of wire.

1.12 TWISTING COUPLE ON A SOLID SHAFT

Consider a solid shaft with high rigidity modulus. It is used to transmit the power from the machine to load. The couple applied at one end of the shaft is transmitted to its other end without any appreciable twist in it. It is rotated on bearing about its own axis. Even for a large couple, the twist in the shaft is very small. The following assumptions are made to drive the relation between torque, angle of twist and shear stress in a solid shaft of diameter *d*:

(i) It consists of a number of thin concentric tubes.
(ii) Each thin tube carries a shear force and is independent from its neighbours.
(iii) The shaft shift is within the elastic limit.
(iv) The radial lines before and after twist remain same.



(a) *Shaft* (b) *Elementary thin tube*

Fig. 1.8 *Twisting couple—solid shaft*

Let *dx* be the thickness of an elementary thin tube and *r* be its radius. Consider that the shaft is twisted through an angle θ.
Therefore, shear force on elementary thin tube = $2\pi r dr \times$ shearing stress (1.61)
where $2\pi rdr$ is the area of cross-section of the thin tube.
Therefore, torque carried by the tube = $2\pi rdr \times$ shearing stress $\times r$ (1.62)
We know that θ is the angle of twist of the tube of length *l*, then substituting the value of Shearing stress = $\frac{\eta r}{l}\theta$ from Eq. (1.55) in Eq. (1.62) we get,

The torque carried out by the tube = $2\pi r^2 dr \times$ shearing stress

i.e., $$= \frac{2\pi r^3 \eta\theta}{l} dr \tag{1.63}$$

The McGraw·Hill Companies

The total torque (moment of the couple) carried out by solid shaft is the sum of all the elementary torques, i.e.,

$$\text{Torque} \quad T = \int_0^{\frac{d}{2}} \frac{2\pi r^3 \eta\theta}{l} dr \tag{1.64}$$

We know that the radial lines before and after twist remains same and hence, angle θ of the twist is constant. Therefore, the rigidity modulus η and the length of the shaft remain constant.

Therefore, Eq. (1.64), can be written as,

$$\text{Torque} \quad T = \frac{\eta\theta}{l} \int_0^{\frac{d}{2}} 2\pi r^3 dr \tag{1.65}$$

$$\text{(or)} \qquad T = \frac{\eta\theta}{l}\left[2\pi\left(\frac{d}{2}\right)^4 \frac{1}{4}\right]$$

Rearranging the above equation, we get,

$$\text{Torque} \quad T = \frac{\theta}{32}\left[\frac{\eta\pi d^4}{l}\right] \tag{1.66}$$

where d is the diameter of the shaft.

The couple per angular twist $\left(C = T/\theta\right)$ is

$$\text{Couple} \quad C = \frac{1}{32}\left[\frac{\eta\pi d^4}{l}\right] = \frac{1}{2}\left[\frac{\eta\pi r^4}{l}\right] \tag{1.67}$$

Equation (1.67) gives a couple per angular twist. Further, it is clear that the angle of twist is directly proportional to the length of the shaft.

1.13 TORSIONAL PENDULUM

Torsional pendulum is a pendulum which performs torsional oscillations. It is used to determine the moment of inertia and torsional rigidity of a given body. The compound and torsional pendulum differ from each other in terms of oscillations. The oscillations are twisted (torsional) in a horizontal plane in torsional pendulum, while they are linear in a compound pendulum.

A torsional pendulum consists of a circular disc D as shown in Fig. 1.9. The metal rod or wire is used to clamp the disc. The metal rod is suspended

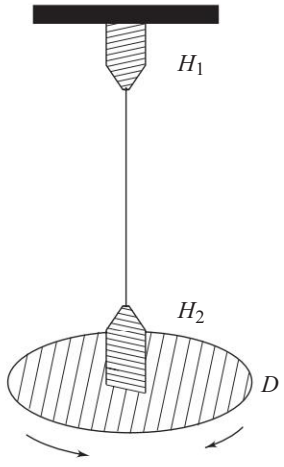


Fig. 1.9 *Torsional pendulum—torsion oscillation*

The McGraw·Hill Companies

symmetrically by clamping at the torsion head H_1. The length of torsion wire is adjusted employing the head H_2. In order to measure the time period, a pointer is marked at the disc. The length of pendulum, i.e., the distance between head H_2 and head H_1, is adjusted by using the head H_2. The head H_1 is not disturbed.

The disc D is rotated through a small angle and left free. As a result, the elements in the wire undergo shearing strain due to the twisting of wire when it is rotated. Therefore, restoring couple which is acting on the wire brings back to the normal original position. This results in torsional oscillation of the disc about the mean position.

Therefore, the restoring couple = $C\theta$, (1.68)

where C is the couple per unit twist.

Therefore, angular acceleration in the wire is due to the restoring couple.

$$a = \frac{d^2\theta}{dt^2} \tag{1.69}$$

If I is the moment of the wire about the axis,

$$I\left(\frac{d^2\theta}{dt^2}\right) = -C\,\theta \tag{1.70}$$

Equation (1.70) indicates that the angular acceleration ($d^2\theta/dt^2$) is directly proportional to angular displacement θ. The negative sign indicates that a decrease in twist on the wire with couple and direct towards the mean position. Therefore, torsional oscillations made by torsional pendulum are simple harmonic motion and the period of oscillation is controlled by the momentum of inertia of the suspended mass about the axis of suspension and a couple per unit twist produced in the wire. Therefore, time period of oscillator is

$$T = 2\,\pi\sqrt{\frac{\text{Displacement}}{\text{Acceleration}}} \tag{1.71}$$

Substituting the values of acceleration and displacement in the above equation, we get,

$$T = 2\,\pi\sqrt{\frac{1}{(C/I)}}$$

or

$$T = 2\,\pi\sqrt{\frac{I}{C}} \tag{1.72}$$

Equation (1.72) gives the time period of torsional oscillation of torsional pendulum.

1.13.1 Application of Torsional Pendulum

Torsional pendulum is mainly used to measure the moment of inertia of an irregular body and torsional rigidity modulus.

The McGraw·Hill Companies

(i) Determination of Moment of Inertia of a Body Torsional pendulum is used to determine the moment of inertia of an irregular body as shown in Fig. 1.10. The same is determined by taking a regular body with known moment of inertia. In the first step, the time period T_0 of the pendulum without any additional mass is determined. Therefore,

$$T_0 = 2\pi\sqrt{\frac{I_0}{C}} \tag{1.73}$$

where I_0 is the moment of inertia of the disc without any additional mass.

In the second step, the regular bodies with known momentum of inertia are placed at a distance d_1 away from the centre of the disc. The time period T_1 of the pendulum with regular mass at a distance d_1 from the centre of the disc is determined. Therefore,

$$T_1 = 2\pi\sqrt{\frac{I_0 + 2I_1}{C}} \tag{1.74}$$

where I_1 is the moment of inertia of the disc with a regular mass at a distance d_1 from the centre of the disc.

During the third step, the regular bodies are replaced by the irregular bodies at the same distance d_1 from the centre of the disc. The time period T_2 for a pendulum with irregular body is determined. Therefore,

$$T_2 = 2\pi\sqrt{\frac{I_0 + 2I_2}{C}} \tag{1.75}$$

where I_2 is the moment of the inertia of the disc with an irregular mass at a distance d_1 from the centre of the disc.

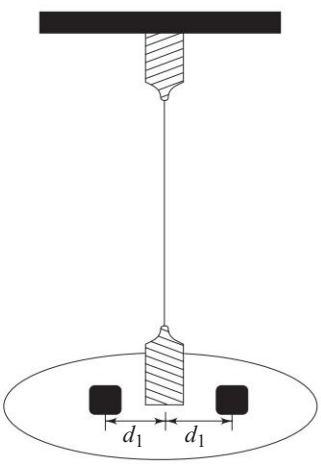


Fig. 1.10 *Torsional pendulum—moment of inertia*

Squaring and subtracting Eq. (1.73) from Eq. (1.74), we get,

$$T_1^2 - T_0^2 = \frac{4\pi^2 (I_0 + 2I_1)}{C} - \frac{4\pi^2 I_0}{C}$$

Simplifying the above equation, we get,

$$T_1^2 - T_0^2 = \frac{8\pi^2 I_1}{C} \tag{1.76}$$

Similarly, squaring and subtracting Eq. (1.73) from Eq. (1.75), we get,

$$T_2^2 - T_0^2 = \frac{4\pi^2 (I_0 + 2I_2)}{C} - \frac{4\pi^2 I_o}{C}$$

The McGraw·Hill Companies

Simplifying the above equation, we get,

$$T_2^2 - T_0^2 = \frac{8\pi^2 I_2}{C} \tag{1.77}$$

Dividing Eq. (1.76) by Eq. (1.77), we get,

$$\frac{T_1^2 - T_0^2}{T_2^2 - T_0^2} = \frac{(8\pi^2 I_1 / C)}{(8\pi^2 I_2 / C)} = \frac{I_1}{I_2} \tag{1.78}$$

Therefore, the moment of inertia of the irregular body is

$$I_2 = I_1 \frac{(T_2^2 - T_0^2)}{(T_2^2 - T_0^2)} \tag{1.79}$$

Equation (1.79) is used to determine the moment of inertia of an irregular body using torsional pendulum.

(ii) Determination of Torsional Rigidity Modulus One can determine the torsional rigidity modulus of the given body by measuring the time period of the disc under three conditions, namely, without any additional masses, two equal masses at a distance d_1 from the centre of the disc and two equal masses at a distance d_2 from the centre of the disc. The schematic representation of the pendulum is shown in Fig. 1.11.

The time period T_0 of the disc without additional mass is,
Therefore,

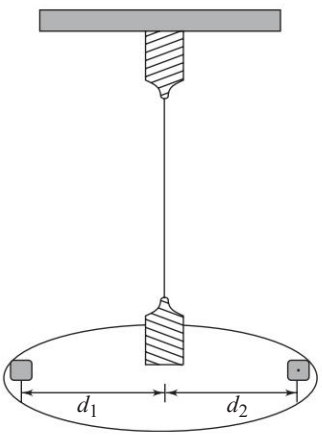


Fig. 1.11 *Torsional Pendulum—rigidity modulus*

$$T_0 = 2\pi\sqrt{\frac{I_0}{C}} \tag{1.80}$$

where I_0 is the moment of inertia of the disc and C, the couple per unit twist.
Let T_1 be the time period of the disc by placing the two equal masses on either side of the disc at a distance d_1 from the centre of the disc.
Therefore, time period of disc with two equal mass at the position d_1

$$T_1 = 2\pi\sqrt{\frac{I_0 + 2md_1^2}{C}} \tag{1.81}$$

where m is the mass of the object placed at a distance d_1.
Similarly let T_2 be the time period of disc by placing the two equal mass on either side at a distance d_2 from the centre of the disc.

Therefore, time period of disc with two equal mass at distance d_2,

$$T_2 = 2\pi\sqrt{\frac{I_0 + 2md_2^2}{C}} \tag{1.82}$$

Squaring Eqs (1.81) *and* (1.82), we get,

$$T_1^2 = 4\pi^2 \frac{(I_0 + 2md_1^2)}{C} \tag{1.83}$$

and

$$T_2^2 = 4\pi^2 \frac{(I_0 + 2md_2^2)}{C} \tag{1.84}$$

Subtracting the above equations, we get,

$$T_2^2 - T_1^2 = \frac{4\pi^2}{C} 2m\left(d_2^2 - d_1^2\right) \tag{1.85}$$

Substituting the value of couple from Eq. (1.60) in Eq. (1.85), we get,

$$T_2^2 - T_1^2 = \frac{4\pi^2 2m\ 2l}{\pi\eta a^4}(d_2^2 - d_1^2)$$

$$= \frac{16\pi ml}{\eta a^4}(d_2^2 - d_1^2)$$

Therefore, torsional rigidity $\eta = \dfrac{16\pi ml}{a^4}\dfrac{(d_2^2 - d_1^2)}{\left(T_2^2 - T_1^2\right)}$ (1.86)

Equation (1.86) gives the rigidity modulus of the given body using torsional pendulum.

1.14 BENDING OF BEAM

A beam is defined as a rod or a bar made of homogenous and isotropic elastic material with uniform cross-section (circular or rectangular). The length of a beam is very large compared to its other dimensions, namely, breadh and thickness. Thus, the shearing stress at any point of the beam is very small and negiliable. Generally, beams are set in a horizontal position with a fixed support at their two ends. The beams are used to carry heavy loads in applications like roofs, bridges, etc.

In order to study flexural rigidity of the beam for heavy loading applications, a simple theory is developed to study the bending of a beam. The following are the important points to be considered:

(i) The weight of the beam should be low.
(ii) The beam should have small curvature.
(iii) The minimum deflection of the beam is small compared to its length.
(iv) There are no shearing forces.
(v) The cross-section of the beam remains unaltered during bending. Hence, the geometrical moment of inertia of the beam remains constant.

The McGraw·Hill Companies

One end of the beam is fixed while a load is applied at the other end. As a result, longitudinal filament is extended at the convex side, while a contraction is obtained at the concave side. However, there is no change in the filament at the centre, i.e., in between the two sides. The central filament is in the neutral stage and hence, it is known as *neutral filament*. The corresponding central axis is known as *neutral axis*.

1.14.1 Bending Moment of a Beam

The beam *ABEF* is fixed at one end *AE*, while a load is applied at the other end *BF*, as shown in Fig.1.12. Consider a small portion of the beam *PP'BF*. Let *CD* be the neutral axis of the beam. It is clear from Fig 1.12, that the filament *PB* which lies above the neutral axis gets extended and experiences an inward pressure. On the other hand, the filament *P'F* which lies below the neutral axis gets shortened and experiences an outward pressure. Tensile and compressive stresses are developed in the respective halves of the beam due to inward and outward pressure in the upper and lower halves of the beam. The above stresses generate a couple and it opposes the bending of the beam. When the beam is at equilibrium, the strength of the bending moment is equal to the restoring couple or moment of resistance to the bending.

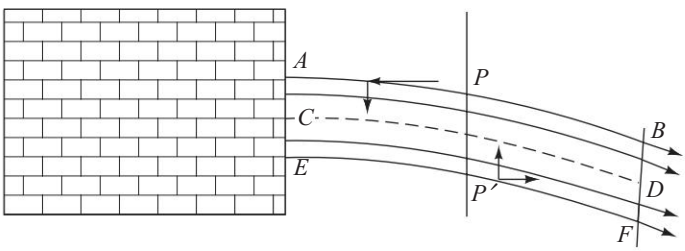


Fig. 1.12 *Bending moment—beam*

Consider a small portion of the beam *ABCD* as shown in Fig.1.13 to obtain the expression for the moment of restoring couple, i.e., bending moment. Consider that R is the radius of curvature of the beam and ϕ, the angle subtended by it at the centre of curvature. Let *GH* be the neutral filament at a distance r from the upper filament *AB*. $R\phi$ is the length of the filament (beam) under zero applied strain. When the strain is applied, the length of the filament is $(R + r)\phi$.

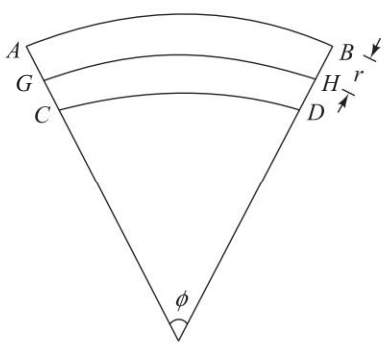


Fig. 1.13 *Bending moment of a beam*

Therefore, the strain at the filament *GH* is

$$\text{Strain} = \frac{\text{Change in length}}{\text{Original length}}$$

Substituting the values, we get,

$$\text{Strain} = (R + r)\,\phi - R\phi$$

$$R\varphi$$

$$\text{Strain} = \frac{r}{R} \tag{1.87}$$

It is clear from (1.87) that the strain is directly proportional to the distance *r* from the neutral axis.

Therefore, Young's modulus is $Y = \frac{\text{Stress}}{\text{Strain}}$

i.e., $\text{Stress} = \text{Y} \times \text{Strain}$

$$\text{Stress} = \text{Y}\ \frac{r}{R}$$

If *a* is the area of cross–section of the beam,

then, $$\text{Stress} = \frac{\text{Force}}{\text{Area}} \tag{1.88}$$

Therefore, $$\text{Force} = \text{Stress} \times \text{Area} \tag{1.89}$$

Substituting the values of stress and area in the above equation, we get,

$$F = Y\frac{r}{R}a \tag{1.90}$$

Therefore, moment of this force about $GH = Y\ \frac{r}{R}ar$ (1.91)

We know that the moment of the force acting on the upper and lower halves of the beam is same. Therefore, total moment of the force is

$$= \sum Y\frac{ar^2}{R} = \frac{Y}{R}\sum ar^2$$

$$= \frac{Y}{R}I_g \tag{1.92}$$

where I_g is known as the *geometrical moment of inertia* and is equal to $\Sigma ar^2 = Ak^2$, where *A* is the total area of the cross–section and *k* the radius of gyration.

Therefore, moment of the force $= \frac{Y}{R}I_g$ (1.93)

Under equilibrium conditions, when a load is applied at one, the bending moment of the beam is equal and opposite to the moment of bending couple.

The McGraw·Hill Companies

Therefore, bending moment of the beam $= \frac{Y}{R} I_g$ (1.94)

In Eq. (1.94), the term YI_g is called *flexural rigidity* of the beam. The flexural rigidity is defined as bending moment of the beam for unit radius of curvature.

1.15 CANTILEVER

A cantilever is a beam which is fixed horizontalies at one end and a load is applied at other end as shown in Fig.1.14. One can determine Young's modulus of the material of a cantilever by knowing the value of depression produced in the cantilever. Let W be the weight applied at one end of the cantilever. Consider PQ is the neutral axis of the cantilever.

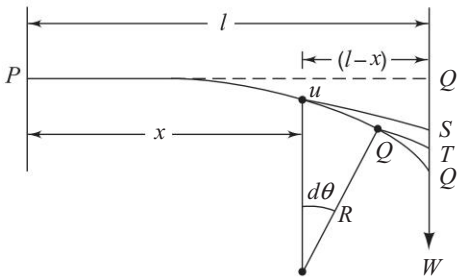


Fig. 1.14 *Cantilever*

When load W is applied at Q, the end Q is deflected to Q'. Thus, the neutral axis takes a new position PQ'. Consider a point U in the cantilever which is at a distance x from the fixed end P. The moment of the couple at this point U due to the load W is

$$W \quad UQ = W(l - x) \tag{1.95}$$

where l is the length of the cantilever and $(l - x)$ the distance of U from Q.

We know that under equilibrium conditions, the moment of the couple, i.e., bending moment is equal to the restoring couple. Therefore, from Eq. (1.95) and (1.94),

$$W(l - x) = \frac{YI_g}{R} \tag{1.96}$$

where Y is the Young's modulus of the rod of cantilever, I_g the geometrical moment of inertia and R the radius of curvature of the neutral axis.

Substituting the value of I_g, we get,

$$W(l - x) = \frac{Yak^2}{R} \tag{1.97}$$

We know that the moment of a couple increases as one moves from U towards the fixed end P. The radius of curvature is different at different points and it decreases as one approaches the point P. Consider a point V at a distance dx from U. The radius of curvature at V is same as that of at U, since the point V is very close to the point U. Therefore,

$$UV = Rd\theta = dx$$

Rearranging the above equation, we get,

$$R = \frac{dx}{d\theta} \tag{1.98}$$

Substituting the value of R from Eq. (1.98) in Eq. (1.97), we get,

$$W(l-x) = \frac{Yak^2}{dx} d\theta$$

Therefore, $$d\theta = \frac{W(l-x)dx}{YaK^2} \tag{1.99}$$

Draw tangents at U and V respectively at S and T which meets the vertical line QQ'. Therefore, the observed dispersion dy of V below U is

$$dy = (l-x)\, d\theta \tag{1.100}$$

Substituting Eq. (1.99) and Eq. (1.100), we get,

$$dy = \frac{W(l-x)^2}{YaK^2} dx \tag{1.101}$$

One can obtain the total depression of the beam Q' by integrating Eq. (1.101) employing the limit 0 to l.

Therefore, total depressions $$y = \int_0^l \frac{W(l-x)^2}{YaK^2} dx \tag{1.102}$$

Simplify the above equation, we get,

or $$= \frac{W}{YaK^2} \int_0^l [l^2 - 2lx + x^2] dx$$

or $$= \frac{W}{YaK^2} \left[l^2 x - 2\frac{lx^2}{2} + \frac{x^3}{3} \right]_0^l$$

$$y = \frac{1}{Y\, a\, K^2} \frac{W l^3}{3} = \frac{W l^3}{3 Y I_g}$$

Therefore, total depressions $$Y = \frac{W l^3}{3 Y I_g} \tag{1.103}$$

Equation (1.103) gives the depression value at the free end of cantilever.

1.15.1 Cantilever–Young's Modulus

Experimental arrangements used to determine the Young's modulus of a rod employing the cantilever is shown in Fig1.15. The given rod or bar whose Young's modulus is

The McGraw·Hill Companies

to be determined is clamped firmly and horizontally at one end. The other end of the rod is left free. A weight hanger is suspended at the free end. Further, a vertical pin is fixed firmly at the free end of the rod as shown in Fig1.15. A traveling microscope (T) is used to focus the tip the pin. The image of the tip of the pin is made to coincide with the horizontal position of intersection of the crosswire. The reading on the microscope is taken and tabulated as shown in Table 1.1. The load is increased gradually in a step-by-step manner. The corresponding readings for the pin are noted and tabulated. After applying the maximum load, the experiment is repeated by decreasing the load in a step-by-step manner in the same way as that of increasing the load. For decrease in each load, the corresponding microscope readings are taken. From the observed results, the depression for M kg, say 50×10^{-3} kg, is determined. The distance from the fixed end of the rod to the point at which the weight hanger is suspended is measured as l.

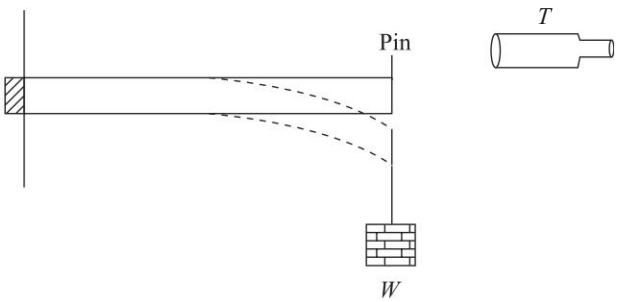


Fig. 1.15 *Cantilever—Young's modulus*

Table 1.1 *Depression for different applied loads*

Load	*Microscope reading*			*Depression, y for M*
	loading m	*unloading* m	*Mean* m	*kg wt in M*
W $W + 50$ $W + 100$				

Therefore, Young's modulus of cantilever is

$$Y = \frac{W l^3}{3 Y I_g}$$

$$Y = \frac{Mgl^3}{3yI_g} \tag{1.104}$$

Equation (1.104) is used to determine the Young's modulus of cantilever. The Young's modulus depends on the value I_g, i.e., the geometrical moment of inertia of

The McGraw·Hill Companies

the material. For example, the values of I_g, for a circular and rectangular dimension of the cantilever are different.

i.e., for circular dimension cantilever $Ig = \dfrac{\pi r^4}{4}$ (1.105)

Similarly, for rectangular dimension cantilever $I_g = \dfrac{bd^3}{12}$ (1.106)

where r is radius of the cantilever, b the breadth and d the thickness of the cantilever.

Therefore, one can determine the Young's modulus of a cantilever based on it dimension employing the value either form Eq. (1.105) and Eq (1.106).

1.15.2 Uniform Bending—Young's Modulus

Consider a cantilever is placed above the knife edge C and D as shown in Fig. 1.16. Consider that an equal load W is applied at the two ends. The bending of the beam is uniform and firms an arc of a circle. Therefore, bending is known as *uniform bending*. Therefore, the beam is experiencing a symmetrical load at the ends and hence, the reaction at each knife edge is equal to W.

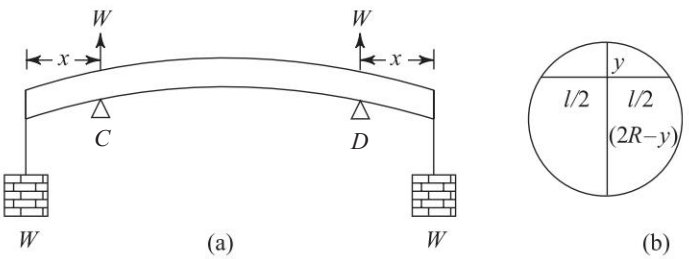


Fig. 1.16 *Uniform bending—Young's modulus*

Therefore, bending moment = $W x$ (1.107)

where x is the distance between the load and knife edge.

We know that, the restoring couple $= \dfrac{Y I_g}{R}$ (1.108)

where I, is the geometrical moment of inertia.

Under equilibrium condition, the bending moment is equal to restoring couple. Therefore, from Eq. (1.107) and Eq. (1.108), we get,

$$Wx = \frac{Y I_g}{R} \tag{1.109}$$

Consider the circle with the radius R as shown in Fig. 1.16b. Consider y is the elevation produced in the beam and l the length of the beam. Then, the centre of the beam is elevated through a distance (*Rule of Sagitha*)

i.e., $$\frac{l}{2}\,\frac{l}{2} = y(2R - y)$$

The McGraw·Hill Companies

or $$\frac{l^2}{4} = 2Ry - y^2 \quad (1.110)$$

where y is the elevation produced in the beam. The value of y^2 may be neglected since y is very small compared to the radius of curvature of circle.

Therefore, Eq. (1.110) can be written as,

$$\frac{l^2}{4} = 2Ry$$

$$R = \frac{l^2}{8y} \quad (1.111)$$

Substituting the value of $1/R$ in Eq. (1.109), we get,

$$Wx = \frac{Y I_g}{l^2} 8y$$

Therefore, Young's modulus $$Y = \frac{Wxl^2}{8yI_g} \quad (1.112)$$

Equation (1.112) is used to determine the Young's modulus of a beam bent uniformly.

1.15.3 Non-Uniform Bending –Young's Modulus

Consider a beam is supported symmetrically on two knife edges C and D as shown in Fig.1.17. At the centre of the beam, a load W is applied. The bending beam is a non-uniform bending, since the bending of beam does not describe an arc of a circle. A depression is produced at the centre depending on the applied load. Due to the application of load at the centre, the beam appears as two separate inverted cantilevers. It appears that the cantilevers are fixed at the centre E and loaded at the ends each with a load $W/2$. The length of the cantilever is equal to $l/2$.

One can determine Young's modulus of the bar under non-uniform bending by substituting $W = W/2$ and $l = l/2$ in Eq. (1.112),

$$Y = \frac{\frac{W}{2}\left(\frac{l}{2}\right)^3}{3YI_g}$$

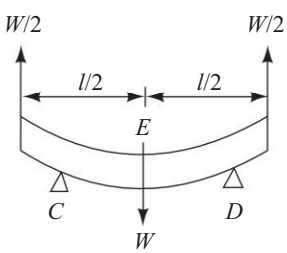


Fig. 1.17 *Non-uniform bending—Young's modulus*

Therefore, Young's modulus $Y = \dfrac{Wl^3}{48\, yI_g}$ (1.113)

Consider that the cantilever is a rod of uniform cross-section.
Therefore, geometrical moment of inertia is,

$$I_g = aK^2$$

$$= \pi r^2 \left(\frac{r}{2}\right)^2$$

$$I_g = \frac{\pi r^4}{4} \quad (1.114)$$

Similarly, if the cantilever is a rectangular bar, then the geometrical moment of inertia is

$$Y = \frac{Wl^3}{12 y \pi r^4} \quad (1.115)$$

Therefore, Young's modulus of the beam for the circular cross-section is,

$$I_g = \frac{bd^3}{12} \quad (1.116)$$

Therefore, Young's modulus of the beam of rectangular bar is

$$Y = \frac{W\, l^3}{4 y\, b\, d^3} \quad (1.117)$$

Equations (1.115) and (1.17) are used to determine Young's modulus respectively of the rod of uniform cross-section and a beam of rectangular bar.

1.154 Applications—I shaped Girder

Consider an *I*-shaped girder is supported at its bends. The weight of the girder is very high and hence, it may bend at its centre. Thus, the I-shaped girder is an example of non-uniform bending.

The depression produced in a non-uniform bending is $y = \dfrac{Mgl^3}{48Y\, I_g}$ (1.118)

where M is the mass of the applied load, l the length of the grids, Y the Young's modulus and I_g the geometrical moment of inertia of the girder.
The geometrical moment of inertia of a girder having a breath b and width d is

$$I_g = \frac{bd^3}{12} \quad (1.119)$$

Therefore, the depression produced in a girder is

$$y = \frac{Mgl^3}{4Ybd^3} \quad (1.120)$$

The McGraw·Hill Companies

Equation (1.121) gives the value of Young's modulus of I-shaped girder. The depression y is inversely proportional to d^3. Therefore, a small increase in the width results a small depression. In order to obtain the same quantity of depression, for a particular value of d, the value of breadth b of the girder should be increased to three times as that of d. On the other hand, to obtain the depression as that of the value of the breadth, the value of the width should be increased three times. Therefore, the girder with minimum depression looks like an I-shaped girder. The I-shaped girder is shown in Fig.1.18.

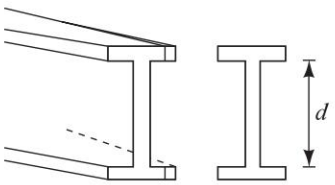


Fig. 1.18 *I-shaped girder*

Keypoints to Remember

- Elastic materials are classified as two types, namely, perfectly elastic and plastic.
- The material which regains its original state after removal of the deforming force is known as a perfectly elastic material.
- The material which does not regain its original state even after removal of deforming force is known as a plastic material.
- The restoring force, or recovering force, per unit area is known as stress.
- The ratio of the change in dimension to its original dimension is known as strain.
- When the stress is remove d, a small amount of strain is present in the material known as residual strain.
- Modulus are classified into three types, namely, Young's, bulk and rigidity modulus.
- The ratio of longitudinal stress to linear strain within the elastic limit is known as Young's modulus.
- Young's modulus of the linear wire is $Y = mgL/\pi r^2 \Delta L$, where m is the mass, g the acceleration due to gravity, r the radius of the wire, L the length of the wire and ΔL the elongation in the length of the wire.
- The ratio of the bulk stress to bulk strain is known as bulk modulus $K = \dfrac{-\Delta pV}{\Delta V}$
- Rigidity modulus is the ratio of the shearing stress to shearing strain i.e., $\eta = FLIA\ \Delta L$
- To ratio of the longitudinal strain per unit stress and lateral strain per unit stress is known as Poisson's ratio $\sigma = \beta/\alpha$
- The relationship between Young's and bulk modulus is $K = \dfrac{Y}{3(1-2\sigma)}$

- The relationship between rigidity and Young's modulus is $\eta = \frac{Y}{2(1+\sigma)}$
- The relationship between the three elastic modulus is $\frac{9}{Y} = \frac{1}{K} + \frac{3}{\eta}$
- The twisting couple per unit angular twist is known as angular twist.
- The couple per unit twist is equal to $\frac{1}{2}\frac{\pi \eta a^4}{l}$
- Time period of torsional oscillations is $T = 2\pi\sqrt{\frac{I}{C}}$
- The moment of inertia of an irregular body is $I_2 = I_1 \frac{(T_2^2 - T_0^2)}{(T_1^2 - T_0^2)}$
- The bending momention of a beam is equal to $\frac{y}{R} I_g$
- The depression produced by a cantilever at the free end is $\frac{Mgl^3}{3Y I_g}$
- Young's modulus of cantilever is equal to $\frac{Mgl^3}{3Y l_g}$
- Young's modulus of uniform bending beam is equal to $\frac{W x l^2}{8Y I_g}$
- Young's modulus of non-uniform bending beam is $\frac{W l^3}{4Y b d^3}$
- The depression produced by girder is equal to $\frac{M g l^3}{4Y b d^3}$

Solved Problems

Example 1.1 A copper wire, 2 m long and 6.25×10^{-5} m^2 in cross-section, is found to stretch 3×10^{-3} m under a tension of 1×10^3 N. What is the Young's modulus of the material of the wire?

Given Data

Length of the wire $L = 2$ m
Cross-sectional area $A = 7.25 \times 10^{-5}$ m^2
Increase in length $\Delta L = 3 \times 10^{-3}$ m
Tension $T = 1 \times 10^3$ N

Solution We know that, Young's modulus

$$Y = \frac{F L}{A \Delta L}$$

The McGraw·Hill Companies

Substituting the values in the above equation, we get,

$$Y = \frac{1\times 10^3 \times 2}{6.25 \times 10^{-5} \times 3 \times 10^{-3}} = 9.195 \times 10^9$$

The Young's modulus of the wire Y is 1.92×10^{10} N m^{-2}.

Example 1.2 A steel wire with 2.75 m length and 1×10^{-3} m diameter is attached to a beam at its upper end. If a load of 1 kg is suspended from the lower end, how much will the wire be extended? (Young's modulus = 2×10^{11} N m^{-2}).

Given Data

Young's modulus of the wire	$Y = 2 \times 10^{11}$ N m^{-2}
Length of the wire	$L = 2.75$ m
Diameter of the wire	$d = 1 \times 10^{-3}$ m
Applied load	$M = 1$ kg

Solution We know that, tension $T = Mg$

Substituting g (acceleration due to gravity) and M values in the above equation, we get,

$$T = 1 \times 9.8 \text{ N}$$

Therefore, tension $T = 9.8$ N

We know that, change in length $\Delta L = \dfrac{FL}{AL} = \dfrac{FL}{\pi r^2 Y}$

Substituting the values in the above equation, we get,

$$\Delta L = \frac{9.8 \times 2.75}{\pi \times (0.5\times 10^{-3})^2 \times 2 \times 10^{11}}$$

$$= 3.43138 \text{ x } 10^{-4}$$

The increase in length of wire is 3.43138 x 10^{-4}.

Example 1.3 A rectangular solid has dimensions $6 \times 6 \times 2$ cm. A force of 0.2 N, applied tangentially to the upper surface as shown in Fig. 1.19, causes a displacement of 5 mm relative to the lower surface. Calculate the shearing strain, stress and the rigidity modulus,

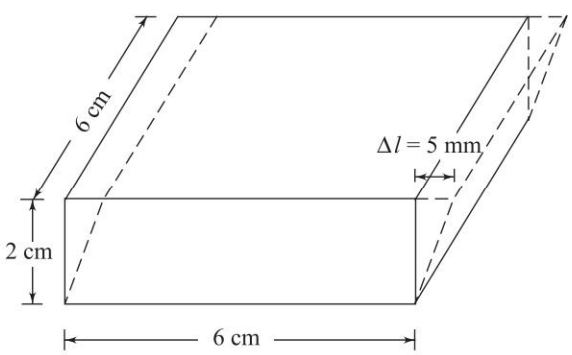


Fig. 1.19 *Solid rectangle*

Given Data

Force $= 0.2$ N

Displacement $= 5 \times 10^{-3}$ m

Length of the solid $= 6 \times 10^{-2}$ m

Breadth of the solid $= 6 \times 10^{-2}$ m

Height of the solid $= 2 \times 10^{-2}$ m

Solution We know that, shear stress $= \dfrac{\text{Force}}{\text{Area}}$

Substituting the values in the above equation, we get,

$$\text{Shear stress} = \frac{0.2\,N}{6 \times 6 \times 10^{-4}} = 55.55\,N\,m^{-2}$$

$$\text{Shear strain, } \theta = \frac{\text{Displacement of the upper layer}}{\text{Height}}$$

Substituting the values in the above equation, we get,

$$\theta = \frac{5 \times 10^{-3}}{2 \times 10^{-2}} = 0.25$$

Therefore, Shear stress $= 0.25$.

$$\text{The rigidity modulus} = \frac{\text{Tangential stress}}{\text{Shear strain}}$$

Substituting the values in the above equation, we get,

$$= \frac{55.55}{0.25} = 222.22$$

Therefore, Shearing stress $= 55.55$ N m^{-2}

Shear strain $= 0.25$

Rigidity modulus $= 222.22$ N m^{-2}.

Example 1.4 Calculate the shearing force required to distort a block of aluminium by 1.5 cm, if its thickness is 30 cm and its surface area is 12 cm^{-2}. The shear modulus for aluminium is 2.5×10^{10} N m^{-2}.

Given Data

Rigidity modulus $= 2.4 \times 10^{10}$ N m^{-2}

Thickness $= 30 \times 10^{-2}$ m

Surface area $= 10 \times 10^{-4}$ m^2

Displacement $= 1.5 \times 10^{-2}$ m

Solution We know that, rigidity modulus $\eta = \dfrac{F\ L}{A\ \Delta\ L}$

Therefore, the shearing forces $F = \dfrac{\eta\ A\ \Delta\ L}{L}$

Substituting the values in the above equation, we get,

$$F = \frac{2.4\times10^{10}\times10\times10^{-4}\times1.5\times10^{-2}}{30\times10^{-2}}$$

$$= 1.25\times10^{6}\ \text{N}$$

The shearing force required, F is 1.25×10^{6} N.

Example 1.5 (a) show that the bulk modules K, Young' modulus Y and Poisson's ratio s are connected together by the relation,

$$K = \frac{Y}{3(1-2\sigma)}$$

(b) Show that the rigidity modulus n and Young's modulus Y are connected by the relation,

$$\eta = \frac{Y}{2(1+\sigma)}, where\ \sigma\ is\ the\ Poisson's\ ratio.$$

Solution (a) The relation between K and α can be written as,

$$K = \frac{1}{3(\alpha-2\beta)}$$

or

$$K = \frac{1}{3\alpha\left(\left(1-2\dfrac{\beta}{\alpha}\right)\right)}$$

(b) The relation between η and α can be written as,

$$\eta = \frac{1}{2(\alpha+\beta)}$$

or

$$K = \frac{1}{3\alpha\left(1-\dfrac{\beta}{2\alpha}\right)} = \frac{Y}{3(1-\sigma)}$$

where$y = 1/\alpha$ and $\sigma = \beta/\alpha$.

Example 1.6 A wire with 3 m length for which Young's modulus is 10×10^{10} N m^{-2}, has a 1 mm diameter. If a load of 10 kg is attached to one end, how much is the extension produced? If Poisson's ratio is 0.26, what is the lateral compression produced?

Given Data

Length of the Cu wire $L = 3$ m

Young's modulus $Y = 10 \times 10$ N m^{-2}

Radius of the wire $r = 5 \times 10^{-4}$ m (Since d = 1 mm)

Poisson's ratio a = 0.26

Solution We know that, Young's modulus $Y = \dfrac{FL}{A\Delta L}$

Therefore $$\Delta L = \frac{FL}{AY} = \frac{MgL}{\pi r^2 Y}$$

Substituting F, A, Y and L values in the above equation, we get,

$$\Delta L = \frac{10 \times 9.8 \times 3}{\pi \times \left(5 \times 10^{-4}\right)^2 \times 10 \times 10^{10}}$$

$$\Delta L = 3.75 \times 10^{-3} \text{ m}$$

Poisson's ratio $\sigma = \dfrac{\text{Lateral strain}}{\text{Longitudinal strain}}$

Therefore, lateral strain $= \sigma \times$ longitudinal strain

$$= \sigma \times \frac{\Delta L}{L}$$

Substituting σ, ΔL and L values in the above equation we get.

$$\text{lateral strain} = \frac{0.26 \times 3.75 \times 10^{-3}}{3}$$

$$= 3.25 \times 10^{-4}$$

This gives the value of lateral strain, that is, ratio of decrease in diameter to original diameter of the wire.

$$\frac{\text{Decrease in diameter}}{\text{Original Diameter}} = 3.25 \times 10^{-4}$$

Therefore, decrease in diameter $= 3.25 \times 10^{-7} \times 1 \times 10^{-3}$

$$= 3.25 \times 10^{-7} \text{ m}$$

Extension produced is 3.25×10^{-3} m

Lateral compression produced is 3.25×10^{-7} m

Example 1.7 What couple must be applied to a 1 m long wire with 1-mm diameter, in order to twist one end of it through 90°, the other end remaining fixed? The rigidity modulus is 2.8×10^{10} N m^{-2}.

The McGraw·Hill Companies

Given Data

Length of the wire $L = 1$ m

Diameter of the wire $= 1 \times 10^{-3}$ m

Rigidity modulus of the wire $\eta = 2.8 \times 10^{10}$ N m^{-2}

Angle of twisting $= 90^{\circ} = \pi/2$ radian

Solution The couple required to twist the wire through 90º

$$C = \frac{\pi \eta r^4}{2l} \phi$$

$$C = \frac{\pi \eta r^4}{2l} \frac{\pi}{2} = \frac{\pi^2 \eta r^4}{4l}$$

Substituting n, r and l values in the above equation, we get,

$$C = \frac{\pi^2 \times 2.8 \times 10^{10} \times \left(5 \times 10^{-4}\right)^{-4}}{4 \times 1} = 4.3179 \times 10^{-3}$$

The couple to be applied is 4.3179×10^{-3} N m.

Example 1.8 A 0.60 mm thick gold wire elongates by l mm when stretched by a force of 160 g-wt and twist through 1 radian when equal and opposite torques of 10×10^{-5} N are applied at its end. Find the value of Poisson's ratio. Given $\eta = 2.34 \times 10^{9}$ N m^{-2}.

Given Data

Diameter of the wire $= 0.60 \times 10^{-3}$ m

Length of elongation produced $= 1 \times 10^{-3}$ m

Mass $= 0.16 \times 10^{-3}$ Rg

The angular twist $= 1$ radian

The torque $= 10 \times 10^{-5}$ N

Solution We know that, Young's modulus $Y = \frac{F L}{A \Delta L} = \frac{M g L}{A \Delta L}$

Substituting M, g, ΔL, L and A values in the above equation, we get,

$$Y = \frac{0.16 \times 9.8 \times L}{\pi \times \left(0.35 \times 10^{-3}\right)^2 \times 10^{-3}}$$

$$= 6.238 \times 10^{9} \text{ L N m}^{-2}$$

Young's modulus, bulk modulus and Poisson's ratio are related by the relation,

$$\eta = \frac{Y}{2(1+\sigma)}$$

Therefore, Poisson's ratio can be written as

$$\sigma = \frac{Y}{2\eta} - 1$$

Substituting Y and n values in the above equation, we get,

$$\sigma = \frac{5.199 \times 10^9}{2 \times 2.26 \times 10^9} - 1 = 1.3329 - 1$$

The Poisson's ratio is 0.3329.

Example 1.9 Calculate the twisting couple on a solid cylindrical shaft of length 1.5 m and radius 80 mm when it is twisted through an angle 0.6º. The rigidity modulus of the material of the shaft is 93 $\times 10^9$ N m^{-2}. If the above solid shaft is converted into a hollow cylindrical shaft with an inner radius of 60 mm and outer radius of 100 mm and the same length, find out the twisting couple on the cylindrical shaft (hollow), when it is twisted through the same angle.

Given Data

The length of the cylindrical shaft	$l = 1.5$ m
The radius of the cylindrical shaft	$r = 80 \times 10^{-3}$ m
The rigidity module of the shaft	$\eta = 93 \times 10^9$ N m^{-2}
The inner radius of the hollow shaft	$r_1 = 60 \times 10^{-3}$m
The outer radius of the hollow shaft	$r_2 = 100 \times 10^{-3}$ m
The angle of twisting	$\phi = 0.6$º or $\pi/300$

Solution We know that, the couple required to twist the cylindrical shaft through 0.6 (π/300)

$$C = \frac{\pi \eta r^4}{2l}\phi$$

$$C = \frac{\pi \eta r^4}{2l}\frac{\pi}{300}$$

$$= \frac{\pi^2 \eta r^4}{600\, l}$$

The McGraw·Hill Companies

Substituting the values, we get,

$$C = \frac{3.14^2 \times 93 \times 10^9 \times (80 \times 10^{-3})^4}{600 \times 1.5}$$

$$= 41.73 \times 10^3 \text{ N m.}$$

Similarly, we know that, the couple required to twist the hollow cylindrical shaft through 0.6 (π/300)

$$C = \frac{\pi\eta(r_2^4 - r_1^4)}{2l}\phi$$

$$= \frac{\pi\eta(r_2^4 - r_1^4)}{2l}\pi / 300$$

$$= \frac{\pi^2\eta(r_2^4 - r_1^4)}{600l}$$

Substituting the values, we get,

$$C = \frac{3.14^2 \times (93 \times 10^9) \times (100 \times 10^{-3})^4 - (60 \times 10^{-3})^4}{600 \times 1.5}$$

$$= 88.678 \times 10^3$$

Thus, the twisting on a solid cylindrical shaft is 41.73 x 10^3 Nm.

The couple twisting on a hollow cylindrical shaft is 88.678 Nm.

Example 1.10 By what fraction will the volume of a steel bar increase as the air is evacuated to 10^2 N m^{-2} from a chamber in which it rests? Standard atmospheric pressure is 1.01 x 10^5 N m^{-2} and the bulk modulus of steel is 16×10^{10} N m^{-2}.

Given Data

Standard atmospheric pressure $= 1.01 \times 10^5$ N m^{-2}

Bulk modulus $K = 16 \times 10^{10}$ N m^{-2}

The change in pressure $= 1.01 \times 10^5 - 10^2$ N m^{-2}

$= 1.009 \times 10^5$ N m^{-2}

Solution We know that, Bulk modulus $K = \frac{-V}{dV}dp$

Therefore, $dV = \frac{-V}{K}dp$

The negative sign indicates that volume decreases when pressure increases. Therefore, the above equation can be written as,

$$\frac{dV}{V} = \frac{dp}{K}$$

Substituting dp and K values of above equation, we get,

$$\frac{dV}{V} = C = \frac{1.009 \times 10^5}{16 \times 10^{10}} = 6.3 \times 10^{-7}$$

The fractional change of volume is 6.3×10^{-7}
The change in volume of steel bar is $6.3 \times 10^{-7} \times V\ m^3$

Example 1.11 A rectangular bar with 2 cm breadth, 1 cm depth and 100 cm length is supported at its ends and a load of 2 kg is applied at its midpoint. Calculate the depression, if the Young's modulus of the material of the bar is 1.013×10^{10} N m^{-2}.

Given Data

Young's modulus of the bar, Y $= 1.025 \times 10^{10}$ N m^{-2}

Breadth of the bar, b $= 2 \times 10^{-2}$ m

Length of the bar, l $= 1$ m

Depth of the bar, d $= 1 \times 10^{-2}$ m

Load applied, M $= 2$ kg

Solution We know that, the depression produced in the bar, $Y = \dfrac{Wl^3}{4Ybd^3}$

Substituting the values of Y, W, b, d and l, we get

$$Y = \frac{2 \times 9.8 \times 1^3}{4 \times 1.25 \times 10^{10} \times 2 \times 10^{-2} \times \left(1 \times 10^{-2}\right)^3}$$

$$Y = 0.0239 \text{ x } 10^{-2} \text{ m}$$

The depression produced in the bar is 2.39×10^{-2} m

Example 1.12 A circular cantilever with 1.2 cm radius and 1.5 m length is fixed at one end. At the other end, a load of 2 kg is applied. The Young's modulus of the cantilever is 19.5×10^{10} N m^{-2}. Find the depression produced.

Given Data

Radius of the cantilever, r $= 1.0$ cm $= 1.2 \times 10^{-2}$ m

Length of the cantilever, l $= 1.5$ m

Young's modulus, Y $= 19.5 \times 10^{10}$ N m^{-2}

Solution The depression produced in the cantilever, $y = \dfrac{4wl^3}{3Y\pi r^4}$

Substituting the values of Y, w, r and l, we get

The McGraw·Hill Companies

$$y = \frac{4 \times 2 \times 9.8 \times 1.5^3}{3 \times 19.5 \times 10^{10} \times \pi \left(1.2 \times 10^{-2}\right)^4}$$

$$y = 6.943 \times 10^{-3} \text{ m}$$

The depression produced in the cantilever is 6.943×10^{-3} m

Objectives Questions

1.1 The restoring force per unit area is known as _____.

1.2 The modulus of elasticity is

a) Stress × Strain b) Stress × E

c) Stress / Strain d) Strain / Stress

1.3 The applied stress is within the _____ limit, it returns to its original state.

1.4 The linear strain is _____.

1.5 Young's modulus of the linear wire

a) $Y = \dfrac{M g L}{\pi r^2 \Delta L}$ b) $Y = \dfrac{M g L^2}{\pi r^2 \Delta L}$

c) $Y = \dfrac{mMg L}{2\pi r^2 \Delta L}$ d) $Y = \dfrac{M g L}{\pi r^2 \Delta L^2}$

1.6 The ratio of the lateral strain to linear strain is known as _____.

1.7 Young's and bulk modulus of the material is related by the equation

a) $K = \dfrac{Y}{3(1+2\sigma)}$ b) $K = \dfrac{Y}{3(1-2\sigma)}$

c) $K = \dfrac{\eta}{3(1+2\sigma)}$ d) $k = \dfrac{\eta}{3(1+2\sigma)}$

1.8 The rigidity and Young's modulus are related by the equation

a) $\eta = \dfrac{Y}{2(1-\sigma)}$ b) $\eta = \dfrac{Y}{2(1-\sigma^2)}$

c) $\eta = \dfrac{1}{2(1-\sigma)}$ d) $\eta = \dfrac{Y}{2(1+\sigma)}$

1.9 The three modul are connected by the relation $\dfrac{9}{Y} = \dfrac{1}{K} +$ _____.

1.10 The restoring couple per unit twist is equal to

a) $\dfrac{1}{2}\dfrac{\pi \eta a^4}{l}$ b) $\dfrac{1}{2}\dfrac{\pi^2 \eta a^4}{l}$

c) $\dfrac{1}{2}\dfrac{\pi \eta^2 a^4}{l}$ d) $\dfrac{1}{2}\dfrac{\pi \eta a^4}{l^2}$

1.11 The moment of inertia of an irregular body is equal to

a) $I_2 = I_1 \dfrac{T_2^2 - T_0^2}{T_2 - T_0}$ b) $I_2 = I_1 \dfrac{T_0^2 - T_1^2}{T_2^2 - T_0}$

c) $I_2 = I_1 \dfrac{T_2 - T_0}{T_1 - T_0}$ d) $I_2 = I_1 \dfrac{T_2^2 - T_0^2}{T_1^2 - T_0^2}$

1.12 The rigidity modulus of the given body employing the torsional pendulous is

a) $\eta = \dfrac{16\pi ml}{a^4} \dfrac{(d_2^2 - d_1^2)}{(T_2^2 - T_1^2)}$ b) $\eta = \dfrac{16\pi a^4 l}{m} \dfrac{(d_2^2 - d_1^2)}{(T^2 - T_1^2)}$

c) $\eta = \dfrac{16\pi ml^2}{a^4} \dfrac{(d_2^2 - d_1^2)}{(T_2^2 - T_1^2)}$ d) $\eta = \dfrac{16\pi}{a^4 ml} \dfrac{(d_2^2 - d_1^2)}{(T_2^2 - T)}$

1.13 The bending moment of the beam is equal to

a) $Y R I_g$ b) $\dfrac{Y}{R} I_g$

c) $Y R / I_g$ d) $\dfrac{Y}{R} I_g^2$

1.14 At the free end of the cantilever, a depression y is produced and is equal to _____.

1.15 Young's modulus of the cantilever is equal to

a) $\dfrac{Mgl^3}{3y I_g}$ b) $\dfrac{Mgl^2}{3y I_g}$

c) $\dfrac{Mgl^3}{3y I_g^2}$ d) $\dfrac{Mgl^3}{3y^2 I_g}$

1.16 Young's modulus of the uniform bending beam is

a) $\dfrac{W x l^2}{8 y Ig}$ b) $\dfrac{W x^2 l^2}{8 y Ig}$

c) $\dfrac{W x l^3}{8 y Ig}$ d) $\dfrac{W x^2 l^2}{8 y Ig}$

1.17 $\dfrac{W l^3}{4 y b d^3}$ is the Young's modulus of the _____ bending beam.

1.18 _____ is the depression produced by a girder.

KEY

1.1. Stress. 1.2. a)
1.3. Elastic limit 1.4. $(\Delta L / L)$
1.5. a) 1.6. Poisson's ratio
1.7. b) 1.8. d)
1.9. $3/\eta$ 1.10 a)

The McGraw·Hill Companies

1.11. d) 1.12 a)

1.13. b) 1.14 $y = \dfrac{M g l^3}{3YI_g}$

1.15. a) 1.16 a)

1.17. non-Uniform 1.18 $\dfrac{M g l^3}{4Yba^3}$

Exercises

1.1 A steel wire with 3-m length and 1-mm^2 cross-section is held between rigid supports with a tension of 50 N. If the middle of the wire is pulled 5 mm sideways, calculate the change in tension. Also calculate change in tension if temperature changes by 10 °*C*. (For steel, $Y = 2 \times 10^{11}$ N m^{-2} and $a = 10^{-5}$ C^{-1}).

1.2 While at 0ºC, a square steel bar of 2 cm side is fixed at both ends so that it cannot expand. If the temperature is raised to 20ºC, what force would be exerted on the support at the ends? Given $Y = 2 \times 10^{11}$ Nm^{-2} and coefficient of linear expansion of steel is 10^{-5} per degree.

1.3 A circular bar, with one-metre length and 0.8 cm diameter, is rigidly clamped at one end in a vertical position. A couple of magnitude 0.035 N m^{-1} is applied at the other end. As a result, a mirror fixed at this end deflects a spot of light by 40 cm on a scale 1 m away. Determine the rigidity modules of the wire.

1.4 A steel wire with 2.75 m length and 1×10^{-3} m diameter is attached to a beam at its upper end. If a load of 1 kg is suspended from the lower end, how much will the wire be extended? (Young's modulus = 2×10^{11} N m^{-2}).

1.5 A wire of a radius *r,* stretched without tension along a straight line is lightly fixed at *A* and *B* as shown in Fig. 1.20. What is the tension in the wire when it is pulled into the shape *ACB*? Assume Young's modules of the material of the wire to be *Y*.

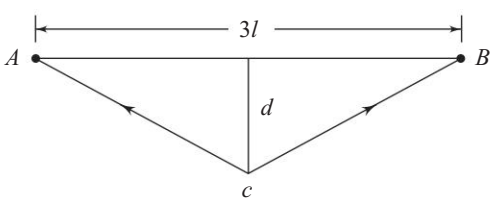


Fig. 1.20

1.6 A material has a Poisson's ratio of 0.25. If a uniform rod of the material suffers longitudinal strain of 24×10^{-3}, calculate the percentage change in its volume, if one end of a wire with 2-mm radius and 1.5 m length is twisted through 450. Calculate the angle of shear on its surface.

The McGraw·Hill Companies

1.7 A wire with 0.3 mm diameter elongates by 1mm when stretched by a force of 0.33 Kg-wt and twists through 90° when equal and opposite torque of 2.15×10^{-5} N m is applied at its ends. Find the Poisson's ratio for the wire.

1.8 A sphere with 800-g mass and 3-cm radius is suspended from a wire with 100-cm length and 0.5-mm radius. If the period of torsional vibration is 1.23 s, calculate the rigidity of the material of the wire.

1.9 A solid ball with 3.3 m diameter is submerged in a lake at such a depth that the pressure exerted by water is 1 N m^{-2}. Find the change in volume of the ball. Given $K = 1.00 \times 10^7$ N m^{-2}.

1.10 A cantilever with 3 cm breadth, 10 mm width and 1.2 m length is fixed at one end. A load of 1.7 kg is applied in the other end. If the Young's modulus of the cantilever is 11.02×10^{10} Nm^{-2}, find the depression produced.

1.11 A cantilever is supported at two ends. The cantilever is a bar with 3.5 cm breadth, 1.2 cm width and 1 m length. If the depression produced in the cantilever is 0.6 mm, find the Young's modulus of the cantilever.

Short Questions

1.1 Define elasticity.
1.2 Distinguish between elastic materials and plastic materials.
1.3 Define the terms stress and strain.
1.4 State and explain Hooke's law of elasticity.
1.5 What is the effect of temperature on elasticity of a material?
1.6 What is the elastic limit?
1.7 What are the three different types of elastic moduli?
1.8 What is linear strain?
1.9 Define the term Young's modulus.
1.10 What is volume strain?
1.11 Define the term 'bulk modulus'.
1.12 What is meant by shear strain?
1.13 Explain the term 'rigidity modulus'.
1.14 Explain the term 'lateral strain'.
1.15 Define the term Poisson's ratio.
1.16 Write down the relation between Y and K and explain the various terms.
1.17 Write down the relation between Y and η and explain the various terms.
1.18 Write down the relation between Y, K and η and explain the various terms.
1.19 What are the limiting values of σ?
1.20 What is meant by torsional pendulum?
1.21 What are the applications of the torsional pendulum?
1.22 What is a cantilever?
1.23 What do you mean by bending moment of a beam?
1.24 What is geometrical moment of inertia?
1.25 Distinguish the uniform and non-uniform bending in a cantilever.
1.26 What is meant by bending moment of a beam?

The McGraw·Hill Companies

1.27 Distinguish the uniform bending and non-uniform bending.
1.28 What is an I-shaped girder?
1.29 How will you reduce the depression in an I-shaped girder?

Descriptive Questions

1.1 Describe about the elastic behaviour of materials.
1.2 Discuss the three types of elastic modulus.
1.3 What are the different types of elastic modulii? Derive the necessary mathematical relation for them.
1.4 Define Young's modulus, bulk modulus and rigidity modulus. If Y, K and η represent these modulii, respectively, prove the relation

$$Y = \frac{9\eta K}{3K + \eta}$$

1.5 Prove that Young's modulus Y, bulk modulus K and rigidity modulus η are related by the relation:

$$\frac{9}{Y} = \frac{1}{K} + \frac{3}{\eta}$$

1.6 Show that Young's modulus, bulk modulus K and rigidity modulus η are related with σ by the following relations given below.

(a) $\frac{Y}{2\eta} = 1 + \sigma$

(b) $\frac{Y}{3K} = 1 - 2\sigma$

1.7 Derive the expression for a couple per angular twist of a twisted cylindrical wire.
1.8 Derive the expression for the time period of a torsional pendulum.
1.9 Derive the expression for the moment of inertia of a disc and rigidity modulus of a cylindrical wire using torsional pendulum.
1.10 How will you determine the rigidity modulus of the wire using torsional pendulum?
1.11 What is a cantilever? Derive an expression to find the depression in a cantilever fixed at one end and loaded at the other end and explain an experimental method used to find the Young's modulus of a cantilever.
1.12 What do you mean by non-uniform bending? Derive an expression for the Young's modulus of a bar length 'l' m width,'d' m and breadth, 'b' non-uniformly and explain how the expression for the Young's modulus of a rectangular beam is determined using non-uniform bending.

Chapter

Viscosity

2

OBJECTIVES

- To understand viscosity and the properties of liquids
- To derive the equation for the coefficient of viscosity
- To explain the role of Poiseuille's equation in liquid flow
- To determine the velocity of a rigid body moving in a viscous medium
- To explore the experimental procedure to determine the viscosity of a liquid
- To explain the importance and application of Stoke's law

2.1 INTRODUCTION

Consider that a beaker containing water is stirred for few seconds and left to itself. The motion of water subsidies after some time. It is due to an internal resisting force acting inside the liquid parallel to its flow and is known as viscosity. It is an inherent property of a liquid.

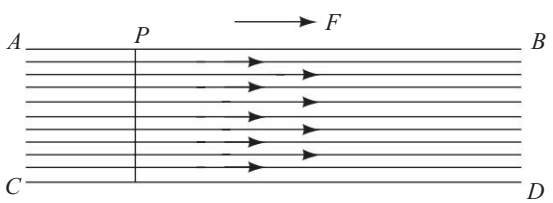


Fig. 2.1 *Flow of liquid between the layers AB and CD*

Consider *AB* and *CD* represent two plates and are separated by a small distance. Consider the liquid present between the two plates is imagined to be divided into different layers as shown in Fig. 2.1. If a force *F* is acting on the upper layer then the upper layer of the liquid is slightly displaced compared to the layer that lies just below it. Hence, the upper layer moves with a constant velocity after exerting the initial acceleration. The next layer moves with a constant velocity slightly less than that of the upper layer. The bottom layer, which is in contact to the plate *CD*, experiences zero velocity. The different layers of the liquid may be regarded as a pile of thin sheets

The McGraw·Hill Companies

placed one above the other just like a deck of playing cards. When a force is applied on the upper liquid, each layer of the liquid looks like a card which drags with the next lower layer. As it happens, if one card moves, then its motion is opposed by the other cards. Similarly, in case of a liquid, when a liquid is flowing, the motion of one layer is opposed by another layer. Therefore, the liquid experiences resistive force or internal friction known as *viscous force* and this property of the liquid is called *viscosity*.

2.2 COEFFICIENT OF VISCOSITY

Newton derived an expression for the coefficient of viscosity. To derive the expression, consider a liquid consisting of different layers as shown in Fig. 2.2. Consider a force F is applied on the upper layer and hence, there is a relative motion between adjacent layers. Let v and $v + dv$ be the velocities of the layers EF and GH respectively. Let dx be the distance between them. The velocity gradient is $\frac{dv}{dx}$. The viscous force is directed opposite to that of the velocity gradient and directly proportional to the surface area of the layer.

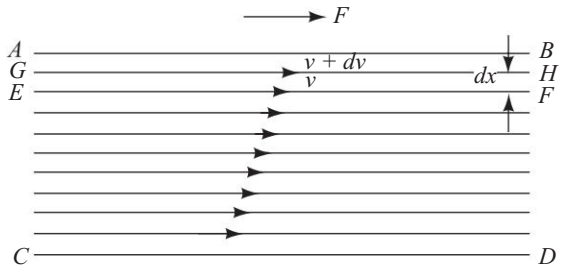


Fig. 2.2 *Flow of liquid—when there is a force F is applied*

i.e., $$F \propto -\frac{dv}{dx}$$
$$\propto A$$

The viscous force can be written by introducing a proportionality constant, as

$$F = -\eta A \frac{dv}{dx} \tag{2.1}$$

where η is known as the *coefficient of viscosity.* It is expressed in poise in the CGS system. In SI unit, it is expressed in N s m^{-2}.

If $\frac{dv}{dx} = 1\ s^{-1}$ and $A = 1\ m^2$, then $F = \eta$. The coefficient of viscosity η is defined as the *tangential force per unit area required to maintain unit relative velocity between its two layers separated by unit distance.*

2.3 STREAMLINE AND TURBULENT FLOW

In a steady (or laminar) flow of a fluid, the velocity of the fluid does not change with time, either in magnitude or in direction. This type of flow is also said to be streamline flow. The gentle flow of water is one example for streamline flow. If the velocity of

The McGraw·Hill Companies

the streamline flow increases, the flow becomes non-steady. The velocity at which the steady, laminar or streamline flow changes into turbulent or eddy flow, is known as critical velocity. If the velocity of the flow further increases and it is greater than the critical velocity, then streamline flow gets changed into non-steady (or turbulent or eddy) flow and it becomes highly irregular.

2.4 REYNOLD'S NUMBER

Reynold, with a number of experiments, found that the critical velocity is directly proportional to the coefficient of viscosity η and inversely proportional to the density of the liquid ρ and radius of the tube, r. According to Reynold, the critical velocity is given by

$$v_c = \frac{k\eta}{\rho r} \tag{2.2}$$

where k is a constant, known as Reynold's number. On the basis of the results of a number of experiments, he found that if k is equal to 2000, the flow of the liquid is said to be streamline flow. Generally, if k is less than 2000, the flow is streamline flow, whereas if k is greater than 3000, the flow is said to be turbulent. The flow is unstable if k lies between 2000 and 3000 and the flow may change from one type to another.

2.5 POISEUILLE'S EQUATION FOR THE FLOW OF A LIQUID THROUGH A TUBE

Consider a capillary tube of length 'l' as shown in Fig. 2.3 and consider there is a steady flow of liquid through it. Let x be the radius of the capillary tube. Let the velocity of the liquid flowing through the capillary tube at all points of the cylindrical layer be constant. The velocity of the flow increases from the surface (wall) of the capillary tube to the axis. Due to the fast-moving liquid at the centre portion compared to the wall of the capillary tube, it experiences a backward tangential force. Therefore, the retarding force acting according to Newton's law of viscous force is

$$F = -\eta A \frac{dv}{dx}$$

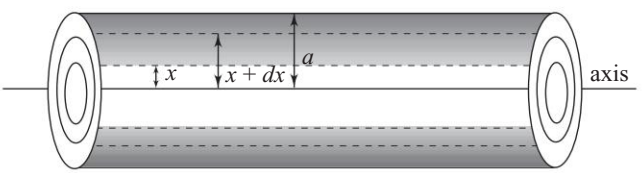


Fig. 2.3 *Poiseuille's equation—Derivation*

Substituting $A = 2\pi x.l$, where A is the surface area, x the radius and l the length of the tube, we get,

$$F = -2\pi x l \eta \frac{dv}{dx}$$

The McGraw·Hill Companies

The flow of liquid in the capillary tube is maintained by a forward force and hence there is a pressure difference between the two ends of the tube. Let P be the pressure difference. The force exerted due to this pressure is $P.\pi x^2$. These two forces are acting opposite to each other, so equating these two forces, we get,

$$-2\pi x l\eta \frac{dv}{dx} = P\pi x^2$$

$$\therefore \quad dv = \frac{-P\pi x^2 dx}{2\pi x l\eta}$$

i.e., $$dv = \frac{-P}{2l\eta} x dx$$

Integrating the above equation, we get,

$$V = \frac{-P}{2l\eta}\frac{x^2}{2} + C \tag{2.3}$$

where C is an integration constant. This constant is evaluated by substituting the boundary conditions, when $x = a$, $v = 0$, i.e., at the wall of the capillary tube, the velocity of the layer is zero.

Substituting $v = 0$ when $x = a$, we get,

$$C = \frac{P}{4l\eta} a^2 \tag{2.4}$$

From Eq. (2.4) and Eq. (2.3), we get,

$$v = \frac{P}{4l\eta}\left(a^2 - x^2\right) \tag{2.5}$$

Equation (2.5) gives the velocity of the flow at a distance of x from the axis of the tube. From Eq. (2.5), it is found that the velocity at $x = a$ is zero and at $x = 0$, the velocity is maximum. Hence, this equation indicates that the profile of the advancing liquid is a parabola and is shown in Fig. 2.4. To find the volume of the liquid flowing through the capillary tube, consider another close co-axial cylinder at $x + dx$ from the axis of the tube. The area of this co-axial cylinder is $2\pi x.dx$. The volume of the liquid flowing through the capillary tube in one second is

dv = cross sectional area × velocity = $2\pi x \cdot dx\, v$

Substituting the value of v from Eq. (2.5), we get

$$dv = 2\pi x\, dx.\frac{P}{4l\eta}\left(a^2 - x^2\right) \tag{2.6}$$

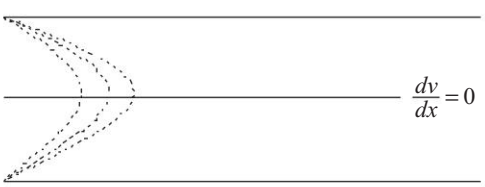


Fig. 2.4 *Profile of a liquid flowing through a capillary tube*

The McGraw·Hill Companies

The volume of the liquid flowing through the entire capillary tube is obtained by integrating the above equation between $x = 0$ and $x = a$.

$$V = \int_0^a \frac{2\pi P}{4l\eta}\left(a^2 - x^2\right) x\, dx$$

$$V = \frac{\pi P}{2l\eta}\int_0^a \left(a^2 x - x^3\right) dx$$

$$V = \frac{\pi P}{2l\eta}\left(\frac{a^4}{2} - \frac{a^4}{4}\right)$$

$$V = \frac{\pi P a^4}{8l\eta} \tag{2.7}$$

Equation (2.7) is known as Poiseuille's formula for the rate of flow of liquid through a capillary tube and is valid for streamline flow.

2.6 MOTION OF A RIGID BODY IN A VISCOUS MEDIUM

Consider a perfectly rigid and smooth body falling through a viscous medium. Its motion is opposed by a force, usually frictional force in nature. When a rigid body is moving through a viscous liquid, it carries a layer of the fluid in contact, along with it. This produces some relative motion between the layers of the fluid. This relative motion is opposed by viscous force of the fluid. This opposing force increases with the velocity of the body. In case of small bodies, this opposing force is just equal to the motive or driving force and the body then attains a constant velocity known as *terminal velocity.*

Stoke showed that this retarding force for a spherical body of radius r moving with a velocity v in a viscous liquid of viscosity η is

$$F = 6\,\pi\, v\, r\, \eta \tag{2.8}$$

Equation (2.8) is known as Stoke's law. This equation can be obtained by considering the dimensions of various quantities on two sides.

For a slow moving body, $F \propto$ terminal velocity of the body

$\propto$ radius of the body

$\propto$ viscosity of the liquid

or $$F = K\, v^a\, r^b\, h^c \tag{2.9}$$

where K is a constant and a, b and c are dimensional coefficients. Substituting the dimensions of various quantities on two sides, we get,

$$\left[M^1L^1T^{-2}\right] = \left[LT^{-1}\right]^a [L]^b \left[ML^{-1}T^{-1}\right]^c$$

$$M^1L^1T^{-2} = \left[L^{a+b-c}\right]\left[T^{-a-c}\right]\left[M^c\right]$$

Equating the coefficients of M, L and T, we have $a = b = c = 1$.

$$F = K\, v\, r\, \eta$$

The value of K found by Stoke was $6\,\pi$. Let r be the density of the spherical body. Then its weight = mass × acceleration due to gravity is,

$$\text{Weight} = \text{Density} \times \text{Volume} \times g = \frac{4}{3}\pi r^3 \rho g$$

If σ is the density of the medium, then upthrust on the body due to the displaced medium $= \frac{4}{3}\pi r^3 \sigma g$

Therefore, the resultant downward force acting on the body

$$= \frac{4}{3}\pi r^3 \rho g - \frac{4}{3}\pi r^3 \sigma g = \frac{4}{3}\pi r^3 g(\rho - \sigma)$$

Equating this against the value of the retarding force from Stoke's law, we get,

$$6\pi v r \eta = \frac{4}{3}\pi r^3 (\rho - \sigma) g$$

$$v = \frac{2}{9}\frac{r^2 g(\rho - \sigma)}{\eta} \tag{2.10}$$

The terminal velocity of body falling in a viscous liquid is (i) inversely proportional to the viscosity of the liquid, η, (ii) directly proportional to the square of the radius of the body, r^2, and (iii) directly proportional to the difference between the density of the body and the medium.

Consider an air bubble is produced at the bottom of a water tank. The density of water is greater than the density of the air bubble. Therefore, $\sigma > \rho$. The terminal velocity is negative. Hence, the body will have an upward terminal velocity. Therefore, an air bubble produced at the bottom of the tank moves in the upward direction.

A raindrop formed in the clouds is an example of a body falling through a viscous medium. The diameter of the smaller raindrops is around 0.001 cm and their velocity, when they are falling through air, for which $\eta = 1.8 \times 10^{-5}$ N s m^{-2}, is about 1.2×10^{-2} m s^{-1}. Thus, they appear as floating in air. The radius of a bigger raindrop is about 0.01 cm and its terminal velocity when falling through air is 1.2 m s^{-1}. Therefore, bigger raindrops fall through the air.

The McGraw·Hill Companies

2.7 EXPERIMENTAL DETERMINATION OF VISCOSITY OF A LIQUID

2.7.1 Poiseuille's Method

This experimental set-up is shown in Fig. 2.5 and it consists of a constant-pressure vessel V and a capillary tube AB with a 1 cm length. The constant pressure vessel is filled with the given liquid up to the level O. The level of the liquid in the vessel V is maintained as constant. If the level of the liquid exceeds the limit, then it flows through the outlet O. A capillary tube AB with 1 cm length is mounted in the bottom portion of constant pressure vessel and held horizontally. A beaker is placed just below the end B of the capillary tube. The liquid flowing through the capillary tube is collected in the beaker for a period of t seconds and then the mass of the liquid collected is found out. The height h between the upper level of the liquid and the capillary tube is measured. The pressure difference P is $h\rho g$, where ρ is the density of the liquid. Then, from Poiseuille's equation, the viscosity of the liquid flowing through the capillary tube in one second is $\eta = (Pa^4\pi)/(8Vl)$.

If the liquid is collected for t second, then, $$\eta = \frac{\pi \rho a^4 t}{8Vl} = \frac{\pi h \rho g a^4 t}{8Vl} \qquad (2.11)$$

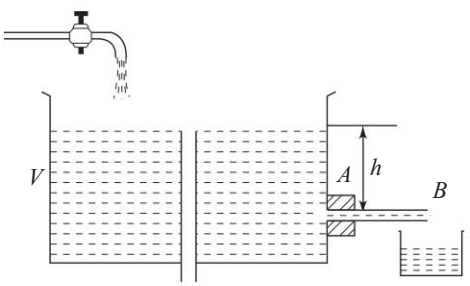


Fig. 2.5 *Poiseuille's method—Determination of η using a pressure vessel*

where V represents the volume of the liquid flowing in one second, a the radius of the capillary tube and l the length of the capillary tube. The radius of the capillary tube is accurately measured using a travelling microscope. The density ρ of the liquid is taken from the table or it is measured using Hare's apparatus.

If a constant-pressure vessel is not available, the above experiment can be done with the help of a burette. At the bottom of the burette, a capillary tube is attached as shown in Fig. 2.6. The given liquid is filled in the burette. The given liquid is allowed to flow from the position h_1 to h_2 and the liquid is collected in a beaker. The time taken for the flow of liquid from h_1 to h_2 is noted. By measuring V, l, h and a, the viscosity of the liquid is determined using the formula $\eta = \frac{\pi h \rho g a^4 t}{8Vl}$.

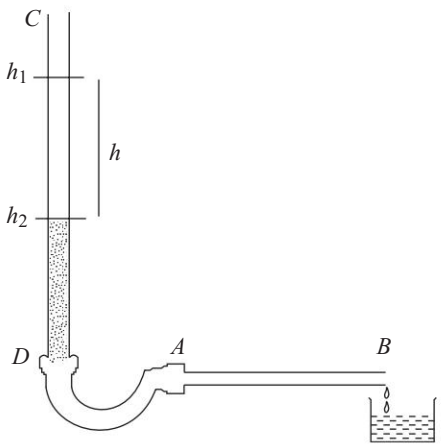


Fig. 2.6 *Poiseuille's method—Determination of η using burette*

2.7.2 Stoke's Law

The experimental set-up used to find viscosity of a highly viscous medium using Stoke's method is shown in Fig. 2.7. It consists of a long glass tube or jar of nearly 1-m length and it is filled with the given highly viscous liquid like castor oil, whose viscosity is to be determined. The temperature of the liquid is measured using a thermometer. A small spherical ball of nearly 2 to 5 mm diameter is dropped into the liquid. After travelling a distance of nearly 10 to 15 cm inside the liquid, the spherical ball attains the terminal velocity. When it attains the terminal velocity, the time taken by the ball to travel from *A* to *B* or a distance of 10 cm is noted. From this observation, the terminal velocity is determined using the formula given below.

Terminal velocity, $v = \frac{AB}{t}$

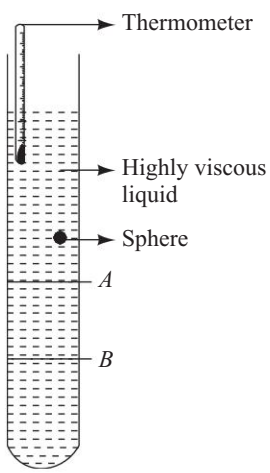


Fig. 2.7 *Stoke's method–Determination of η*

The McGraw·Hill Companies

The experiment is repeated by dropping the spherical ball, three to five times. Then the average value of the terminal velocity is noted. By measuring the radius of the sphere using a travelling microscope accurately, the viscosity of the given liquid is determined using the relation

$$\eta = \frac{2}{9}\frac{r^2(\rho - \sigma)g}{v} \tag{2.12}$$

where ρ and σ are the densities of the sphere and the liquid respectively.

Ladenburg has shown that two different corrections known as *wall effect* and *end effect* have to be applied, because of the finite size of the glass tube or jar. According to Ladenburg, the velocity v_∞ of the ball due to *wall effect* would be

$$v_\infty = v\left(1 + \frac{2.4r}{R}\right) \tag{2.13}$$

and the velocity v_∞ of the ball due to *end effect* would be

$$v_\infty = v\left(1 + \frac{3.3r}{l}\right) \tag{2.14}$$

where v is the observed velocity of the sphere, R the radius of the glass tube or jar and l the full depth of the liquid column in the glass tube or jar.

Incorporating these two corrections, the viscosity of the liquid is given by

$$\eta = \frac{2}{9}\frac{r^2 g(\rho - \sigma)}{v\left(1 + \frac{2.4r}{R}\right)\left(1 + 3.3\frac{r}{l}\right)} \tag{2.15}$$

Using this method, the radius of a small drop falling through air is determined if the viscosity of the air medium is known.

2.7.3 Ostwald's Viscometer—Comparison of Viscosities

This experimental set-up is used to compare the viscosity of any two given liquids. The schematic representation of this viscometer is shown in Fig. 2.8. It consists of two glass bulbs *A* and *B* and suitable markings are made on the glass bulbs such as *C*, *D* on the bulb *A* and *E*, *F* on the bulb *B*. These markings represent the volume of the bulb *A* between *C* and *D*, and the volume of the bulb *B* between *E* and *F,* which are same. The ends *D* and *F* of the glass bulbs, *A* and *B*, are connected to a capillary tube *DGF*. The end *E* of the bulb *B* is connected to a rubber tube and a stopcock *S* is used to close the rubber tube. The end *C* of the bulb is connected to a funnel. The entire set up is immersed in a liquid bath at a temperature t_1K. The given liquid, whose viscosity is to be determined, is poured into the funnel up to the marking *E* in the bulb *B*. Then the stopcock is opened and the rubber tube is connected to a vacuum pump and the air present in the tube is evacuated. Therefore the level of the liquid rises up above the mark *E* and then, the stopcock is closed. On opening the stopcock, the level

of the liquid decreases and the time taken to decrease the level from E to F is noted. Then the viscometer is filled with another given liquid and the entire experiment is repeated. Let t_1 and t_2 be the time taken by the liquid to decrease from the level E to F. Then, from Poiseuille's formula, for the first liquid, the volume of liquid flowing from the level E to F is

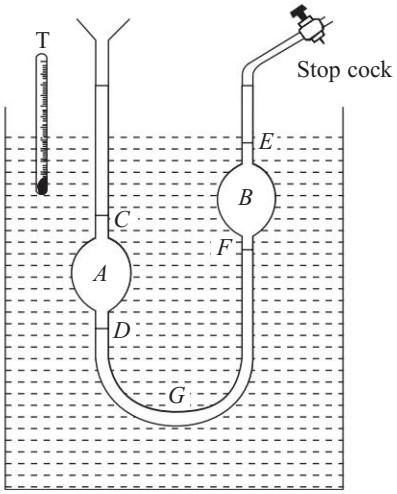


Fig. 2.8 *Ostwald's viscometer*

$$V_1 = \frac{\pi P_1 a^4}{8\eta_1 l} \tag{2.16}$$

Similarly, the volume of the second liquid, flowing from E to F is

$$V_2 = \frac{\pi P_2 a^4}{8\eta_2 l} \tag{2.17}$$

The pressure $P = h\rho g$. Therefore, $P_1 \alpha \rho_1$ and $P_2 \alpha \rho_2$
From Eq. (2.16) and Eq. (2.17), we get,

$$\frac{V_1}{V_2} = \frac{P_1}{P_2}\frac{\eta_2}{\eta_1} \tag{2.18}$$

Let Q be the quantity of the first liquid flowing from E to F in t_1 second and the same quantity of the other liquid is also flowing in t_2 second. Therefore,

$$V_1 = \frac{Q}{t_1} \quad \text{and} \quad V_2 = \frac{Q}{t_2} \tag{2.19}$$

From Eq. (2.18) and Eq. (2.19), we get,

$$\frac{V_1}{V_2} = \frac{\rho_1}{\rho_2}\frac{\eta_2}{\eta_1} \Rightarrow \frac{t_2}{t_1} = \frac{\rho_1}{\rho_2}\frac{\eta_2}{\eta_1}$$

The McGraw·Hill Companies

i.e., $$\frac{\eta_1}{\eta_2} = \frac{\rho_1 t_1}{\rho_2 t_2} \quad (2.20)$$

By knowing the densities of the liquids, the ratio of their viscosities is determined. If the density of the liquids is not known, Hare's apparatus is used to find the density of the liquids. If the viscosity of any one liquid is known, then the viscosity of the other liquid is determined, using Eq. (2.20).

Keypoints to Remember

- The internal resisting force acting inside the liquid makes it flow. This property is known as viscosity.
- The viscous force is the resistive force or internal friction experienced by the liquid.
- The coefficient of viscosity is defined as the tangential force per unit area required to maintain unit relative velocity between its two layers separated by unit distance.
- During the laminar flow of a fluid, the velocity of the fluid does not change with time either in magnitude or direction.
- According to Reynold, the critical velocity is equal to $k\eta/\rho r$, where k is known as Reynold's number ρ, the density of the liquid, and r, the radius of tube.
- Poiseuille's formula for the rate of flow of a liquid through a capillary tube is equal to $\pi Pa^4/8l\eta$ where P is the pressure difference, a is the radius of the capillary tube, l is the length of the tube and η is the viscosity
- Poiseuille's formula for the viscosity of the given liquid is $\pi h\rho g a^4 t/8Vl$, where h is the height of the liquid, v is the volume of the liquid flowing in one second and t is the time of flow of water.

Solved Problems

Example 2.1 Calculate the mass of water flowing in 5 minutes through a tube with 0.054 cm radius and 56 cm length, if there is a constant pressure head of 34 cm of water. The coefficient of viscosity of water is 8.9×10^{-4} N s m^{-2}.

Given Data

Coefficient of viscosity of water	$\eta = 8.9 \times 10^{-4}$ N s m^{-2}
Radius of the capillary tube	$a = 0.054 \times 10^{-2}$ m
Length of the capillary tube	$l = 56 \times 10^{-2}$ m
Height of the pressure head	$h = 34 \times 10^{-2}$ m
Time of flow	$t = 5$ minutes $= 300$ s

Solution From Poiseuille's formula, the volume of the water flowing in one second,

$$V = \frac{\pi P a^4}{8\eta l}$$

where P is the pressure of water. It is given by $P = h\rho g$, where h is the height of the pressure vessel and ρ the density of water (ρ for water = 10^3 kg/m^3). Substituting the values, we get the volume of water flowing in one second,

$$V = \frac{\pi h g P a^4}{8\eta l}$$

$$V = \frac{\pi \times 3.4 \times 10^{-3} \times 10^3 \times 9.8 \times (0.054 \times 10^{-2})^4}{8 \times 8.9 \times 10^{-4} \times 56 \times 10^{-2}}$$

$$= 2.2323 \times 10^{-9} \text{ m}^3$$

Volume of water flowing in 5 minutes

$$V = 2.2323 \times 10^{-9} \times 300 \text{ m}^3$$

$$= 6.6969 \times 10^{-7} \text{ m}^3$$

Mass of water flowing in 5 minutes

$$\text{Mass} = \text{density} \times \text{volume}$$

$$= 6.6969 \times 10^{-7} \times 10^3$$

$$= 6.6969 \times 10^{-4} \text{ kg}$$

The mass of water flowing in 5 minutes is 0.66969 g.

Example 2.2 In an experiment using Ostwald's viscometer, the viscosity of water and alcohol are compared. The data obtained in this experiment is (a) the time taken by water to flow from one marking A to another marking $B = 6$ seconds and (b) the time taken by alcohol to flow from A to $B = 3.6$ seconds. If ρ_1 and ρ_2 for water and alcohol are 10^3 kg m^{-3} and 800 kg m^{-3} respectively, calculate the ratio of their viscosities? If the viscosity of water is 8.9×10^{-4} N s m^{-2}, calculate the viscosity of alcohol?

Given Data

Time taken by water	= 6 s
Time taken by alcohol	= 3.6 s
Density of water, ρ_1	= 10^3 kg m^{-3}
Density of alcohol, ρ_2	= 800 kg m^{-3}
Viscosity of water, η_1	= 8.9×10^{-4} N s m^{-2}

Solution The ratio of viscosities of any two liquids, $\frac{\eta_2}{\eta_1} = \frac{t_2}{t_1}\frac{\rho_2}{\rho_1}$

Substituting the values of ρ_1, ρ_2, t_1 and t_2, we get $\frac{\eta_2}{\eta_1} = \frac{3.6}{6} \times \frac{800}{1000} = 0.48$

Therefore, $\eta_2 = 0.48 \times \eta_2$
$= 0.48 \times 8.9 \times 10^{-4}$
$= 4.272 \times 10^{-4}$

The viscosity of the alcohol = 4.272×10^{-4} N s m^{-2}.

Example 2.3 A spherical body of radius 0.2 cm is falling through a medium of density 1.26×10^3 kg m^{-3} and viscosity 2×10^{-5} N s m^{-2}. Find the terminal velocity of the body. The density of the spherical body = 8000 kg m^{-3}.

Given Data

Density of the medium, σ $= 1.26 \times 10^3$ kg m^{-3}
Viscosity of the medium, η $= 2 \times 10^{-5}$ N s m^{-2}
Radius of the body, r $= 0.2$ cm $= 0.2 \times 10^{-2}$ m
Density of the body, ρ $= 8 \times 10^{-3}$ kg m^{-3}

Solution The terminal velocity $v = \dfrac{2r^2(\rho - \sigma)g}{9\eta}$

$$v = \frac{2 \times (0.2 \times 10^{-2})^2 \times (8 \times 10^{-3} - 1.26 \times 10^{-3}) \times 9.8}{9 \times 2 \times 10^{-5}}$$

$= 2.874 \times 10^{-3}$

The terminal velocity of the spherical body is 2.87×10^{-3} ms^{-1}.

Objective Questions

2.1 The viscosity of a liquid is due to the _____ force experienced by the liquid.

2.2 The unit for viscosity is _____.

2.3 According to Reynold, the critical velocity is equal to.

a) $v_c = \dfrac{k\eta^2}{\rho r}$ b) $v_c = \dfrac{k\eta}{\rho r^2}$

c) $v_c = \dfrac{k\eta}{\rho r}$ d) $v_c = \dfrac{k\eta^2}{\rho^2 r}$

2.4 According to Poiseuille, the volume of a rate of flow of a liquid through a capillary tube is equal to

a) $V = \dfrac{\pi Pa^4}{4l\eta}$ b) $V = \dfrac{\pi Pa}{8l\eta}$

c) $V = \dfrac{\pi Pa^4}{6l\eta}$ d) $V = \dfrac{\pi Pa^4}{8l\eta}$

The McGraw·Hill Companies

2.5 Which is not correct? The terminal velocity of a body falling in a viscous liquid is

a) inversely proportional to the viscosity of liquid

b) directly proportional to the square of the radius of the body

c) directly proportional to the difference between the density of the body and the medium.

d) none of the above

2.6 Viscosity of liquid using the Poiseuille's method is equal to

a) $\eta = \frac{\pi h \rho g a^4 t}{8vl}$ b) $\eta = \frac{\pi h \rho g a^2 t}{8vl}$

c) $\eta = \frac{\pi h \rho g a}{8vl}$ d) $\eta = \frac{\pi h \rho g a^4}{8vl}$

2.7 The velocity of liquid using Ostwald's viscosity is equal to

a) $\frac{\eta_1}{\eta_2} = \frac{\rho_1 t_2}{\rho_2 t_1}$ b) $\frac{\eta_1}{\eta_2} = \frac{\rho_2 t_2}{\rho_1 t_1}$

c) $\frac{\eta_1}{\eta_2} = \frac{\rho_1 t_1}{\rho_2 t_2}$ d) $\frac{\eta_1}{\eta_2} = \frac{\rho_2 t_1}{\rho_1 t_2}$

KEY

2.1 resistive force 2.2 N s m^{-2}

2.3 c) 2.4 d)

2.5 d) 2.6 a)

2.7 c)

Exercises

2.1 A drop of water of 0.12 mm radius is falling through a medium of density 1.2 kg m^{-3}. If the terminal velocity of the water is 1.2×10^{-2} m s^{-1}, find the viscosity of the medium.

2.2 A pressure head with 20 cm height is used to find the viscosity of a liquid. If a capillary tube with 0.12 cm diameter and 26 cm length is used, find the viscosity of liquid. Given that the volume of liquid collected in 60 seconds is 8.1×10^{-7} m^3. The density of the liquid = 900 kg m^{-3}.

2.3 A gas bubble with 1 cm diameter rises steadily through a liquid of density 1.0×10^3 kg m^{-3} at the rate of 0.35 cm s^{-1}. Calculate the coefficient of viscosity of the liquid. Density of the gas bubble =1.29 kg m^{-3}.

The McGraw·Hill Companies

Short Questions

2.1 What do you mean by viscosity of liquids?
2.2 Distinguish between streamline and turbulent flow.
2.3 Define the term 'coefficient of viscosity'.
2.4 What is Reynold's number?
2.5 What is a viscometer?
2.6 What are the advantages of an Ostwald's viscometer?
2.7 State Stoke's law.
2.8 Obtain Stoke's law using dimensional coefficients.
2.9 Derive an expression for coefficient of viscosity.
2.10 What is meant by terminal velocity?

Descriptive Questions

2.1 Derive Poiseuille's expression for the volume of the liquid collected, when it is flowing through a capillary tube of length '*l*' m in '*t*' seconds.
2.2 Explain any one experimental method used to find the viscosity of a liquid.
2.3 Explain with a neat sketch, the construction and working of an Ostwald's viscometer.
2.4 Explain how the viscosity of a given liquid is determined using Stoke's method.
2.5 Explain how the viscosity of a given liquid is determined using Poiseuille's method.

The McGraw·Hill Companies

Chapter

Vacuum Technique

3

OBJECTIVES

- To explore the knowledge on vacuum techniques and the ways to attain the low pressure in a vessel
- To understand the mechanism involved in different types of vacuum pumps
- To study instruments used for the measurement of vacuum at low pressure in a system
- The explore importance of measuring the vacuum / low pressure in a system
- To discuss the role of vacuum in industries and their applications

3.1 INTRODUCTION

Vacuum technique is one of the important techniques for scientific investigations and industrial developments. The term vacuum technique is used to represent the pressure below the normal atmospheric pressure of 760 torr, where 1 torr = 1 mm of mercury column. The pressure is also represented by other units like pascal, bar, etc. The equivalent of 1 bar is 760 torr and 1 torr is equivalent to 133 pascal. The vacuum attained is represented by the total pressure of the residual gases in the vessel, which is pumped.

3.2 TYPES OF PUMPS

The exhaust or vacuum pumps are classified into different categories based on the vacuum required in a system. The classification of the vacuum pumps and their method of operation along with their pressure level to be obtained in a system are shown in Table 3.1.

The McGraw·Hill Companies

Table 3.1 *Different vacuum pumps along with their vacuum levels*

S. No.	*Category*	*Type*	*Pressure level*
1.	Oil pumps	Piston type	Up to 10^{-2} mm
		Rotary type	Up to 10^{-4} mm
2.	Mercury pumps	Piston type	Up to 10^{-2} mm
		Rotary type	Down to 10^{-3} mm
3.	Molecular pumps	Rotary type	Up to 10^{-5} mm
4.	Diffusion pumps	-	Up to 10^{-6} mm

3.3 VACUUM SYSTEMS

Generally, the vacuum tubes are classified in two broad categories, namely, static and kinetic systems. *In static systems*, all the systems are tight, i.e., sealed hermetically except the exit from the pump. The static systems are mainly used to attain very low pressure. In kinetic systems, very fast pumps are used for the continuous evolution of the gas or air from various parts of the apparatus.

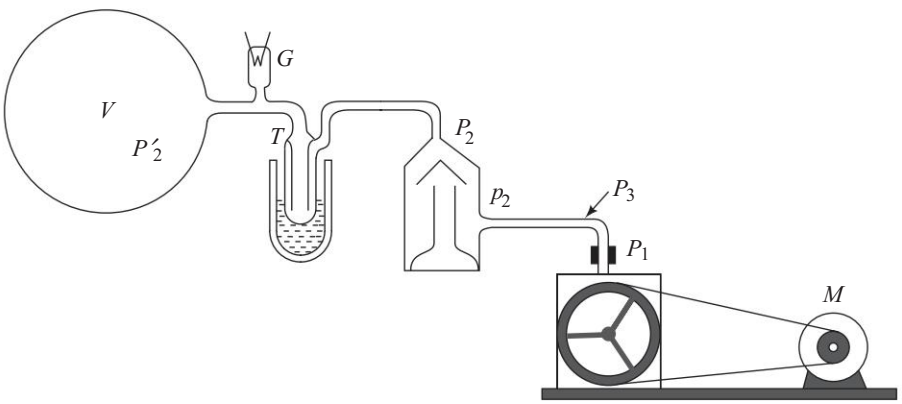


Fig. 3.1 *Block diagram of vacuum systems*

A schematic diagram of the vacuum system is shown in Fig. 3.1. The system is operating based on the kinetic method. V is the volume of the system in which high vacuum is to be obtained. The pump P_1 is known as *backing pump* and it operates against the atmoshperic pressure. Due to its continuous operation, it reduces the pressure sufficiently to a low value. This helps in the operation of the pump P_2. The pump P_1 reduces the pressure initially at the P_1 area and thereby reduces at P_2. Thus, due to combined action of P_1 and P_2, the required high vacuum is attained in V. A mercury–vapour trap T can be used between V and the pump P_1, to avoid the entry of mercury vapour to reach V during vacuum process. A resistance to the flow of gas can be achieved by connecting a tube between V and the pump P_1. This can be reduced by using a connecting tube with larger width and shorter length. The actual pressure attained in V can be measured by connecting a low–pressure gauge G, as shown in Fig. 3.1. The above system is used in most of the case wherein a constant vacuum is required. Thus, by operating the pump continuously, one can obtain the constant vacuum in any system.

The McGraw·Hill Companies

3.4 PRODUCTION OF VACUUM

Let us discuss briefly the working principle, construction, operation and the pressure limitations of vacuum pumps, namely, rotary vacuum pump, diffusion pump and turbo molecular pump in the following section:

3.4.1 Rotary Pump (or) Gaede Potary Oil Pump

This pump is used to produce vacuum in the order of 10^{-3} torr. It is based on the principle of displacement of a gas. This pump consists of a rotating disc *R* called *rotor*. It is rotating eccentrically inside a hollow cylinder *S*, called *stator*. The stator has an inlet *I* connected to vessel, which is to be exhausted and it also has an outlet *O*, which is also an exit for the compressed gases. The portion inside the stator *S* is splitted into two separate compartments with the help of a valve, *V*.

The vessel to be evacuated is connected to a stopcock. The entire set up is immersed in oil as shown in Fig. 3.2. If the pump is stopped, the stopcock will turn to the left and it will allow the atmospheric air to enter into the pump.

Since the rotor moves in immersed oil, the oil fills the dead space and thus allows a high compression ratio to be achieved. It also seals a moving disc against back leakage, besides lubricating and cooling the pump.

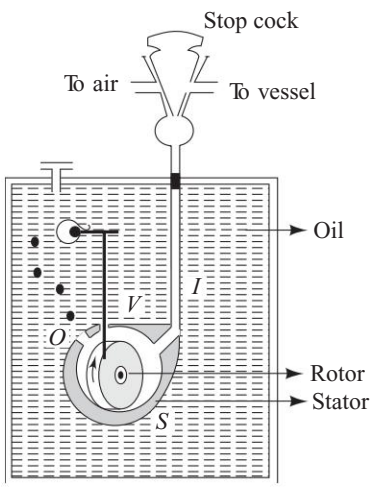


Fig. 3.2 *Rotary pump*

The rotor is operated externally by means of an electric motor. The rotor is capable of rotating in the clockwise direction. When the rotor is rotating, the space in the outlet portion decreases and the gases present in this portion get compressed. The compressed gases go out through the outlet *O*. When the rotor rotates again and again, the pressure inside the stator decreases and hence, the air inside the vessel is drawn. This air is also compressed and exhausted through the outlet *O*. Thus, during each revolution of the pump, the air is drawn from the vessel and is removed from the vessel continuously.

3.4.2 Stationary Vane or Cenco-Hyvac Rotary Oil Pump

A massive cylindrical shaft called rotor *R* revolves eccentrically inside a hollow stout steel cylinder called stator *S* compressing the gas or entering air, and finally ejects it through a 'non-return' outlet valve. The whole pumping set-up is kept immersed in oil for (i) providing automatic lubrication, (ii) preventing leakage of gas (or) air into the high vacuum created, and (iii) making for efficient cooling of the pump.

The stationary vane or the Cenco-Hyvac rotary oil pump is the most commonly used oil pump in vacuum industries. The schematic representation of the vacuum systems is shown in Fig. 3.3. The rotor with the help of a spring *S* through the partition vane *V* maintains the air or gas flow inside the cylinder. This helps to introduce the fresh incoming air or gas to the cylinder. Hence, the system is known as stationary rotary oil pump.

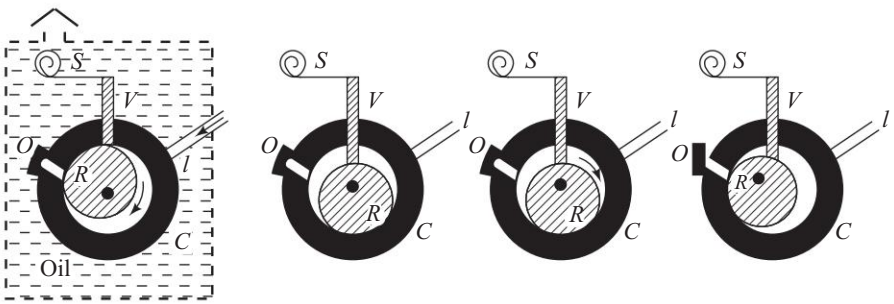


Fig. 3.3 *Cenco-Hyvac rotary oil pump*

The cylinder consists of an inlet tube *I* and an outlet tube *O*. The inlet tube *I* is connected to the vessel which is to be exhausted, while the outlet tube is provided with a valve opening outwards. In order to prevent leakage, the whole system is immersed in oil. A very high-speed motor is used to rotate the rotor *R* in the same direction as shown by the arrow head. The external system to be exhausted is connected to the inlet tube *I* as shown in Fig. 3.3a. The air or gas from the vessel is just admitted into the cylinder *C* and the vacuum process is ready to start. The compression of the air or gas takes place (Fig. 3.3b) immediately after starting the rotation of the rotor. The vane *V* helps in keeping apart from the fresh air or gas coming from the vessel. The process is continued and is shown in Fig. 3.3c. During the final stage of compression (Fig. 3.3d), the value at the outlet *O* is forced to open and hence the air or gas is expelled out. This process is continued until a high vacuum is obtained in the vessel. One can use this vacuum pump to exhaust vessels of larger volumes upto a pressure level of 10^{-3} mm.

The following are the advantages of a rotary oil pump:

- It is an automated and electrical motor-driven method and hence the operation is quick and faster.
- One can achieve a high vacuum of 10^{-3} mm at a faster rate.
- It facilitates to prevent leakage of gas or vapour with the exhaust vessel to obtain an efficient cooling and automatic lubrication due to immersion of the system in oil.

The McGraw·Hill Companies

Pumping Speed It is defined as the rate of decrease of mass of the gas or air in a given volume (i.e., *dm/dt*) or as the rate of change of volume of the gas or air in the vessel at any given instant, the volume being measured at the pressure obtaining at that very instant as *dv/dt*. The speed of the pump may be defined as the refractive rate of reduction of pressure in a given volume. The limiting value of the attainable pressure (P_o) is never zero, i.e., the prefect vacuum is not attainable by means of any exhaust pump .

3.4.3 Diffusion Pump

A diffusion pump works based on the principle of diffusion. A rotary pump continuously backs this pump. The vessel to be evacuated is connected to the diffusion pump and then, the rotary pump is switched on so as to produce a rough vacuum in the order of 10^{-3} torr. Then, the diffusion pump is switched on.

The schematic representation of a diffusion pump is shown in Fig. 3.4. The diffusion pump consists of a conical shaped vessel *C* in which a low vapour pressure oil like silicon oil or mercury is heated by a heater *H*. When mercury gets heated, it evaporates and the mercury vapour goes up. The asbestos lining *A* is used to prevent the cooling of mercury vapours on their way up. The mercury vapour flows out through the nozzle *B* and moves towards *D*. According to Bernoulli's principle, the pressure of air in *B* is very less compared to the vessel to be exhausted. Hence, the air near *B* is carried away by the mercury vapours to *D*. When mercury vapour is moving towards *D* from the nozzle *B*, it is cooled down by the circulating water in the outer jacket. Therefore, the mercury vapour condenses into mercury and is then reused. The rotary pump removes the air carried away by the mercury vapour. Using this diffusion pump, one can produce vacuum in the order of 10^{-8} torr in a clean vessel.

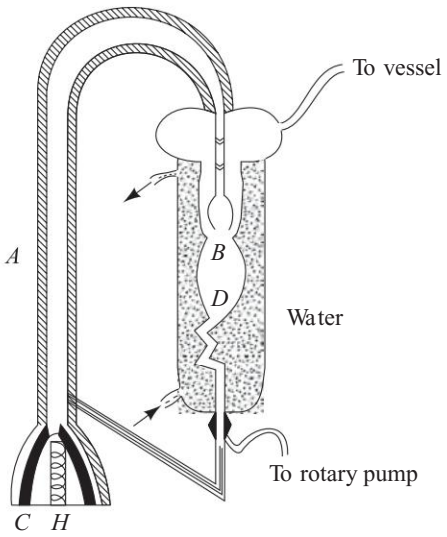


Fig. 3.4 *Diffusion pump*

The McGraw·Hill Companies

3.4.4 Molecular Pump

When the space between the rapidly rotating surface, i.e., rotor and the adjacent stationary stator is as small as 0.03 mm, it exerts a dragging force on the molecule of the gas or air in that space due to viscosity, in the direction of its own motion.

When the pressure is low, the mean free path (λ) of the molecules of the gas or air is greater than the linear dimension of the air gap between the rotor and stator. Therefore, the collision between molecules and molecules are very less when compared with the collision between molecules and walls of the gap. The molecules acquires the drift velocity and are not altered by any subsequent collisions among themselves. The molecular or diffusion pumps require a reduced pressure or fore vacuum of the order 10^{-1} to 10^{-2} mm in the exhaust system. Therefore, a backing pump will be connected to the exhaust port which creates the required initial reduced pressure.

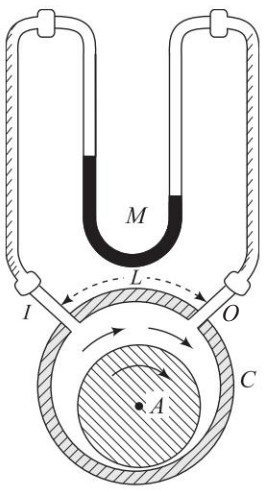


Fig. 3.5 *Molecular pump*

The schematic diagram of a molecular vacuum pump is shown in Fig. 3.5. It consists of a rotor *A* which revolves about an axis through its centre. Generally, the speed of the rotor is 5000 rpm. The rotor is sealed hermitically by a stator *C* and *O* are the respective inlet and outlet in the stator. The inlet and outlet are fixed in such a way that the annular gap between them is wider than all other points in the stator.

Employing the backing pump, the required reduced pressure is created in the exhaust system. The rotor is rotated using an electricity driven motor with very high speed. When the rotor revolves, it drags the air or vapour from the inlet *I* to the output port *O*. The same is indicated by the manometer *M*. The obtained pressure difference between the *I* and *O* is given by the relation

$$p_1 - p_2 = \eta 6 \ell u / h^2 \tag{3.1}$$

where η is the viscosity of the gas, u the speed of the rotor, ℓ and h are the length and radial depth of the slot between *I* and *O*, respectively.

At ordinary pressure, the pressure difference remains constant since there is no change in viscosity for a constant speed of the rotor. On the other hand at low pressure, the pressure ratio p_1/p_2 remains constant for a corresponding speed of the rotor. Further, the pressure is also independent of the reduced pressure or fore vacuum.

The McGraw·Hill Companies

Therefore, the pressure ratio is

$$p_1/p_2 = e^{cu} \tag{3.2}$$

where c is a constant which depends on both the nature of the gas or vapour and the dimension of the slot. Generally, a set of twelve slots will be provided at the circumference of the rotor. These slots will provide the necessary pressure variation between the centre and to the ends where the backing pump is connected.

Thus, it is clear that the pressure retained by this method depends on the speed of the rotor and the fore vacuum. A vacuum of 10^{-6} mm was obtained within a minute using a rotor speed of 10,000 rpm and fore vacuum of 2 mm. The vacuum is further exhausted to 10^{-7} mm by evacuating nitrogen gas with a rough vacuum of 15 mm and a rotor speed of 4500 rpm.

The limitations of the molecular pump are given below:

(i) There can be occurrence of mechanical trouble due to the small clearance between the rotor and stator.
(ii) It is difficult to use slow vaporizing substances like mercury and grease.
(iii) A backing pump is required to obtain the fore vacuum.

3.5 MEASUREMENTS OF VACUUM

We know that suitable measuring instruments are required to measure the vacuum or low pressure in a system. The low pressure measuring devices are classified into two categories, namely, manometers and gauges. The instruments which are in the manometer category are mercury, mechanical and viscosity manometers. The vacuum gauges are different types, namely, radiometer, conductivity, ionization and effusion gauges. Let us discuss only about Pirani (thermal vacuum detector) gauge and Penning (ionization vacuum detector) gauge.

3.5.1 Pirani Resistance Gauge

This gauge is based on the principle that the thermal conductivity K of a gas at a pressure below 10^{-2} mm of Hg is directly proportional to the pressure P, i.e., K is a linear function of P, or $K = \alpha P$, where α is a constant.

The schematic representation of a Pirani gauge is shown in Fig. 3.6. The pirani gauge (PG) consists of tungsten or platinum filament F, enclosed in a glass bulb B. The bulb is open at the lower end that is connected to the vessel in which the pressure is to be measured.

The Pirani gauge is connected to a Wheatstone bridge consisting of two fixed resistances R_1 and R_2. The resistance R_3 is a variable one. These three resistances being made of an alloy like *manganin* and *minalpha*, having an almost zero coefficient of temperature. The ends A and C of the bridge are connected to a battery and a rheostat. The potential difference between the ends A and C is measured with a voltmeter.

To measure the pressure of the vessel, the glass bulb of the gauge is kept in the path of the exhausted gas from the vessel. Since the vessel is at low pressure, the thermal conductivity of the gas changes, i.e., it varies linearly with the pressure. Due to the variation in the thermal conductivity of the gas, the temperature of the gas also changes.

If the Pirani gauge is kept at the vessel having low pressure, due to the change in the temperature of the gas, the temperature of the filament *F* (platinum wire) changes. Due to the change of the temperature of the filament, there is a change of its resistance. The change in resistance leads to an unabalance in the bridge circuit. The deflection shown in the galvanometer is a measure of the pressure of the vessel. The bridge is balanced by varying the potential across *A* and *C* with the help of the rheostat. The galvanometer scale divisions are used to measure the pressure directly by calibrating the galvanometer in terms of pressure using a McLeod gauge.

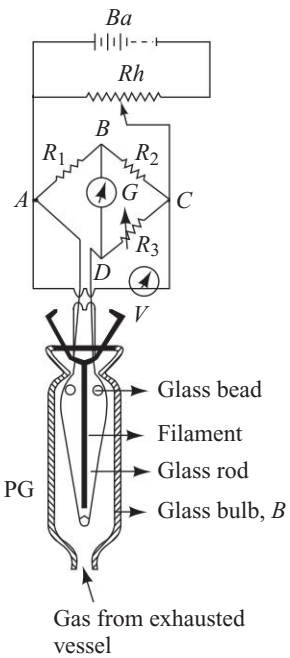


Fig. 3.6 *Pirani guage*

Uses and Drawbacks of the Gauge This gauge is used to measure the pressure between 10^{-2} to 10^{-4} mm of Hg. This gauge is very useful to measure the pressure fluctuations, since the change of resistance of the filament with pressure is instantaneous.

The drawbacks of this gauge are listed as follows:

(i) It is not suitable for use with organic vapours, since they position the filament.
(ii) It is more sensitive to mechanical shocks, accidental vibration, care should be taken while using this instrument.
(iii) It is calibrated using a McLeod gauge, this is not an absolute gauge.
(iv) Pressures below 10^{-4} mm of Hg cannot be measured using this gauge and
(v) At pressures below 10^{-3} mm of Hg, the heat loss due to radiation is higher than conduction. Therefore, the accuracy of this gauge gets affected below 10^{-3} mm of Hg.

The McGraw·Hill Companies

3.5.2 Cold Cathode Ionization Gauge

This gauge is also called Philips ionization gauge after their first manufacturer. F.M. Penning first explained the operating principle of this gauge in 1937 and hence, this gauge is also known as Penning gauge.

The schematic representation of a cold cathode ionization gauge is shown in Fig. 3.7. In this cold cathode ionization gauge, the electrons are withdrawn from the cathode by a high potential difference. In a hot cathode ionization gauge, the electrons are emitted from the filament by thermionic emission. These two ionization gauges differ from each other in the method by which the ions are produced. In a cold cathode ionization gauge, since the electrons are withdrawn from the cathode, the electron emission is very low and hence, the collision between the gas molecules and the electrons is also low, if the electrons travel in straight line. Therefore, to increase the path length of the electrons, a magnetic field is applied around the tube to deflect the electrons. Due to the application of the magnetic field, the electrons travel in spiral path as they move across a magnetic field to the anode. The spiraling action of the electrons increases the path of travel of the electron and hence, the chance of the collision between the gas molecules gets increased. Due to these collisions, the gaseous ions are produced. A microammeter measures the ion current, which is directly proportional to the pressure of the gas, whose pressure is being measured.

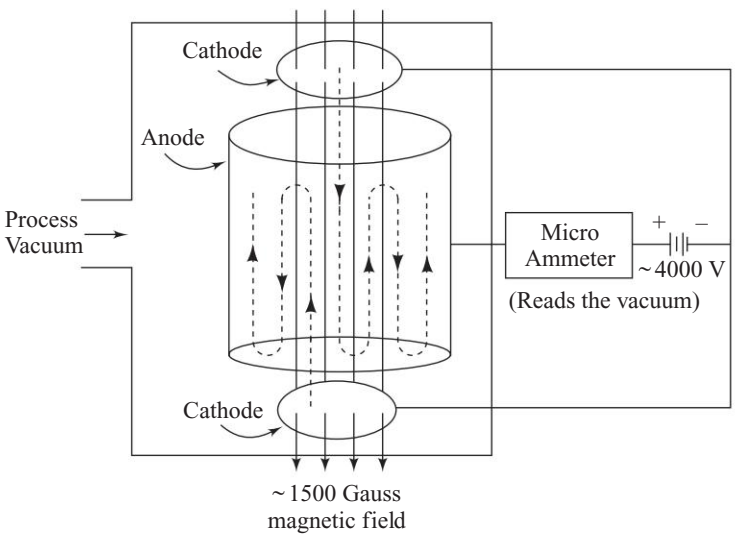


Fig. 3.7 *Penning guage*

The detectable vacuum range is from 10^{-2} to 10^{-7} mm of Hg (1.3 to 1.3×10^{-5} Pa) and the minimum life span of this instrument is one decade. The inaccuracy of the cold cathode unit is about ± 20% at an absolute pressure of 10^{-5} mm of Hg.

The advantages of the cold cathode gauges, compared to hot cathode ones, include their lower cost and that they do not burn out or do not subject the process gas to thermal stress. The disadvantages include their lower accuracy, nonlinear output signal, and gas takes–up caused by high voltage operation. In order to remove polymerized organic contaminants, periodic cleaning of the electrodes and vacuum chamber is required.

The McGraw·Hill Companies

3.6 IMPORTANCE OFVACUUM IN INDUSTRIES

The following are importance applications of vacuum technology in industries:

- Vacuum cleaners are specially designed to cater to the major cleaning processes in most of the industries, such as textile, food, electronics, pharmaceuticals, environment, etc.
- In metallurgy, the high-vacuum furnaces are mainly used for the fabrication of ultra-pure materials and alloys.
- Vacuum drying helps to obtain the residue–free drying of metallic components.
- Vacuum generation is essentially required to obtain empty industrial gas cylinders, decarburizing the drinking water bottles, etc.
- The high vacuum is a must for the production of fluorescent lamps, such as halogen, xenon and incandescent lamps.
- High vacuum is essentially required for the production of x-rays, electrons, etc.
- In industry, the leak detection is very much required both for safety and as well as for clean environment. The vacuum technologies are used to explore the leak detection with very high accuracy.
- In materials characterisation, studies such as SEM, TEM and STEM, a high vacuum is essentially required in the sample region for the stability of the sample and controlling the charge effect.

3.7 VACUUM TECHNIQUES APPLICATIONS

The vacuum technology finds applications in many fields.

(i) Optics To prepare optical mirrors, antireflection coatings on glass, interference filters, etc.

(ii) Solid state devices To prepare solar cells, electrical resistors, capacitors and in microelectronics, to prepare integrated circuits and the production of semiconductor active devices by the method of ion implantation.

(iii) Electron physics To prepare cathode ray tubes especially in domestic television, receiver tubes in colour television, cathode ray oscilloscope, electron microscopes and X-ray machines, etc.

(iv) In industry The electron beam welding used in automobile manufacturing industry, vacuum drying, vacuum freeze-drying, distillation, molecular distillation, vacuum metallurgy (including metal degassing), etc.

(v) Surface studies To study about the emission properties of the surfaces using X-ray photoelectron spectroscopy (XPS), X-ray fluorescent spectroscopy (XFS), etc., Scanning electron microscope (SEM) and atomic force microscopy (AFM) of the surface is used to study dimensional analysis of coatings.

(vi) Mass spectroscopy To study about the mass of the isotopes and to separate the isotopes, etc.

The other uses of vacuum techniques include the areas of rough vacuum used in the field of oil explorations under the seas, and so on.

The McGraw·Hill Companies

Keypoints to Remember

- The unit of pressure is torr. 1 torr is equivalent to 1 mm of mercury column
- The pressure level is also represented by other units like pascal, bar, etc.
- The equivalent of 1 bar is 760 torr and 1 torr is 130 pascal.
- Different categories of vacuum pumps are oil pumps, mercury pumps, molecular pumps and diffusion pumps.
- Rotary pump is used to obtain vacuum in the order of 10^{-3} torr.
- Rotary oil pump is used to obtain very high vacuum in the order of 10^{-3} mm at a faster rate.
- Diffusion pump is used to obtain vacuum in the order of 10^{-8} torr in a clean vessel.
- Molecular pumps are used to obtain vacuum of 10^{-7} mm.
- Low pressures are measured using two categoryies of instruments, namely, manometers and gauges.
- Instruments under the manometer category are mercury, mechanical and viscosity manometers.
- Instruments under gauges are radiometer, conductivity, ionization and diffusion gauges.
- Pirani gauge is used to measure the pressure between 10^{-2} to 10^{-4} mm of Hg.
- Cold cathode ionization gauge is used to measure the pressure in the order of 10^{-2} to 10^{-7} mm of Hg.
- Vacuum techniques are used to explore the leak detection to a very high accuracy.
- High vacuum is essentially required in the sample region to obtain the required stability and conductivity of the charge effect in studies like TEM, STM and STEM.

Objective Questions

3.1 One torr = _____ pascal.
3.2 The pressure level produced by oil pumps is up to_____.
3.3 Molecular pumps are used to obtain a pressure level up to _____.
3.4 _____ is the pressure level obtained for a diffusion pump.
3.5 _____ operation of a vacuum pump is required to obtain constant vacuum in a vessel.
3.6 The principle behind the rotary pump is_____.
3.7 An efficient high vacuum is obtained by immersing the rotary oil pump in a _____

The McGraw·Hill Companies

3.8 Rotary vacuum pump is used to obtain high vacuum of _____ mm at a faster rate.
3.9 The molecular diffusion pump requires a _____ of the order of 10^{-1} to 10^{-2} mm in the exhaust system.
3.10 _____ pump is used to obtain the fore vacuum.
3.11 Measurement on Pirani resistance gauge is based on _____ principle.
3.12 _____ and _____ alloys are used to make resistance in Pirani gauge for vacuum measurements.
3.13 Pirani gauge can not measure the pressure below _____ mm of Hg.
3.14 Pirani gauge is used to measure the vacuum in the range of _____ to _____ mm of Hg.

KEY

3.1 133	3.2 10^{-2} mm
3.3 10^{-5} mm	3.4 10^{-6} mm
3.5 Continuous	3.6 displacement of gas
3.7 oil	3.8 10^{-3} mm
3.9 fore vacuum	3.10 Backing
3.11 Thermal conductivity	3.12 manganin, minalpha
3.13 10^{-4} mm	3.14 10^{-2} to 10^{-7}

Short Answer Questions

3.1 What is meant by vacuum?
3.2 Mention the two methods to produce the high vacuum.
3.3 Mention any two applications of vacuum technology?
3.4 What are the different types of vacuum pumps?
3.5 What is the principle behind rotary pumps?
3.6 Mention the applications of rotary pump.
3.7 What is the limitation of the rotary pump?
3.8 What is meant by pumping speed?
3.9 Explain the role of backing pump.
3.10 What is meant by roughing?
3.11 Define throughput.
3.12 Explain the working principle of diffusion pump.
3.13 What are the advantages of diffusion pump?
3.14 What is the turbo molecular pump?
3.15 Explain the principle of molecular pump?
3.16 What are the limitations of molecular pump?
3.17 Mention the instruments used to measure vacuum.
3.18 What is the basic principle of pirani gauges?
3.19 Explain the uses and drawbacks of pirani gauge.
3.20 What principle the penning gauge is based?

The McGraw·Hill Companies

3.21 Mention the few applications of penning gauge.
3.22 What are the disadvantages of penning gauge?.
3.23 What is the limitations of the pirani gauge?
3.24 Explain the salient features of vacuum technology in industries.

Descriptive-Type Questions

3.1 Explain with a neat sketch the construction and working of rotary pump and diffusion pump.
3.2 Explain the construction and working of molecular pump with suitable schematic diagram.
3.3 Explain with a neat sketch how the construction and working of Pirani and Penning gauges are used to measure the pressure.
3.4 Explain the construction, working and advantages of mercury diffusion pump and Pirani guage.

Chapter

4

Acoustics

OBJECTIVES

- To understand the various categories of sound waves
- To study the different phenomena associated with sound waves
- To derive Sabine's formula for reverberation time and to study the decay and growth of energy in a hall
- To determine the absorption coefficient of sound absorbing materials
- To discuss the principles behind the design of a good auditorium hall
- To analysis the sound insulation and noise pollution

4.1 INTRODUCTION

Sound waves have the same properties as that of ordinary light waves. Hence, they exhibit most of the phenomena of light waves. Due to the available instrumentations facility, and wide range of frequency and energy, sound waves find potential applications in industries. Acoustics is a branch of physics which deals with the generation, propagation, reception and analysis of sound waves. The important applications of acoustics are acoustical instrument design, electro-acoustics, architectural acoustics, bioacoustics, medical acoustics and musical acoustics.

In this chapter, classification, characteristics, properties and applications of acoustics, in particular architectural acoustics, are discussed briefly:

4.1 CLASSIFICATION OF SOUND

Sound is a form of energy and it is produced by a vibrating body. For example, when a hammer strikes the gong of a bell, it produces sound. Sound waves are classified into the three categories based on their frequencies:

(i) Low frequency inaudible sound—Infrasound (< 20 Hz)
(ii) Audible sound (20 Hz–20 kHz)
(iii) High frequency inaudible sound—Ultrasound (> 20 kHz)

Sound waves whose frequencies are less than 20 Hz are known as *infrasound*. They are used to monitor natural events such as volcanic eruption, earthquakes and tsunamis. Sound waves whose frequencies lie between 20 Hz to 20 kHz are known as *audible sound waves*. The sound waves whose frequencies are above 20 kHz are known as *ultrasound*. In the following sections, audible sound waves are discussed in detail along with applications while the ultrasound waves and their applications are given in Chapter 5.

The audibles sound waves are further classified into two categories based on their properties as,

(i) Musical sound
(ii) Noise

4.2.1 Musical Sound

Sound waves which produce a pleasing effect to our ears, while receiving then are known as *musical sound waves*. The typical waveform of the musical sound waves produced by musical instruments is shown in Fig. 4.1. Musical sound waves are produced when a series of similar impulses follow each other at regular intervals without any sudden change in their amplitude.

Examples of musical sounds are sounds produced by a tuning fork, sitar, violin and piano.

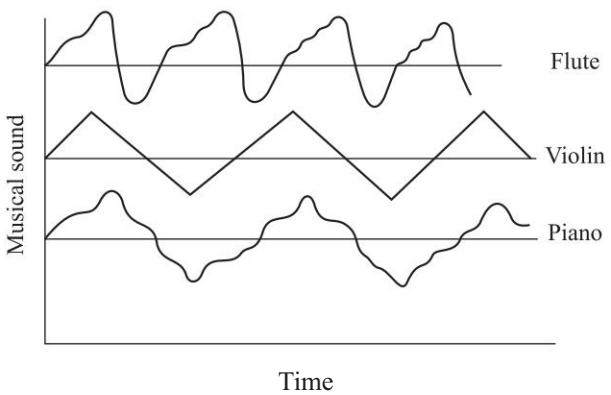


Fig. 4.1 *Musical–sound Waveform*

The important characteristic properties of a musical sound are as follows:

(i) It has a regular waveform.
(ii) It has a definite periodicity.
(iii) There is no sudden change in amplitude.

4.2.2 Noise

Sound waves which produce an unpleasant or displeasing effect or jazzing effect to our ears are known as noise. A typical waveform for noise is shown in Fig. 4.2. Noise sounds of high intensity may are produce permanent or temporary deafness.

The McGraw·Hill Companies

Examples of noise road traffic, moving furniture, explosion of bombs, etc.
A noise sound is characterised by the following properties:

(i) It has an irregularies shaped waveform.
(ii) It lacks periodicity.
(iii) There are sudden changes in amplitude and frequency.

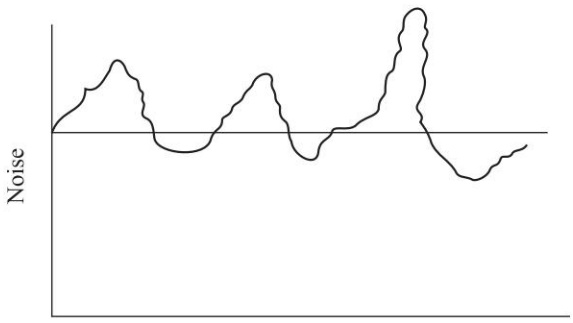


Fig. 4.2 *Noise waveform*

4.3 CHARACTERISTICS OF MUSICAL SOUNDS

Musical sounds consist of two or more than two notes or voices. Therefore, a way to distinguish one from each other is essential. The following are the three characteristic parameters used to analyse musical sound waves:

(i) Frequency or pitch
(ii) Intensity or loudness,
(iii) Quality or timbre

Let us discuss the above parameters briefly in the following sections:

4.3.1 Frequency Or Pitch

Frequency is defined as the number of cycles or oscillations produced per second. The pitch is directly proportional to the frequency. Frequency is a physical quantity and hence, it is measured accurately. The unit used to measure the frequency is Hertz (Hz). Pitch is the characteristic of sound which distinguishes between a shrill sound and a grave sound. The pitch is frequency dependent and hence, it is a psychological quantity. The pitch of a note is a psychological quantity which is merely a mental sensation to convey to our brain when sound waves reaches our ears. The high-pitch sound corresponds to a high-frequency sound and vice versa. The voice produced by women and children are high in pitch due to their high frequency. Similarly, the voice of an aged male has low pitch.

Generally, musicians identify pitch of a musical note by a set of letters. For example, the pitch of note produced by a 256-Hz sound is marked as a *middle* C. The pitches of notes produced by different frequencies of sound is given Table 4.1.

Table 4.1 *Pitches of notes for different frequencies*

Frequency Hz	*Pitch*
256	Middle C
288	D
320	E
342	F
384	G
426	H
480	I
512	J

4.3.1 Intensity or Loudness

Intensity is a physical quantity and hence, it is measurable. It is defined as the ratio of sound energy flowing per unit second to unit area at a point held normal to the direction of the propagation of sound waves. The unit for the intensity is W m^{-2} or J s^{-1} m^{-2}.

Let Q be the energy of sound and it is produced by the source of area A. Therefore, the intensity of the sound wave

$$I = \frac{Q}{tA} \tag{4.1}$$

where t is the time.
Similarly, the intensity of sound in terms of power P of a source is

$$I = \frac{P}{A} \quad \text{and} \tag{4.2}$$

where the power

$$P = \frac{Q}{t} \tag{4.3}$$

Loudness is a psychological quantity. It is defined as the degree of sensation produced on the ears and it depends on the listener. Therefore, it is difficult to measure accurately since it depends on the individual listener.

Weber–Fecher Law

Loudness and intensity of sound waves are related to each other by the Weber – Fecher law. According to this law, loudness L of the source of sound is directly proportional to logarithm of intensity or

$$L \propto \log I \tag{4.4}$$

The McGraw·Hill Companies

Greater the intensity of a sound wave, greater the loudness. Loudness in intensity depends on amplitude, surface area, distance between source and listener, density of medium and air.

Measurement of Intensity of a Sound Wave

The intensity of sound is defined as the average rate of transfer of sound energy per unit area. Even though intensity is a physical quantity to measure absolutely; the relative intensity measurement is widely used for practical purposes than absolute measurements. In order to measure relative intensity, a standard intensity is required. The standard intensity is known *as threshold of audibility* (0 dB) for a note of frequency of 1000 Hz. The value of standard intensity is I=10^{-16} W cm^{-2} or 10^{-12} W m^{-2}.

$$I_0 = 10^{-16} \tag{4.5}$$

Therefore, the intensity level of a sound wave is defined as the ratio of intensity to the standard intensity.

Therefore, the relative intensity $$\beta = \log_{10}\left[\frac{I}{I_0}\right] \tag{4.6}$$

where I_o is the standard intensity and I, the intensity of the source.

The unit for intensity level is *bel*. It is named as bel to honour the inventor of the telephone, Alexander Graham Bell.

Relative intensity $$\beta = \log_{10}\left[\frac{I}{I_0}\right] \text{ bel} \tag{4.7}$$

The relative intensity between two sound notes is defined as 1 bel when the intensity of one sound is 10 times intense than the other.

Therefore, the intensity of sound $$I = 10\, I_0 \tag{4.8}$$

Substituting the I value in Eq. (4.7), we get the relative intensity in bels.

$$\beta = \log\left[\frac{10I_0}{I_0}\right] \text{ bel} \tag{4.9}$$

Simplifying Eq. (4.9), we get,

$$\beta = 1 \text{ bel} \tag{4.10}$$

The relative intensity β takes the value 2 bels, 3 bels etc., when $I = 100\ I_0$, $I = 1000\ I_0$, etc. For practical applications,one cannot use a large unit like bel and hence, a small unit known as *decibel* (dB) is used for relative intensity measurements,

i.e., $$1 \text{ bel} = 10 \text{ decibel} = 10 \text{ dB} \tag{4.11}$$

In order to study the physical significance of one decibel, let us take $\beta = 1$ decibel.

Relative intensity $$\beta\,(\text{in decibel}) = 10\log_{10}\left[\frac{I}{I_0}\right] \tag{4.12}$$

where, $\beta = 1$ decibel.

Equation 4.12 becomes $$1 = 10\log_{10}\left[\frac{I}{I_0}\right]$$

or $$\log_{10}\left[\frac{I}{I_0}\right] = \frac{1}{10} = 0.1$$

Therefore, $$\frac{I}{I_0} = 1.26$$

or $$I = 1.26\, I_o \tag{4.13}$$

The increase in sound intensity level by 1 dB increases the intensity by 26% level. Similarly, an increase of 2 dB sound intensity level leads to an increase in intensity level of $(1.26)^2$.

The intensity level of 0 dB to 120 dB is normally detected by our ears. When the loudness reaches 120 dB, it produces a feeling of pain in our ears. Therefore, it is called as *threshold of feeling* or *pain threshold*. The intensity levels of different sources of sound and their loudness is given in Table 4.2.

Table 4.2 *Intensity and loudness level of sound sources*

S.No	*Source of Sound*	*Intensity level*	*Nature*
1.	Threshold of hearing	0	0
2.	Rulstle of leaves	10	Feeble
3.	Whisper	15 – 20	Feeble
4.	Average house	40	Feeble
5.	Quiet automobile	50	Moderate
6.	Average conversation	60 – 65	Moderate
7.	Heavy traffic road	70 – 80	Loud
8.	Roaring of lion at 20 ft (or) factory noises	90	Loud
9.	Orchestra	100	Very loud
10.	Thunder	100 – 110	Very loud
11.	Threshold of pain	120	Very loud

Phon

Phon is the unit for loudness. A standard source of frequency of 1,000 Hz is used. The loudness is measured by relative measurement only to measure the loudness of a sound wave. The sound wave whose loudness is to be measured is placed near the standard source. The intensity of the standard sound source is first measured. The loudness of the standard source is increased until its loudness value is equal to that of the unknown source. Let *N* decibels be the intensity level above the standard values ,then, the loudness of the source is *N* phons. The interval of loudness corresponding to one decibel on the intensity scale is called one phon. We know that decibel is the unit for intensity or power of sound, whereas phon is a unit of intensity for loudness.

Sound Pressure Level

Let P_1 be the given pressure of sound and $P_{ref,}$ the pressure of reference sound level. The pressure of the reference sound level is taken as standard and is equal to 2×10^{-5} N m^{-2}. 20 times the logarithmic ratio in base 10 of sound pressure of a sound to a reference pressure is known as Sound Pressure Level (SPL). Therefore, sound pressure level,

$$\text{SPL} = 20 \log \frac{P}{P_{ref.}} \tag{4.14}$$

The SPL is the most-often used term to measure the magnitude of sound. The sound pressure level is a measure of acoustic force per unit area. The unit for SPL is dB. The sound pressure level is represented along the distance of measurement from the source.

Sone

We know that loudness is a physical quantity. Phon is a unit for loudness and is not a true scale of loudness. Therefore, another unit known as sone is used to define loudness of 1000 Hz of sound produced at 40 dB sound-pressure level which is sensed by both ears. When SPL crosses 40 dB, loudness in sone will double for an increment of about 10 dB. The relationship between SPL and the common sone are given in Table. 4.3.

Table 4. 3 Relationship between decibel and loudness scales

S.No	*Examples*	*Sound Pressure Level (dB)*	*Loudness in sone (dB)*
1	Jet plate 30 m away	140	1024
2	Threshold of pain	130	512
3	Threshold of discomfort .	120	256
4	Chainsaw at 1 m distance	110	128
5	.Disco, 1 m from the speaker	100	64
6	Diesel Truck 10 m away	90	32
7	Kerbside of busy road, 5 m	80	16

8	Vacuum cleaner, distance 1 m	70	8
9	Conservational speech, 1 m	60	4
10	Average home	50	2
11	Quiet home	40	1
12	Quiet bedroom at night	30	–
13	Background in TV studio	20	–
14	Rusteling leaf	10	–
15	Threshold of hearing	0	–

4.3.3 Timbre (Quality)

The characteristic of a sound wave by which we can distinguish between any two or more than two musical notes or voices having the same pitch and intensity is known as quality or timbre. Most sound waves, including musical notes, are not pure tones *M* they are rather a mixture of different frequencies (tones). Thus, all sound waves, including musical and vocal sounds, have complex waveforms. Consider a musical tone produced by a piano, obtained by regular vibrations of the source. Thus, a musical tone gives a waveform at a specific frequency. Consider another musical tone produced by an instrument at a specific frequency which is different from the first one. When the above two tones are sounded together, a complex waveform is obtained. The combined musical tone has the same frequency as that of the first musical tone with an increase in amplitude of waveform. An observer can easily distinguish the first and second musical tone when listening to a complex waveform. This is possible by considering the sound quality of musical tones like pitch and loudness.

4.4 TRANSMISSION OF SOUND

Sound waves are longitudinal in nature. When sound waves are propagated through a medium, alternative compression and rarefaction take place. The compression of waves produce heat while rarefaction produces cooling effects which finds no time to dissipate to the surrounding medium. A sound wave exhibits properties like reflection, refraction, diffraction, etc., as that of an electromagnetic radiation.

Consider that a sound wave propagating in medium I is made incident on Medium II. During this process, a part of the incident sound wave is reflected back, a part of the sound is absorbed and the remaining part is transmitted through the second medium. The intensity of the transmitted sound wave depends on the incident sound wave, characteristic impedance of the two mediums and the angle of the incident sound wave. The characteristic impedance *Z* or specific *acoustic impedance,* is a frequency-dependent parameter. It is defined as the ratio of acoustic pressure *P* in a medium to the associated particle velocity *u*.

Let *c* be the velocity of sound in a medium and ρ_0 the equilibrium density. Then the

$$\text{acoustic impedance } = \rho_0 c = Z = \frac{P}{u} = \frac{\rho_0 c u}{u} \tag{4.15}$$

The McGraw·Hill Companies

Consider a sound wave is made incident on a medium, under normal incidence. Let V_1 and V_2 be the velocity of sound waves respectively for mediums I and II. Let ρ_1 and ρ_2 are the densities of mediums I and II respectively. Therefore, the amplitude of the reflection coefficient is

$$a_r = \frac{\rho_2 V_2 - \rho_1 V_1}{\rho_1 V_1 + \rho_2 V_2} = \frac{Z_2 - Z_1}{Z_1 + Z_2} \tag{4.16}$$

Similarly, the amplitude of the transmission coefficient is

$$a_t = \frac{2\rho_2 V_2}{\rho_1 V_1 + \rho_2 V_2} = \frac{2Z_2}{Z_1 + Z_2} \tag{4.17}$$

Case I When the characteristic acoustic impedance of Medium *I* is greater than that of Medium *II* ($Z_2 > Z_1$), the reflection coefficient a_r is positive. Therefore, the reflected and transmitted waves are in the same phase. For example, when a sound wave is made incident on a rigid surface, there is no phase change in the reflected waves. A pulse of compression waves will be reflected from the wall as another pulse of compression waves. Similarly, a pulse of rarefaction will be reflected from the wall as another pulse of rarefaction. The above result confirms that there will not be any phase change when the sound wave is incident on a rigid body.

Case II When characteristic impedance of Medium *I is* less than Medium *II* ($Z_2 < Z_1$), the reflection coefficient is negative. As a result, with reference to the incident beam, the reflected wave is 180^0 out of phase. Consider that a sound wave is made incident on a yielding surface. The rarefaction wave is reflected as a compressed wave, while the compressed wave is reflected as a rarefaction. Therefore, the sound wave incident on a yielding surface makes a phase change. The phase of the reflected wave is opposite to that of incident wave.

Case III When the characteristic impedance of Medium *I* is equal to Medium *II* ($Z_1 = Z_2$), the reflection coefficient is zero. Under this condition, all the incident sound energies are transmitted into the second medium. This means that there is no boundary between medium I and II. For example, an open window transmits the entire sound energy through it.

Case IV When the ratio of characteristic impedance of Medium *I* to Medium *II* is infinite ($Z_1 / Z_2 = \infty$), the reflected and transmitted waves are in the same phase. The reflection is from a highly compressive or very dense medium. Therefore, the waves are reflected only with a slight reduction in amplitude without any change in phase. Thus, it establishes a pattern of standing wave in Medium I, with antinodes at the boundary with double the amplitude of the incident wave.

Case V When the ratio of characteristic impedance of Medium *I* to Medium *II* is zero ($Z_1/Z_2 = 0$), the reflected and transmitted waves are in the same phase. The reflection is from an easily incompressible or rarefaction medium. Therefore, the amplitude of reflected wave is almost equal to that of the incident wave. Further, the reflected waves establish a pattern of standing wave in Medium I with a node of zero amplitude of incident wave.

The McGraw·Hill Companies

4.5 TRANSMISSION LOSS

During transmission of sound waves, a part of the amount of sound energy is lost in the form of heat and non-geometrical effects. The losses are known as transmission loss. The transmission loss for random incident sound waves is less than that of a normal incident wave. Let I_i be the intensity of incident sound waves and I_t be the intensity of transmitted sound waves.

Therefore, the transition loss for a normal incident is $TL = 10 \log \frac{I_i}{I_t}$ (4.18)

Let α_t be the transmission coefficient of sound energy, which is defined as the ratio of the intensity of transmitted to incident sound.

Therefore, $\alpha_t = \frac{I_t}{I_i}$ (4.19)

Substituting Eq. (4.19) in Eq. (4.18), we get,

Transmission loss $TL = 10 \log (1/\alpha_t)$ (4.20)

Consider that an enclosure consists of different kinds of surfaces with different transmission coefficients. Therefore, the transmission loss is

$$TL = 10 \log \left[\frac{s}{s_1 \alpha_{t1} + s_2 \alpha_{t2} + s_3 \alpha_{t3} + \ldots.} \right] \text{dB} \quad (4.21)$$

where s is the total surface area of all surfaces in the enclosure s_1, s_2, s_3, etc., and α_{t1}, α_{t2}, etc., are the surface areas and transmission coefficients of the sources 1, 2, 3, etc. Generally, transmission loss is less for a randomly incident sound wave than that of a normal incident wave.

4.6 ACOUSTICS OF BUILDINGS

Today sound engineering has become an important requirement in areas like film-making, broad-casting, traditional music shows, dance and drama. Acoustics of buildings is branch of physics which deals with the control of sound in a room or hall or an auditorium. The purpose of this control is to create the required conditions wherein the audience gets to here the sound comfortably. The acoustical properties of rooms or halls have considerable effect on the clarity and intelligibility of speech and music. The branch of science which deals with the design and construction of rooms, halls or auditoria with an optimum acoustical efficiency (i.e., best sound effect) is known as *architectural acoustics* or *acoustics of buildings.* The phenomena like reflection, interference, absorption of sound waves and reverberation are described in detail in architectural acoustics. Wallace Clement Sabine did pioneering work in architectural acoustics for more than two decades. According to his research, the following are the conditions required for a hall or an auditorium to consider it acoustically fit.

(i) The uniformly distributed intensity of sound heard by the audience must be sufficiently loud irrespective of location.

The McGraw·Hill Companies

(ii) The tonal quality of speech and music must be unchanged.
(iii) For the sake of clarity, successive sounds of the speech and music should be uniformly distributed.
(iv) There should not be any overlapping of sound in a hall.

The general design of a hall to get uniform distribution of sound is shown in Fig.4.3. The loudspeakers of the auditorium have a parabolic shape. When a person speaks, sound waves are reflected from the parabolic wall in almost all directions. The reflected waves are parallel beams and get distributed uniformly throughout the hall.

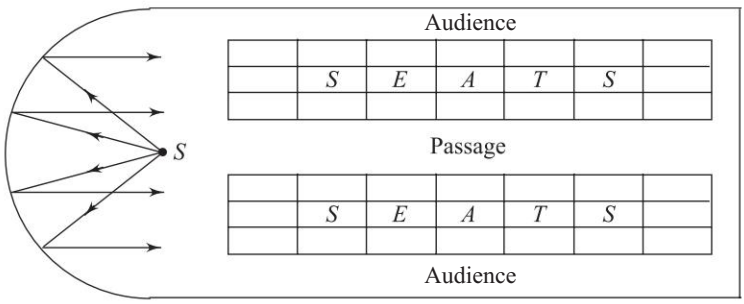


Fig. 4.3 *Sound — uniform distribution in a hall*

Rooms or halls with a high absorption rate of sound waves are called as *dead* or *anechoic rooms*. Rooms with low absorption rate of sound waves are called *live* or *reverberant anechoic rooms*. Therefore, in order to obtain good acoustic property, the hall or room should be constructed by considering the following factors:

(i) Reverberation or prolonged reflection of sound
(ii) Loudness or intensity of sound
(iii) Echelon effect
(iv) Extraeous noises
(v) Resonance within the building
(vi) Noise
(vii) Focusing

The factors which are used to construct a hall or room with good acoustical properties are discussed in detail under the following headings.

4.6.1 Reverberation

Reverberation is an important factor to be considered in the design of a hall/theatre/music hall with good acoustical properties. When a source of sound is operated continuously in a room, it generates sound waves in all directions like the walls, floors, etc. These sound waves undergo two to three hundred multiple reflections, before they become inaudible. As a result, even though the source is cut off, sound waves persist for a short period of time. Therefore, listeners sense the persistence of a sound wave until it becomes very weak. The existence or prolongation of sound in a room even when the sound source is cut off is known as *reverberation.*

The McGraw·Hill Companies

Reverberation Time It is the time taken by a sound wave to fall from an average value to inaudibility. In order to define reverberation time, Sabine used the sound wave produced by an organ pipe, as reference. The intensity of sound waves produced by an organ pipe, operated at a frequency of 512 Hz, becomes inaudible when its intensity is reduced to one millionth of its intensity just before stopping the organ pipe. Therefore, time taken by a sound to reduce its intensity to one-millionth of its intensity when the source is cut off is known as *standard reverberation time*.

According to Sabine, reverberation time for a hall of volume V is

$$T = \frac{0.163V}{A} \; S \tag{4.22}$$

where A is the total absorption in Open Window Unit (OWU) or sabine.

$\therefore$ the total absorption $$A = \Sigma as \tag{4.23}$$

where s is the total surface area of the sound-absorbing material present in a hall in m^2 and a, the coefficient of absorption or average sabine absorptivity.

The problem of specifying an optimum reverberation time for rooms or halls is more complex for specific applications. For example, the optimum time required for a music hall is different from that of a lecture hall. Therefore, the music rooms should be more reverberant than lecture halls. The optimum value of reverberation time required for a lecture hall is 0.5 second while for a music hall it ranges between 1 and 2 seconds. Similarly, for medium sized theatres, the optimum value of reverberation time ranges between 1.1 to 1.5 seconds. However, for a larger one, its value goes up to 2.3 seconds.

Factors Controlling Reverberation Time

The following are the factors which play a major role in controlling reverberation time of a hall:

(i) Reflection of Sound The persistence of sound energy inside a hall depends on the reflection coefficient of sound energy from the walls, ceiling and floors. For example, when the reflection coefficient is very low, the sound energy persists for a short period of time and this becomes inaudible. On the other hand, if the reflection coefficient is very high, the sound energy exists in hall for a longer period of time.

(ii) Absorption Coefficient of Materials present in a Hall The coefficient of absorption of material like curtains, cushions, chair, wood floor, etc., is high. Therefore, inclusion of the above materials inside a hall results in absorption of more energy. As a result, reverberation time is reduced.

(iii) Intensity of Sound When the intensity of sound is large, it exists for a longer period of time, while for low intensity of sound it can stay only for a short period of time.

(iv) Frequency of Sound High-frequency sound waves are easily absorbed than low-frequency sound. When the frequency of a sound wave is low, reflection is also low.

The McGraw·Hill Companies

(v) Volume of a Hall The reverberation time (T = 0.161 V/A) is directly proportional to the volume of a hall. Hence, the influence of the volume of a hall is large on reverberation time.

4.6.2 Coefficient of Absorption

sound absorbing materials play an important role in controlling the sound energy in a hall. The absorption or reflection of sound energy depends on sound absorbing materials used inside the building. Thus, knowledge on absorption of sound energy of sound absorbing materials is more essential. It is measured by the relative measurement method, i.e., the unit area of an open window is used as reference. since it transmits the entire sound energy falling on it and no sound wave is reflected back. It helps to study sound absorbing properties of materials.

Sound Absorption Coefficient It is defined as the inverse of the area of the sound-absorbing material which absorbs an equivalent of sound energy as that of 1 m^2 of an open window. Generally, coefficient of absorption amount of materials lies between 0 and 1. For example, a sound absorbing material with an area of 3 m^2 absorbs an equivalent amount of sound energy as that of a 1 m^2 open window. Therefore, the coefficient of the absorption of the absorbing material is 1/3 = 0.33 sabine. The coefficient of absorption of an average–built human is 0.45 sabine.

4.6.3 Noise Reduction (NR)

The intensity of sound level in a hall depends on the level of absorption which takes place inside the hall. The intensity of sound level is reduced due to multiple reflections.

Noise reduction in a hall is

$$NR = 10 \log \frac{A_2}{A_1} \; dB \tag{4.24}$$

where A_1 and A_2 are total absorption in sabines of a hall or room before and after the existence of the sound wave. Thus, using Eq. (4.24), the total absorption in a hall or a room can be controlled.

4.6.4 Echo

The sound wave reflected or returned with sufficient magnitude and delay from a distant extended surface like a cliff, hill, well, building etc., is known as echo. The transmitted and reflected sound waves differ in quality as well as characteristics. Consider that a person produces a sound inside a hall. The distance between the person and the wall is x. The sound is reflected from the wall and it takes t seconds to travel the to and fro distance.

Therefore, velocity of sound in air $v = \frac{2x}{t}$

or $$x = \frac{vt}{2} \tag{4.25}$$

The McGraw·Hill Companies

We know that the value of velocity of sound in air is 340 m s^{-1}. One can distinguish the incident sound and reflected sound if there is a small time gap between them. The minimum time gap required to distinguish the two sounds waves is $(1/10)^{th}$ of a second.

Therefore, substituting the values of v and t in Eq. (4.25), we get,

Velocity of sound in air $$x = \frac{340(1/10)}{2} \quad (4.26)$$

$$= 17 \text{ m} \quad (56 \text{ ft})$$

Equation (4.26) clearly states that when the distance between sound source and the reflecting wall is less than 17 m, we cannot hear an echo. On the other hand, if the distance is more than 17 m, we will hear the echo. This will helps to construct an auditorium or a hall which is free from echo.

4.6.5 Creep

Generally, creep occurs when the sound source is close to the surface like domes, walls, etc. The creep occurs mainly due to the multiple reflections of sound waves on a curved surface. The energy of the sound wave remains constant in a hall. It is heard distinctly at any point on the surface. The formation of creep on a curved surface is shown in Fig.4.4.

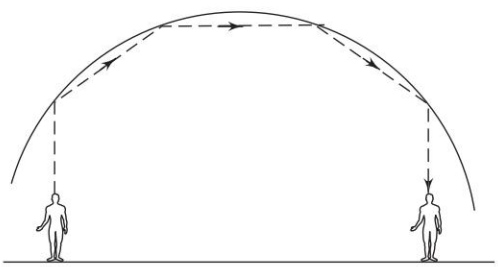


Fig. 4.4 *Creep—curved surface*

4.6.6 Focussing

When the surface of a building is concave ,the sound waves get reflected from surface. The reflected sound waves are focused at one point instead of getting distributed. The intensity of sound waves at the focused point is maximum while it is minimum in other places. This effect is known as focusing effect.

4.6.7 Standing wave

In a small hall or a room, due to multiple reflections, nodes and antinode are built by sound waves. A maximum and minimum in sound pressure levels are obtained in the respective nodes and antinode regions. The formation of standing waves takes place inside a hall. The standing waves does not affect the acoustics of a building at high

The McGraw·Hill Companies

frequencies, rather they will have little effect at low frequency. The standing wave ratio is defined as the ratio of amplitude of antinodes and nodes. It is used to determine the load impedance of a building.

4.7 SABINE'S FORMULA FOR REVERBERATION

Sabine derived reverberation formula for a hall or an auditorium to express the rise and fall of sound intensity by considering the following assumptions:

- Sound energy is uniformly distributed throughout a hall.
- There is interference between sound waves.
- The absorption coefficient is independent of sound intensity.
- The rate of emission of sound energy from the source is constant.

4.7.1 Sabine formula

Consider one side of the plane wall *EG* in a hall. Let *ds* be the small element on the plane *EG* as shown in Fig. 4.5.

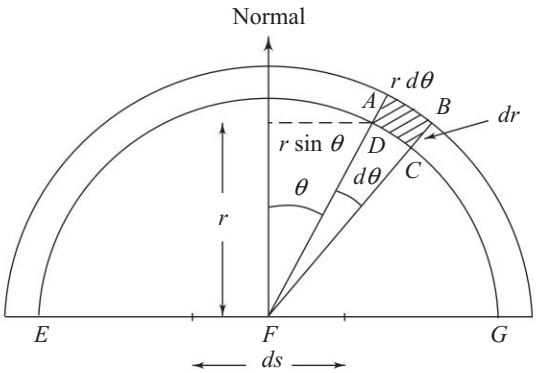


Fig 4.5 *Element* ds *in a plane wall*

Consider a hall of volume *V*. Sound energy is distributed uniformly throughout the hall and hence, element *ds* also receives energy *E*. Draw two semicircles with radii *r* and $r + dr$ from the centre point *F* of the element *ds*. Consider a small shaded region *ABCD* lying between two semicircles drawn with angles θ and $\theta + d\theta$ with the normal to *ds.* Therefore,

$$\text{The area of shaded region } ABCD = r\, d\theta\, dr \tag{4.27}$$

where *dr* and $r\, d\theta$ are respectively the radial and arc length of the shaded portion.

Assume that the whole figure is rotated through an angle $d\phi$ about the normal. Thus, the shaded portion will travel through a small distance. Hence, a region is traced out with volume *dv*

i.e., $dv =$ Area of the region $\times$ circumferential path

$$= r\, d\theta\, dr \times r \sin\theta\, d\phi$$

Therefore, $$dv = r^2 \sin\theta\, dr\, d\theta\, d\phi \tag{4.28}$$

We known that E is the average sound energy per unit volume, and hence, the sound energy present in volume dv, is

$$= E\, dv \tag{4.29}$$

Substituting the dv value from Eq. (4.28), we get sound energy present in volume dv

$$= E\, r^2 \sin\theta\, d\theta\, dr\, d\phi \tag{4.30}$$

The solid angle subtended by ds at the volume element $ABCD$,

$$d\omega = \frac{ds\cos\theta}{r^2} \tag{4.31}$$

A fraction of this sound energy incident on ds from the volume element $ABCD$ is given by,

$$\frac{d\omega}{4\pi} Edv = \frac{ds\cos\theta}{r^2 4\pi} \times E\, r^2 \sin\theta\ dr\, d\theta\, d\phi \tag{4.32}$$

Simplifying the above equation, we get,

$$= \frac{E\, ds}{4\pi} \sin\theta\cos\theta\, dr\, d\theta\ d\phi \tag{4.33}$$

The total sound energy received by the small element ds per second is obtained from the above equation by integrating with respect to r, θ and ϕ. The radius can take the value from 0 to v (the entire volume lying within a distance r of ds). The angle θ changes the value from 0 to $\pi/2$ (0 to 180^o) and angle ϕ changes from 0 to 2π (0 to 360^o).

Integrating Eq.(4.33), w.r.t ϕ we get,

Therefore, energy received in unit time,

$$= \frac{E\, ds}{4\pi} \sin\theta\cos\theta\, dr\, d\theta \int_0^{2\pi} d\varphi$$

Integrating $d\phi$ over the limit 0 to 2π, we get,

$$\frac{E\, ds}{4\pi} \sin\theta\cos\theta\, dr\, d\theta\ \ 2\pi$$

Simplifing the above equation, we get,

$$= \frac{E\, ds}{2} \sin\theta\cos\theta\, dr\, d\theta \tag{4.34}$$

Similarly, integrating Eq. (4.34) w. r. t θ, we get,

$$= \frac{E\,ds}{2}\,dr \int_0^{\pi/2} \sin\theta \cos\theta \; d\theta$$

$$= \frac{Eds}{4} dr \tag{4.35}$$

Integrating Eq. (4.35) *w.r.t v*, we get,

$$\text{Energy received in unit time} = \frac{Eds}{4} \int_0^v dr$$

$$= \frac{E\,v\,ds}{4} \tag{4.36}$$

Consider a is the absorption coefficient of small element ds.

Therefore, sound energy absorbed by ds per unit time.

$$= \frac{E\,v}{4} ds\,a \tag{4.37}$$

Therefore, total sound energy absorbed by the room per unit time.

$$= \frac{1}{4} E\,v \Sigma\, adS$$

$$= \frac{1}{4} E\,v\,A \tag{4.38}$$

where A is the total absorption of all the substances in the hall and is equal to

$$\Sigma\, a\, ds = a_1 s_1 + a_2 s_2 + a_3 s_3 + \ldots \tag{4.39}$$

Where a_1, a_2, a_3, etc., are the absorption coefficients of substances 1,2,3… and s_1, s_2, s_3, etc., are the surface area of substances 1,2,3…

Let P be the power of the sound source, then the maximum sound energy density E_m from the source is

$$P = \frac{1}{4} E_m V\,A$$

$$E_m = \frac{4P}{V\,A} \tag{4.40}$$

If V is the volume of a hall, then total sound energy in a hall is EV.

$$\text{Therefore, the rate of increase in sound energy} = \frac{d}{dt}(EV) = V\frac{dE}{dt} \tag{4.41}$$

$$\text{Similarly, the rate of absorption of sound energy} = \frac{1}{4} E\,v\,A \tag{4.42}$$

The rate of emission of sound energy = Rate of growth of sound energy + Rate of absorption of sound energy (4.43)

Substituting the values of rate of growth and absorption of sound energies respectively from Eq. (4.41) and (4.42), in Eq (4.43), we get,

Therefore, total sound energy supplied form the source per second,

$$P = V\frac{dE}{dt} + \frac{1}{4}E\,v\,A \tag{4.44}$$

Dividing the above equation by V, we get

$$\frac{P}{V} = \frac{dE}{dt} + \frac{v\,A}{4V}E$$

$$\frac{4P}{v\,A}\alpha = \frac{dE}{dt} + \alpha.E \tag{4.45}$$

Substituting $\alpha = \frac{v\,A}{4V}$ and multiplying the above expression by $e^{\alpha t}$, on both sids, we get,

$$\frac{4P}{v\,A}\alpha e^{\alpha t} = \left(\frac{dE}{dt} + \alpha\,E\right)e^{\alpha t}$$

$$\frac{4P}{v\,A}\alpha e^{\alpha t} = \frac{d}{dt}\left(Ee^{\alpha t}\right) \tag{4.46}$$

Integrating the above expression, we get,

$$\frac{4P}{v\,A}\alpha\, e^{\alpha t} + K = Ee^{\alpha t} \tag{4.47}$$

where K is an integration constant which is obtained using the following boundary conditions, namely, growth and decay of energies.

Growth of Sound Energy

Let t be the counted time when the source begins to emit sound. When the source is on, the growth of sound starts at $t = 0$ and $E = 0$.

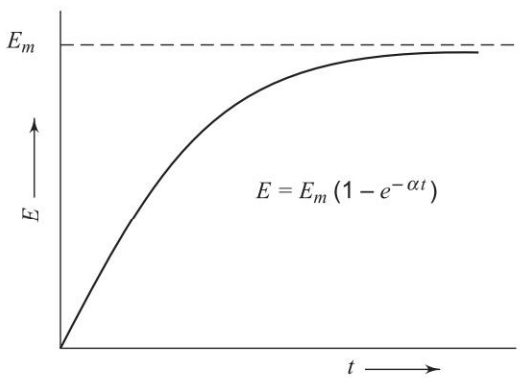


Fig. 4.6 *Growth of energy*

Substituting the boundary condition in Eq. (4.37), we get,

$$\frac{4P}{vA} + K = 0$$

or
$$K = -\frac{4P}{vA} \tag{4.48}$$

Substituting this value in Eq. (4.47), we get,

$$\frac{4P}{vA}e^{\alpha t} - \frac{4P}{vA} = Ee^{\alpha t}$$

Rearranging, we get,

$$\frac{4P}{CA}\left(\ e^{\alpha t} - 1\ \right) = Ee^{\alpha t} \tag{4.49}$$

We know that, $E_{\mathrm{m}} = \dfrac{4P}{vA}$

$$E_{\mathrm{m}}\left(e^{\alpha t} - 1\right) = Ee^{\alpha t}$$

Therefore,
$$E = E_{\mathrm{m}}\,(1 - e^{-\alpha t}) \tag{4.50}$$

Equation (4.50) shows the rate of growth of sound energy with time t in a room of volume V. The exponential function of the growth of energy is shown in Fig.4. 6.

Decay of Sound Energy

When sound source is cut off, the rate of emission of sound energy at $P = 0$. At this point, energy has reached the maximum value, i.e., $E = E_{\mathrm{m}}$.

When $P = 0$ and $t = 0$, Eq. (4.47) becomes,

$0 + K = E_{\mathrm{m}}$

$\therefore\ K = E_{\mathrm{m}}$

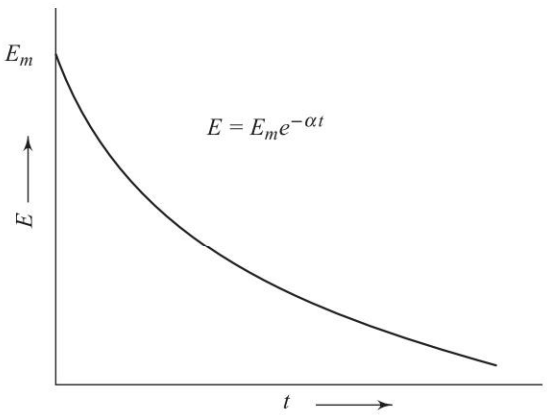


Fig. 4.7 *Decay of sound energy*

Substituting this value in Eq. (4.47), we get,

$$0 + E_m = E\, e^{\alpha t}$$

$$\therefore E = E_m\, e^{-\alpha t} \tag{4.51}$$

Equation (4.51) gives the rate of decay of sound energy with time, when the sound source is switched off. The exponential decay of sound energy in a hall of volume V is shown Fig. 4.7.

Standard Reverberation Time

According to Sabine, the standard reverberation time is the time taken by a sound wave to fall to one millionth of its original intensity even after the source is cut off. E_m is the sound energy of the source before cut off and during standard reverberation time, it reduces to $E_m/10^6$.

Therefore, $$E = \frac{E_m}{10^6}$$

$$= E_m\, 10^{-6} \tag{4.52}$$

we know that, at $t = T$ $$E = E_m\, e^{-\alpha t} \tag{4.53}$$

Substituting the value of $E = E_m 10^{-6}$) in the above equation, we get,

$$E_m\, 10^{-6} = E_m\, e^{-\alpha t}$$

or $$10^{-6} = e^{-\alpha t}$$

$$10^6 = e^{\alpha T}$$

Taking logarithm on both sides,

$$\alpha\, T = 6 \log_e 10$$

or $$T = \frac{6 \times 2.3026 \times 1}{\alpha} \tag{4.54}$$

Substituting the value of $\alpha = \frac{v\,A}{4\,V}$ in the above equation, we get,

$$= 6 \times 2.3026 \times \frac{4\,V}{v\,A}$$

we know that the velocity $v = 340$ m s^{-1}

Therefore, standard reverberation time

$$= \frac{6 \times 2.303 \times 4V}{340 \times A}$$

$$= \frac{0.163V}{A} \text{ s}$$

The McGraw·Hill Companies

Therefore, the standard reverberation time is

$$= \frac{0.163\,V}{A} \tag{4.55}$$

Equation (4.55) gives the equation for standard reverberation time T for a hall or room or an auditorium. The Sabine formula was accepted as a general formula in architectural acoustics to carry out acoustical analysis. Sabine's formula is not applicable for a dead room. For a dead room, the absorption coefficient is 1. Similarly, when sound energy is completely absorbed, $a = 1$, i.e., the reverberation time is zero but according to Sabine,s formula, reverberation time is not zero.

Solid Angle

Let r be the radius of a sphere with the total solid angle ω subtended by the sphere at its centre.

Therefore, the solid angle $\omega = \dfrac{\text{surface area}}{(\text{radius})^2}$

$$= \frac{4\pi r^2}{r^2} = 4\pi \tag{4.56}$$

Consider the surface RS on the sphere at a distance r from the centre C as shown in Fig.4.8.

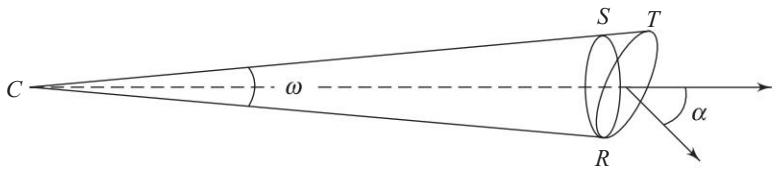


Fig. 4.8 *Solid angle*

The total solid angle subtended by the surface RS at its centre C,

$$= \frac{\text{area}}{r^2}$$

$$= \frac{\text{area of } RS}{r^2} \tag{4.57}$$

Since the surface RS is at a distance r from the centre C, the solid angle subtended by RS at its centre C is

$$= \frac{\text{area of } RT}{r^2}$$

$$\therefore \text{solid angle} \quad \omega = \frac{\text{area of } RS \cos\alpha}{r^2} \tag{4.58}$$

Equation(4.58) gives the solid angle of the sphere.

4.8 MEASUREMENT OF ABSORPTION COEFFICIENT

The measurement of absorption coefficient of absorbing materials is required essentially to optimize standard reverberation time of a hall or room. There are two different methods used to measure the sound absorption coefficient of materials.

4.8.1 Method I

In this method, the absorption coefficient of a sound–absorbing material is determined by relative measurements. Let T_1 be the reverberation time of a hall of volume V which is free from any sound absorbing material. S is the surface area of the hall.

Therefore, reverberation time

$$T_1 = \frac{0.163V}{\sum as} \tag{4.59}$$

A sound absorbing material with absorption coefficient a_1 and surface area s_1 is introduced inside the hall of volume V.

Therefore, the reverberation time $T_2 = \dfrac{0.163V}{\sum as + a_1 s_1}$ (4.60)

Subtracting the inverse of Eq. (4.60) and Eq. (4.59), we get,

$$\frac{1}{T_2} - \frac{1}{T_1} = \frac{\sum as + a_1 s_1}{0.163V} - \frac{\sum as}{0.163V}$$

or

$$\frac{1}{T_2} - \frac{1}{T_1} = \frac{\sum a_1 s_1}{0.163V}$$

The absorption coefficient,

$$a_1 = \frac{0.163V}{s_1}\left[\frac{1}{T_2} - \frac{1}{T_1}\right] \tag{4.61}$$

Equation (4.61) is used to determine the absorption coefficient of materials based on a standard source of frequency of 512 Hz. However, the absorption coefficient of a material is frequency dependent.

4.8.2 Method II

It is also a relative measurement method. Consider the power outputs of two sources as P_1 and P_2. The relative measurement requires, only the ratio of two powers, i.e., P_1/P_2 and not the individual power output. The maximum sound energy densities of two sound emitting sources are,

$$\frac{4P_1}{vA} \tag{4.62}$$

$$\frac{4P_2}{vA} \tag{4.63}$$

Let T_1 and T_2 be the time required for sound waves to become inaudible during decay.

Therefore,

$$E_0 = \frac{4P_1}{vA} e^{-\alpha T_1} \tag{4.64}$$

$$E_0 = \frac{4P_2}{vA} e^{-\alpha T_2} \tag{4.65}$$

Dividing Eq. (4.64) by Eq. (4.65), we get,

$$\frac{P_1}{P_2} = e^{-\alpha(T_1 - T_2)} \tag{4.66}$$

Taking log on both sides of Eq. (4.66), we get

$$\log_e \frac{P_1}{P_2} = \alpha(T_1 - T_2)$$

or

$$\alpha = \frac{\log_e P_1 - \log_e P_2}{(T_1 - T_2)} \tag{4.67}$$

Substituting the value of α in Eq. (4.67), we get,

$$\frac{vA}{4V} = \frac{\log_e P_1 - \log_e P_2}{(T_1 - T_2)}$$

$$A = \frac{4V}{v}\left[\frac{\log_e P_1 - \log_e P_2}{(T_1 - T_2)}\right] \tag{4.68}$$

We know that $A = as$, Eq. (4.68) becomes,

$$as = \frac{4V}{v}\left[\frac{\log_e P_1 - \log_e P_2}{(T_1 - T_2)}\right]$$

Rearrange the above equation, we get,

The coefficient of absorption

$$a = \frac{4V}{vs}\left[\frac{\log_e P_1 - \log_e P_2}{(T_1 - T_2)}\right] \tag{4.69}$$

Equation (4.69) is used to determine sound – absorption coefficient of any material by knowing the surface area, volume of a hall, the reverberation time and ratio between the two powers.

4.9 SOUNDABSORBING MATERIALS

Sound absorbing materials are important in acoustic design of a hall or theatre. Sound absorbing materials are used to reduce reflections of sound waves or increase the sound absorption in a hall. Thus, they helps to obtain the optimum reverberation time depending on requirements of halls or theatres for a particular application. The materials have unique properties and behaviour, suitable for acoustical design. The following are the properties of a sound–absorbing material:

(i) It should be highly porous, inelastic and highly compressive.
(ii) It should be cheap and easily available, easy to fix, good looking, durable, light in weight, of sufficient structural strength, waterproof, of economical maintenance etc.
(iii) It should have good resistance to fire.
(iv) It should be efficient over a wide range of frequencies.

Sound–absorbing materials with the above properties are used for acoustical design. Based on the above properties, sound–absorbing materials are classified into four broad categories, namely,

(i) Porous absorbents
(ii) Cavity resonators
(iii) Resonant absorbents, or panel absorbents
(iv) Composite types of absorbents

4.9.1 Porous Absorbents

When a sound wave strikes a porous material, a small part of the wave gets reflected, while the major part of the sound waves enters into the porous material. This part is converted into heat energy. The small part of the reflected wave from the porous material has less energy and hence, it becomes inaudible and does not cause any interference. The porous absorbents are available in the form of prefabricated, assumable materials, scuh as fiber boards, soft plastics, rock wood, wood wool, glass silk, asbestos fiber spray etc.

4.9.2 Cavity Resonator

A chamber or container with a small opening is known as a cavity resonator. The sound waves which enter the resonator undergo multiple reflections inside the chamber. Thus, energy of sound waves decreases and becomes inaudible due to multiple reflections. The resonators are designed with suitable design and volume to absorb sound waves of any frequency. The drawback of a cavity resonator is that it is not suitable for multiple frequencies. Moreover it is used only for an individual machine, air–conditioning plant, etc.

4.9.3 Resonant Absorbents or Panel Absorbents

A resonant absorbent, or panel absorbent, is fixed on a frame with an air space between the frame and walk. It acts as a panel absorber. When sound waves strike this absorbent, a certain amount of energy is converted into heat energy due to the flexural vibration of the panel. Example are gypsum boards, wood and hardboard panels, suspended plaster, ceilings, rigid plastic board panels, windows, doors, etc.

The McGraw·Hill Companies

4.9.4 Composite Types of Absorbents

When the functions of all the three types described above are combined in a single unit, it is known as a composite absorbent. The composite absorbent consists of a perforated panel fixed over an air space containing porous absorbents. The panel may be metal, wood, plywood, hardboard, plastic board, etc. When the sound strikes the panel, it passes through it and is damped by resonance of the air in the cavity. Examples of such absorbents are empty jars and bottles, perforated cardboard backed by perforated fiber boot, etc.

4.10 FACTORS AFFECTING ACOUSTICS OF BUILDINGS AND THEIR REMEDIES

We know that the factors such as reverberation time, loudness, focusing due to walls and ceilings, echelon effect, extraneous noises and resonance within a building affect the acoustics of a building. In the following section, influences of the above factors on the construction of an auditorium or a hall and their remedies are discussed in detail.

4.10.1 Reverberation Time

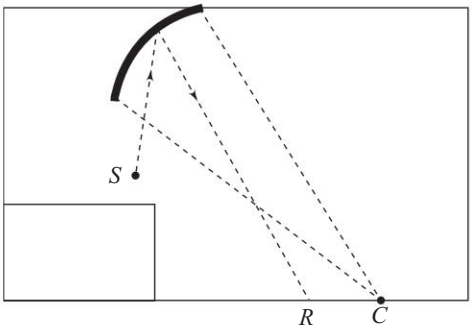
In order to construct an acoustically fit auditorium or hall, the reverberation time of a hall or an auditorium should be optimiesd. For example, when the reverberation time is too high, the sound produced by the speaker will persist for a long period of time. As a result, overlapping of sound waves takes place. The sound wave undergoes multiple reflections before it becomes eradible. Therefore, presence of overlapping, formation of standing waves etc., reduces the quality of sound waves distributed in hall. Thus, listeners are not able to hear the sound clearly. In order to overcome this effect sound-absorbing materials are placed inside the hall. Thus, they help to bring the reverberation time to the required optimum value. Similarly, when reverberation time is too low, sound produced by the speaker dies quickly and it becomes inaudible in a short period of time. In order to improve the sound, reverberation time of a hall should be increased to an optimum value. This can be achieved by decreasing total the absorption coefficient of the volume of the hall i.e., sound reflection boards are placed inside a hall. This helps to increase the reverberation time of the hall. Therefore, by keeping an optimisied reverberation time, the sound level and quality are made to meet the listener requirements. Therefore, in order to make the sound by the time should kept at an value.

4.10.2 Loudness

Similar to reverberation time, loudness of a sound in a hall should be made as an optimum value. This helps us to make the sound audible. We know that reverberation time of hall is directly proportional to loudness. Consequently, low loudness results in the existence of sound for a short period, while a high loudness leads to existence of sound for a longer period of time. A loud sound is painful to the listener. Therefore, sound produced by the speaker should be within audible range. Generally, sound-absorbing boards are used to reduce the loudness of sound, while reflecting boards are used to increase the loudness.

4.10.3 Focusing due to Walls and Ceilings

In an acoustically fit auditorium or hall, sound waves are distributed throughout the hall uniformly. However, sound produced by the speaker undergoes multiple reflections at the ceilings and walls. The reflected sound from the ceiling and walls should not be focused at a particular point, rather it has to be disturbed throughout a hall. Generally, a plane surface reflects the sound uniformly throughout a hall while a curved surface does not distribute it evenly. Therefore, the reflection from a curved surface produces a harmful effect. Hence, the distribution of sound from a curved surface must be tackled carefully.



C - Center of curved surface
R - Reflected sound
S - Speaker

Fig. 4.9 *Focusing of sound—curved surface*

The reflection of sound from a curved surface is shown in Fig. 4.9. The reflected sound waves are focused at the front. The focusing point *R* of the reflected sound waves is much closer to the speaker and is away from the centre of the curved surface. Thus, the sound waves are not distributed uniformly due to the reflection from the curved surface.

In order to avoid these problems, the auditorium should be constructed by keeping in mind that the radius of curvature of a concave ceiling should be nearly twice the height of the building as shown in Fig.4.10. The distribution of sound from a concave surface is uniform. Therefore, to avoid bad distribution of sound waves from a ceiling, a careful ceiling design is required.

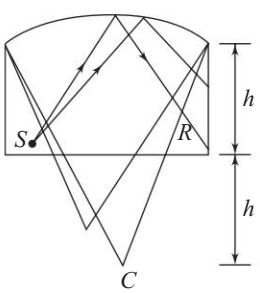


Fig. 4.10 *Focusing of sound—Concave surface*

4.10.4 Echelon Effect

Whenever people walk on a staircase or floors or hard paved paths, unwanted sounds are produced. These sound waves are produced due to poor finishing of

The McGraw·Hill Companies

the floor surface, structural effects, etc. The above unwanted sound is known as *echelon effect*. Consider a person with, steel-tipped shoes climbing up a staircase, as shown in Fig.4.11. Whenever the person makes a step, in staircase produces an unwanted sound.

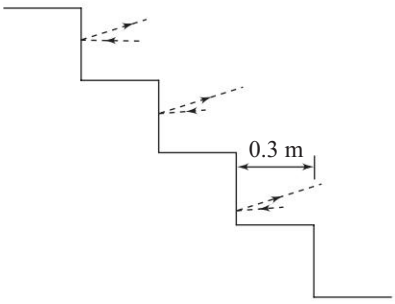


Fig. 4.11 *Echelon effect—taircase*

Consider that the width of the stairs is 0.3 m. Hence, the total distance traveled by the sound is 2 x 0.3 m. Therefore, the frequency of sound produced is,

$$f = \frac{v}{d} \tag{4.70}$$

We know that v is the velocity of sound at room temperature and is equal to 340 m s^{-1}.

Substituting the values of v and d in the above equation, we get,

$$f = \frac{340}{2 \times 0.3} \tag{4.71}$$

$$= 567 \text{ Hz}$$

Therefore, the person who is climbing up the staircase produces an unwanted sound at a frequency of nearly 600 Hz. These waves are known as echelon effect. The echelon effect is minimised by finishing the surface of the floor or stairs very finely and using a carpet to cover the floor and stairs.

4.10.5 Resonance Within the Building

We know that for the propagation of sound waves, a medium is essential. During the propagation of sound waves in a hall, air particles in the medium vibrate at their resonance frequencies. When the frequency of vibration of air particles matches with the hall's natural frequency of vibration, resonance is obtained. At this resonance condition, sound waves get amplified. Thus, it results in an unwanted sound effect inside a hall. Therefore, the resonance within the building has to be minimised.

The same can be achieved by studying the resonance effect on a model auditorium or hall. This model hall or an auditorium is kept inside a vessel which contains water. The water-wave particle movements are studied and are used for the construction of the actual hall or auditorium.

4.11 PRINCIPLES TO BE OBSERVED IN THE ACOUSTICAL DESIGN OF AN AUDITORIUM

The design of an auditorium requires a smooth growth of sound energy. In an auditorium, sound produced by sources must be distinguished and distributed clearly over the whole area of the auditorium. In order to ensure the above important features, factors which influence the acoustical design such as volume, shape and sound absorption are to be considered effectively while designing a hall.

4.11.1 Volume

The requirement of volume of a hall/auditorium is based on its purpose or use. For example, a music hall enquires a large volume than a speech /theatre hall. Similarly, an auditorium requires a large volume to accommodate more strength. The important factors to be considered while designing the volume of a hall/auditorium is the uniform distribution of sound without any compromise in the quality of sound waves. The height of a hall plays a dominant role in deciding the volume of the hall than the width or breadth of the hall. The thumb rule for the ratio of ceiling height and width of a hall is 2:3. The general guideline used for designing the volume of a hall/auditorium are given in Table 4.4.

Table 4.4 *Surface area required per person for construction*

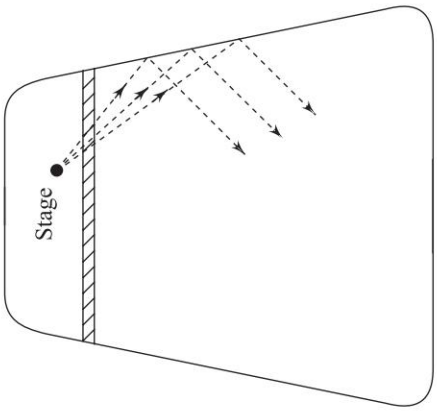
S.No.	*Requirement*	*Area required per person (m^3)*
1	Cinema theatre	3.74 to 4.2
2	Public cultural hall	2.8 to 3.7
3	Concert hall	4.2 to 5.6

An auditorium may be used as a music/concert hall or a speech hall. Hence, in an auditorium, the volume available for a person should be chosen so as to have a value intermediate between a speech hall and a music hall.

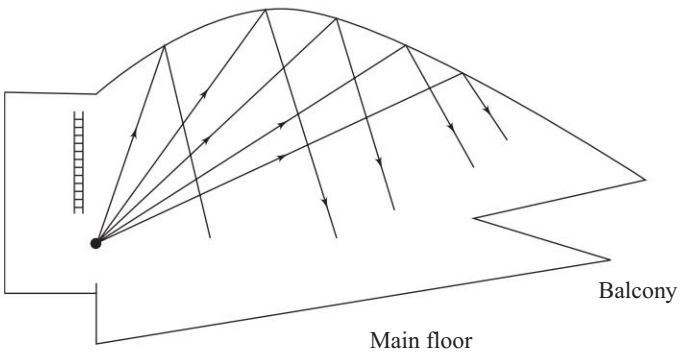
4.11.2 Shape

Similar to the volume of a hall, the shape of the walls and ceiling play an important role in optimising the distribution of sound waves. A concave surface reflects more amount of sound energy and focuses the reflected sound in a particular place. The concave surface and plain surface also distribute sound to the entire hall. The shape of walls and ceilings of a hall controls the echo and reflection of sound waves. When sound energy from the source is incident on a concave surface, it reflects more amount of sound energy and focuses the same on a particular place. Similarly, the plain surface distributes sound to the entire hall. The multiple reflections from a concave ceiling are reduced by keeping the radius of curvature of a concave surface as twice the height of a hall.

The McGraw·Hill Companies



(a) *Fan shapped*



(b) *Dome shaped*

Fig. 4.12 *Shape of wall ceiling*

The reflections of sound waves from the side walls of fan-shaped and dome-shaped floors are shown in Fig. 4.12.

4.11.3 Sound Absorption

The reverberation time of a hall/auditorium is optimised based on the applications. For example, the optimum value of reverberation time for a lecture hall is 0.5 second while for a music hall, the value lies between 1 and 2 seconds. The sound-absorbing materials are introduced into a hall *at* proper places, and thus it helps to obtain the required reverberation time. When sound-absorbing materials are placed at the right places in a hall, they absorb sound energy. The important factors to be considered to optimise reverberation time are the sound absorbing coefficient of the audience, arrangements of seats in a hall, and various sound-absorbing materials used inside the hall.

The seats are arranged in such a way that the passage of direct sound is not affected by the heights of heads of the audience in the other row. A typical arrangement of seats is shown in Fig.4.13.

The McGraw·Hill Companies

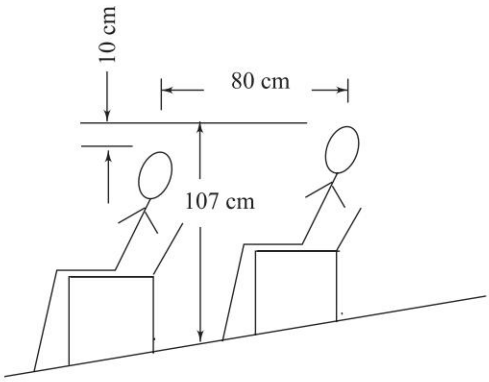


Fig. 4.13 *Auditorium—Seating arrangement*

4.12 ACOUSTICAL ANALYSIS AND ITS CORRECTION

Sabine's reverberation formula is used to verify the acoustic fitness of a given room or hall. In order to verify it, the absorption coefficient and surface area of each and every sound-absorbing materials present in a hall are measured. Let a_1, a_2, a_3 etc., and s_1, s_2, s_3 etc., be respectively the absorption coefficient and surface area of sound-absorbing materials present in a hall.

Therefore, the total absorption

$$A = a_1 s + a_2 s_2 + a_3 s_3 + \ldots\ldots \tag{4.72}$$

Equation (4.72), gives the total absorption coefficient of a hall of volume V. Thus, by knowing total absorption coefficient A and volume of hall V, reverberation time is determined using standard relation given in Eq.(4.70).

4.13 SOUND INSULATION

Sound insulation is the prevention of transmission of sound energy into a hall. Generally, there are two types of external noise which transits into a hall and disturb sound quality inside a hall. They are *airborne sound* and *structural bone sound*. In the following heading, origin, transmission and prevention of external sounds are discussed in brief:

4.13.1 Air-borne Sound

The sound waves which are originated from outside sources are propagated into a hall through air by open windows, doors and ventilators and are known as air-borne sound. For example, sound produced by a vehicle moving outside a hall is air-borne sound. The air borne sound is minimised employing the following ways:

The McGraw·Hill Companies

(i) One can use double doors and windows with separate frames by placing a sound absorbing material in between them.
(ii) One can use the paper for doors and windows.
(iii) One can close the windows and ventilator doors. However, it creates loss in transmitted light. The same can be overcome by using alternate ventilation.
(iv) We can make the hall air-conditioned.

4.13.2 Structure-borne Sound (Impact Sound Insulation)

Sound waves which are generated inside a hall are known as structure borne sound. These sound waves are propagated through the walls and floors. These sound waves are produced due to apparent motion such as of benches and footsteps. The following are reasons for the production of structure-borne sound inside a hall due to the movent of benches:

(i) It transfers vibration into walls and hence, the vibration on the walls creates sound waves on the other side.
(ii) It happens during the elastic wave motion in which compression and rarefaction of sound energy are transmitted into a hall from particle to particle by alternate compression and rarefactions. The same process is repeated to transfer the energy from the air on the other side.

The effect on structure-borne type of sound waves are minimised employing the following ways:

(i) We can use rigid structures so as to rest the vibrations.
(ii) We can introduce discontinuities in the path of sound. The discontinuities are provided in the form of construction joints in walls to reduce the passage of sound.
(iii) We can convert the floor and ceiling with suitable sound absorbing materials or anti-vibration mounts.

For example, a sound produced by water flowing in a pipe is avoided by using rubber or insulating gasket at the junctions.

4.14 NOISE POLLUTION

The unwanted sound waves which are dumped into the atmosphere are known as noise pollution. These sound waves produce or unpleasant effect, i.e., they are a nuisance and also an adverse effect to human health. Noise pollution leads to creation of either physical or mental problems. The persons working in the surroundings are liable to make errors. The series of problems expected are reduction in working efficiency, personal comfort, industrial accidents etc. Noise pollution is of various types, namely, domestic noise, trafic noise, aircraft noise, etc. These noises are produced by different sources. However, the types of sources are classified into two types, namely, industrial noises and non-industrial noises.

The *industrial noises* are produced due to the movements of cars, motor tracks, motor cycles, aircraft, machines, explosions, etc. The *non-industrial noises* are produced by loudspeakers, radios, construction work, etc. The sound pressure level, in dB, of industrial

and non-industrial noise is measured by relative measurements. The source for the domestic and industrial noises along with the sound pressure levels is given in Table 4.5.

Table 4.5 *Sound pressure level for domestic and industrial noise*

Source of noise	*Sound pressure Level dB*	*Source of noise*	*Sound pressure Level dB*
Threshold of audibility	0	Electric generator	76
Garden	20	Garbage collector	85
Bedrooms	25	Air compressor	85
Libraries	30	Concrete mixer	85
Living rooms	40	Jack hammer	88
Conversational speech	60	Scrapper	88
Vacuum cleaner	70	Dumb truck	88
Business office	67-70	Power saw	90
Ringing alarm clock	80	Rock drill	98
Average street traffic	80-85	Noisy factory	100
Heavy traffic	90	Pile driver	101
Pop music	110	Riveter	115
Threshold of pain	140	–	–

4.15 NOISE CONTROL IN MACHINES

In most industries, noise pollution is controlled to an optimium level. The uncontrolled noise in industry leads to a variety of problems. When noise level exceeds threshold of pain (120 dB), it leads to problems such as permanent or temporary deafness or industrial guidant's, discomfort, reduces the workers efficiency, etc. Therefore, necessary steps are taken to design the machines and workplace for quieter operations. The following are guidelines and procedures to be adopted in controlling noise pollution in machines and workplaces for quieter operations.

(i) By redesigning the source of sound for quieter operation
(ii) By insulating the source with sound-reducing houses
(iii) By enclosing high impacting parts of machines
(iv) By using high sound-absorbing materials as enclosures.
(v) By selecting sound absorbing materials should based on environmental conditions
(vi) By providing a dynamic balance to the vibrating machines
(vii) By using fixed bed to fix the rotating or impacting machines
(viii) By providing connections such as electricity, gas, air, etc., employing vibration decouplers.
(ix) By using pressure release chambers for providing water connections
(x) By silecing processes like internal combustion in engines
(xi) By using large work area, which consists of good sound-absorbing materials, as an enclosure to keep the machine.
(xii) By reducing structure-borne noise using double doors or walls with air spaces between them

(xiii) By providing sound reducing ear muffs and plugs to all workers
(xiv) By isolating the machine area from the office, show rooms etc.
(xv) By providing discontinuities in the construction of the building, the noise pollution due to vibrations of the machines is avoided

The other source of noise pollution in machines is the vibration of machines. The impact noise produced by a machine under operating conditions will be transferred to the building in the form of vibrations. These vibrations run throughout the building and hence, this problem will be extended to the remote parts of the building through the walls. The same can be reduced simply by providing discontinuities in the construction of the building.

Keypoints to Remember

- Sound waves are classified into three categories, namely infrasound (<20 kHz), audible sound (20 Hz–20 kHz) and ultrasound (>20 kHz)
- Musical sound waves produce a pleasant effect to our ears.
- Noise produces displeasure to our ears.
- Frequency, intensity and quality are the characteristics of musical sounds.
- The standard intensity level is known as threshold of audibility and is equal to $I_o = 10^{-16}$ W m^{-2}.
- The reverberation time is the time taken by a sound wave to fall from an average value to an inaudibility limit.
- The absorption coefficient of a material is defined as the inverse of the area of sound absorbing material which absorbs the same amount of energy as that of 1 m^2 of open window unit.
- Echo is the reflection and returning of sound waves from a distant extended surface like a cliff, hill, building, etc., with sufficient magnitude and delay.
- The sabine formula for reverberation time is $\frac{0.163V}{A}$ where V is the volume of the hall and A the total absorption coefficient.
- Growth of energy in a hall is equal to $E = E_m(1-e^{-\alpha t})$
- Decay of sound in a hall is equal to $E = E_m e^{-\alpha t}$
- Coefficient absorption of material is $a_1 = \frac{0.163V}{s_1}\left[\frac{1}{T_2} - \frac{1}{T_1}\right]$
- Important factors to be considered for acoustical design are volume, shape, and sound absorption.
- The total absorption coefficient is equal to $A = a_1 s_1 + a_2 s_2 + a_1 s_2 + \ldots$ where a_1, a_2, a_3 and, s_1, s_2, s_3 etc., are respectively the absorption coefficient and surface area of the sound-absorbing materials, 1, 2, 3, etc.

- The sound waves which are produced inside the hall are known as structure–borne sound waves.
- The sound waves which are produced outside and propagate into the hall are known as air-borne sound waves.
- There are two different types of noise pollution, namely, industrial noise and non-industrial noise.

Solved Problems

Example 4.1 The intensity of sound produced by thunder is 0.1 W m^{-2}. Calculate the intensity level in decibels.

Given Data

The intensity of sound produced by thunder $I_1 = 0.1$ W m^{-2}

Solution

The intensity level or relative intensity $\beta = 10 \log_{10} \left[\frac{I_1}{I_0}\right]$

$$= 10 \log_{10} \left[\frac{0.1}{10^{-12}}\right]$$

$$= 10 \log_{10} (10^{11}) = 10 \text{ x } 11 = 110 \text{ dB}$$

The sound intensity produced by thunder is 110 dB.

Example 4.2 Calculate the intensity level of a turbine whose sound intensity is 100 W m^{-2} when it is under operation. Given that the standard intensity level is 10^{-12} W m^{-2}.

Given data

The Intensity of sound of the turbine $I_1 = 100$ W m^{-2}

The standard intensity $I_1 = 10^{-12}$ W m^{-2}

Solution

We know that the intensity level or relative intensity $\beta = 10 \log_{10} \left[\frac{I_I}{I_0}\right]$

Substituting the values of I_0 and I_1, we get,

$$\beta = 10 \log_{10} \left[\frac{100}{10^{-12}}\right]$$

(or) $$\beta = 10 \log_{10} 10^{14}$$

Therefore, $$\beta = 10 \times 14 = 140 \text{ dB}$$

The intensity level or relative intensity of a turbine is 140 dB.

Example 4.3 Calculate the increase in the acoustic intensity level when the sound intensity is doubled.

Given Data

The intensity ratio $\left[\frac{I_1}{I_0}\right] = 2$

Solution

The relative intensity $\beta = 10 \log_{10} \left[\frac{I_1}{I_0}\right]$

$= 10 \log_{10} (2) = 3.0102999$ dB

The increase in sound intensity level is 3.01 dB.

Example: 4.4 A loudspeaker emits energy in all directions at the rate of 1.5 J s^{-1}. What is the intensity level at a distance of 20 m? Standard intensity level of sound = 10^{-12} W m^{-2}.

Given Data

The power radiated by the loudspeaker $Q = 1.5$ J s^{-1}

Radius $r = 20$ m

Standard intensity level $I_0 = 10^{-12}$ W m^{-2}.

Solution

The intensity of the sound produced by the loudspeaker

$$I_2 = \frac{\text{Power radiated}}{\text{area}}$$

$$= \frac{1.5}{4 \times 3.14 \times 20^2} = 0.0003$$

The intensity level $\beta = 10 \log_{10} \left[\frac{I_1}{I_0}\right]$

$$= 10 \log_{10} \left[\frac{0.0003}{10^{-12}}\right]$$

$= 84.7$ dB

The intensity level of the sound produced by the loudspeaker is 84.7 db.

Example 4.5 An electric generator operates at a sound intensity level of 80 dB. If it is operated in a room with an existing sound intensity level of 70 dB, what will be the resultant intensity level?

Given Data

Intensity level of the sound produced by the electric generator $\beta = 80$ dB

Intensity level of the room = 70 dB

The McGraw·Hill Companies

Solution

(i) Intensity level of the sound produced by electric generator = 80 dB

i.e., The relative intensity $\beta = 10 \log_{10}\left[\frac{I_1}{I_0}\right] = 80$ dB

$$80 = 10 \log\left[\frac{I_1}{10^{-12}}\right]$$

where I_1 is the standard sound intensity level = 10^{-12} W m^{-2}

Therefore, $I_1 = 10^{-4}$ W m^{-2}

The intensity of the sound produced by the electric generator

$$= 10^{-4} \text{ W m}^{-2}$$

(ii) Intensity level of the room $\beta_2 = 70$ dB

i.e., $$\beta_2 = 10 \log_{10}\left[\frac{I_2}{I_0}\right] = 70 \text{dB}$$

The intensity of the sound existing in the room

$$I_2 = I_0 \log^{-1}\left[\frac{\beta}{10}\right]$$

Here, I_0 is the standard sound intensity

Therefore, $= 10^{-12}$ W m^{-2}

$$I_2 = I_0 \log^{-1}\left[\frac{I_1}{I_0}\right]$$

$$= 10^{-5} \text{ W m}^{-2}$$

The intensity of sound existing in the room

$$I_2 = 10^{-5} \text{ W m}^{-2}$$

The total sound intensity when the generator is operating

$$= I_1 + I_2$$

$$= 10^{-4} + 10^{-5} \ W\,m^{-2}$$

$$= 1.1 \times 10^{-4} \ W\,m^{-2}$$

The relative intensity $\beta = 10 \log_{10}\left[\frac{I_1}{I_0}\right]$

$$= 10 \log_{10}\left[\frac{1.1\times10^{-4}}{10^{-12}}\right]$$

$$= 80.4139268$$

The resultant intensity level of the sound is 80.414 dB.

The McGraw·Hill Companies

Example 4.6 Calculate the reverberation time of hall with volume of 1500 m^3 and total absorption is equivalent to 100 m^2 sabine.

Given Data

Volume of the hall $V = 1500\ m^3$

The absorption of sound by the hall $A_1 = 80\ m^2$ O W U or sabine

Solution

The reverberation time of the hall

$$= \frac{0.163V}{\sum as} = \frac{0.163 \times 1500}{80} = 3\ s$$

The reverberation time of the hall is 3.0 s.

Example 4.7 A hall has a volume of 1, 20,000 m^3. It has a reverberation time of 1.5 seconds. What is the average absorbing power of the surface if the total absorbing surface area is 25,000 m^2?

Given data

Volume of hall $V = 1,\ 20{,}000\ m^3$

The Reverberation time $T = 1.5$ s

Total absorbing surface area $\sum s = 25{,}000\ m^2$

Solution

We know that,

the reverberation time $T = \dfrac{0.163\,V}{\sum a\,s}$

Rearrange the above equation, we get,

total absorption $\sum as = \dfrac{0.163\,V}{T}$

Substituting the values of the V and T, we get,

$$\sum as = \frac{0.163 \times 1{,}20{,}000}{1.5}$$

$$\sum as = 13{,}360 \text{ O W U}$$

The total absorption of the hall is 13,360 open window units (O W U)

Therefore, the average absorbing power of the surface $\bar{a} = \dfrac{\sum as}{\sum s}$

$$\bar{a} = \frac{13{,}360}{25{,}000}$$

$$\bar{a} = 0.5344 \text{ sabine}$$

The average absorbing power of the surface is 0.5344 sabine.

The McGraw·Hill Companies

Example 4.8 Calculate the intensity level of a jet plane just leaving the runway having a sound intensity of about 100 W m^{-2}.

Given Data

The sound intensity $I_2 = 100$ W m^{-2}

Solution

The relative intensity $\beta = 10 \log_{10} \left[\frac{I_2}{I_0}\right]$

$$\beta = 10 \log_{10} \left[\frac{100}{10^{-12}}\right] = 140 dB$$

The sound intensity level of a jet plate is 140 dB.

Example 4.9 A hall has a volume of 12500 m^3 and a reverberation time of 1.5 s. If 200 cushioned chairs are additionally placed in the hall, what will be the new reverberation time of the hall? The absorption of each chair is 1 O W U.

Given data

Volume of the hall $V = 12550$ m^3

The reverberation time $T = 1.5$ s

The absorption of cushioned chair $\sum a_1 s_1 = 200$ O W U

Solution

The absorption of the hall before introducing the cushioned chair is

$$\sum as = \frac{0.163V}{T}$$

Substituting the value of V and T in above equation, we get,

$$\sum as = \frac{0.163 \times 12500}{1.5}$$

$$= 1391.66 \text{ O W U}$$

The absorption coefficient of the hall is 1391.66 O W U.

We know that, the reverberation time of the hall after introducing the cushioned chair is

$$T_2 = \frac{0.163V}{\sum as + \sum a_1 s_1}$$

Substituting the value V, $\sum as$, and $\sum a_1 s_1$, we get,

$$T_2 = \frac{0.163 \times 12500}{(1391.66 + 200)}$$

$$T_2 = 1.311 \text{ s}$$

The new reverberation time of the hall after adding the cushioned chairs is 1.311 s.

Example 4.10 The volume of an auditorium is 12000 m^3. Its reverberation time is 1.5 second. If the average absorption coefficient of interior surfaces is 0.4 sabine m^{-2}, find the area of interior surfaces.

Given data

Volume of the auditorium $V = 12000\ m^3$

The reverberation time $T = 1.5$ S

The average absorption coefficient of interior surface $\bar{a} = 0.4$ sabine m^{-2}

Solution

We know that, the average surface area of the hall is

$$s = \frac{0.163V}{\bar{a}T}$$

Substitute the value of V, a and T, we get,

$$s = \frac{0.163 \times 12000}{0.4 \times 1.5}$$

$$s = 3340\ m^2$$

Therefore, the total surface area s is 3340 m^2 .

Example 4.11 The volume of a room is 1500 m^3. The wall area of the room is 260 m^2, the floor area is 140 m^2 and the ceiling area is 140 m^2. The sound–absorption coefficient for the wall is 0.03, for the ceiling is 0.8 and for the floor is 0.06. Calculate the average absorption coefficient and the reverberation time.

Given d ata

Volume of the hall $V = 1500\ m^3$

Surface	*Area or number (m^2)*	*Coefficient of absorption*
ceiling	140	0.8
walls	260	0.03
floor	140	0.06

Solution

We know that the reverberation time $T = \dfrac{0.163\,V}{a_1 s_1 + a_2 s_2 + a_3 s_3}$

Substituting the V, a and s values, we get,

$$= \frac{0.163 \times 1500}{(0.03\times260)+(0.8\times140)+(0.06\times140)}$$

$$= \frac{250.5}{7.8+112+8.4}$$

$$= \frac{250.5}{128.2}$$

$$T = 1.9539 \text{ s}$$

Therefore, the reveberation time T is 1.9539 s.

The average absorption coefficient is determined by two methods.

Method I

We know that the average absorption coefficient $\bar{a} = \frac{0.163V}{sT}$

where S is the total surface of the room and is given by sum of the surface area of the wall area, floor area and ceiling area.

Sudstiuting the values, we get,

$$= \frac{0.163\times1500}{(260+140+140)\times1.9539}$$

$$= \frac{250.5}{540 \times 1.9539}$$

$$= \frac{250.5}{1055.106}$$

$$\bar{a} = 0.2374 \text{ sabine}$$

Method II

We know that, the average absorption coefficient $\bar{a} = \frac{a_1s_1 + a_2s_2 + a_3s_3}{s_1 + s_2 + s_3}$

By substituting the values, we get,

$$\bar{a} = \frac{(0.03\times260)+(0.8\times140)+(0.06\times140)}{(260+140+140)}$$

$$\bar{a} = \frac{128.2}{540}$$

$$\bar{a} = 0.2374 \text{ sabine}$$

Therefore, the average absorption coefficient $\bar{a}$ is 0.2374 sabines

Example 4.12 A window whose area is 1.4 m^2 opens on a street where the street noise results in an intensity level at the window of 60 decibels. How much acoustic power enters via the window?

Given data

The area of the window $s = 1.4\ m^2$

The intensity level at the window $\beta = 60$ decibels.

The standard intensity of sound $I_o = 1 \times 10^{-12}\ W\ m^{-2}$

Solution

We know that the intensity level or relative intensity $\beta = 10 \log_{10} \left[\frac{I_I}{I_0}\right] dB$

Substituting the value of the I_0 and β, we get,

$$60 = 10 \log_{10} \left[\frac{I_I}{10^{-12}}\right]$$

Simplifying the above equation, we get,

The intensity of the sound I_1 $= 1 \times 10^{-6}\ W\ m^{-2}$

We know that Power = Intensity x Area

$= 1 \times 10^{-6}$ x 1.4

$= 1.4 \times 10^{-6}$

The acoustic power that entering via the window is 1.4 x 10^{-6} watts.

Example 4.13 A hall with a volume of 1000 m^3 has a sound–absorbing surface of area of 400 m^2. If the average absorption coefficient of the hall is 0.2 sabines, what is the reverberation time of the hall?

Given data

The volume of the hall $V = 1000\ m^3$

The surface area of the hall $s = 400\ m^2$

Average absorption coefficient of the hall $\bar{a} = 0.2$ Sabine

Solution

We know that the reverberation time $T = \frac{0.163}{\sum as}$

Substituting the values of the V, ā and s, we get,

$$T = \frac{0.167 \times 1000}{80} = 2.0875\ S$$

Reverberation time of the hall = 2.0875 s.

The McGraw·Hill Companies

Exercises

4.1 Two sound waves of same frequency have intensities 10^{-16} and 10^{-12} W cm^{-2}. What is the difference between the intensity levels of these sounds?

4.2 Calculate the acoustic intensity of a plane acoustic wave of intensity 10^{-15} and 10^{-2} with reference to a standard acoustic intensity of 10^{-12} W m^{-2}.

4.3 The intensity level of a sound is increased by 10 dB. If the original intensity was 10^{-2} W m^{-2}, what is the new intensity?

4.4 Two sounds have intensities of 2×10^{-10} and 2×10^{-7} W m^{-2} respectively. How many decibels is one sound above the other?

4.5 The volume of a hall is 2,000 m^3 and its total absorption equals 100 m^2 of open window. Entry of people in the hall raises the absorption by 2,000 m^2. Determine the change in the reverberation time.

4.6 The volume of a room is 800 m^3. The wall area of the room is 300 m^2, the floor area is 150 m^2 and the ceiling area is 150 m^2. The average sound absorption coefficient (i) for the wall is 0.03, (ii) for the ceiling is, 0.8 and (iii) for the floor is 0.06. Calculate the average sound absorption–coefficient and the reverberation time.

4.7 A lecture hall with a volume of 1,215 m^3 has a reverberation time of 1.5 seconds. What is the total absorption power of all the surfaces in the hall? If the area of the sound absorbing surface is 720 m^2, calculate the average absorption coefficient.

4.8 A hall with a volume of 85,000 cu.ft. is found to have a reverberation time of 2.2 seconds. If the area of the sound–absorbing surface is 7,500 sq.ft., calculate the average sound absorption coefficient.

4.9 The power of a sound from a speaker of a radio receiver is of 25 mW. By tuning the volume control, the power of the sound is increased to 250 mW. Calculate the power increase in decibles.

4.10 A loudspeaker emits energy in all directions at the rate of 1.5 J s^{-1}. What is the intensity level at a distance of 20 m? Standard intensity level of sound = 10^{-12} Wm^{-2}.

Objectives Questions

4.1 _____ is the frequency of the audible sound waves.

4.2 Musical sound waves produce _____ effect to our ears.

4.3 _____ sound waves produce an unpleasant effect to our ears.

4.4 The standard intensity of sound waves is _____ W m^{-2}.

4.5 The value of 1 bel is equal to _____ decibel.

4.6 The audible intensity level is _____ to _____ dB.

4.7 If loudness exceeds _____ dB, it produces a pain in the ear.
4.8 According to Sabine, the reverberation time T is equal to _____ second.
4.9 The coefficient of absorption of an average–built human is _____ sabine.
4.10 The sound waves _____ from a distant extended surface are known as _____.
4.11 Equation for the growth of energy is _____
a) $E = E_m (1-e^{-\alpha}t)$ b) $E_m = E_m (1+ e^{-\alpha t})$
c) $E = E_m (1-e^{\alpha}t)$ d) $E = E_m (1 + e^{-\alpha t})$
4.12 The decay of energy in a hall is E = _____
4.13 The volume required for concert hall is _____ per person.
4.14 The total absorption coefficient equal to _____
4.15 _____ sound waves are propagated into the hall through air.
4.16 _____ sound waves are produced inside the hall.
4.17 The classifications of noise pollution are _____ and _____
4.18 _____ dB is the sound level for the threshold of pain.

KEY

4.1 20 Hz –20 kHz	4.2 Pleasing	4.3 Noise
4.4 10 – 16	4.5 10	4.6 0 to 120
4.7 120	4.8 (0.163 V)/A	4.9 0.45
4.10 Reflected, echo	4.11 a	4.12 $E_m e^{-\alpha t}$
4.13 4.2 t^3	4.14 $a_1s_1 + a_2s_2 + a_3s_3 + \ldots$	4.15 air–borne
4.16 Structure–borne	4.17 industrial noise and non - industrial noise	
4.18 140		

Short Answer Questions

4.1 How are sound waves classified?
4.2 Differentiate the music and noise sound waves?
4.3 What is meant by pitch?
4.4 Distinguish between the intensity and loudness of the sound?
4.5 How the intensity of sound is measured? Explain.
4.6 What is meant by loudness?
4.7 State Weber–Fechner law.
4.8 What is meant by threshold of audibility?
4.9 Define the term bel.
4.10 What is the difference between bel and decibel?
4.11 What is meant by threshold of feeling?
4.12 Define the term phon.
4.13 What is meant by timbre?
4.14 What is meant by transmission loss?
4.15 What is meant by acoustics of building?
4.16 Define the term ‘reverberation time’.

4.17 What are the factors affecting the acoustics quality of the building?
4.18 Define absorption coefficient. State its unit.
4.19 What is meant by noise reduction?
4.20 What is meant by echo?
4.21 Explain the term creep?
4.22 What is meant by standing waves?
4.23 Write Sabine's mathematical relation for reverberation time and explain the various terms.
4.24 Explain the experimental determination of the absorption coefficient of a material.
4.25 What are sound-absorbing materials?
4.26 Give a few examples for sound–absorbing materials.
4.27 What is meant by sound insulation?
4.28 What is meant by air-borne sound?
4.29 Define structure-borne sound.
4.30 Distinguish between air-borne sound and structure-borne sound.
4.31 What is meant by noise pollution?
4.32 What are different types of noise pollutions?
4.33 Explain noise control in machine.
4.34 Mention some of the points to be followed in controlling noise pollution in machines.

Descriptive Questions

4.1 Derive the expression for growth and decay of the energy density inside a hall and hence deduce Sabine's mathematical relation for reverberation time of the hall.
4.2 Explain the various factors affecting the architectural acoustics of a building and their remedy.
4.3 Define the term absorption coefficient. Explain in detail the experimental methods used to measure the absorption coefficient.
4.4 Explain in detail the phenomena of transmission sound. What is meant by transmission loss? How is the transmission loss experimentally measured?
4.5 Describe the sound insulation and sound–absorbing materials.
4.6 Describe noise pollution and noise control in machines.

Chapter

Ultrasonics

5

OBJECTIVES

- To explore the knowledge on high frequency sound waves and their applications.
- To study the different methods of production and detection of ultrasonic waves.
- To study the characteristic properties of ultrasonic waves.
- To study the applications of ultrasonic in industry, medicine and other areas.

5.1 INTRODUCTION

In Chapter 1, we have studied the effects of reflection of sound in architectural acoustics. In this chapter, we will study the generation, detection and application of the ultrasonic waves.

Sound waves whose frequencies lie above the audible frequency of 20,000 Hz are called *ultrasonic waves*. Their wavelengths are small and hence, exhibit some unique phenomena in addition to the properties of audible sound waves. The sound waves whose frequencies lie below the audible limit are called *infrasonic waves*, whereas supersonic waves refer to the sound waves whose frequencies are higher than the audible waves.

Let us first discuss the classification of ultrasonic waves.

5.2 CLASSIFICATION OF ULTRASONIC WAVES

Ultrasonic waves are classified into three categories based on the mode of vibration of the particle in the medium, with respect to the direction of propagation of the initial waves. They are as given below:

(i) Longitudinal waves
(ii) Transverse or Shear waves
(iii) Surface or Rayleigh waves, and
(iv) Plate or Lamb waves

Let us discuss the different types of ultrasonic waves in detail.

5.2.1 Longitudinal or Compressional Waves

The nature of transmission of longitudinal ultrasonic waves is similar to longitudinal sound waves. In this mode, the particle motion is parallel to the direction of propagation of the incident wave. Due to the vibrations of the particles, alternate compression and rarefaction zones are produced, as shown in Fig. 5.1.

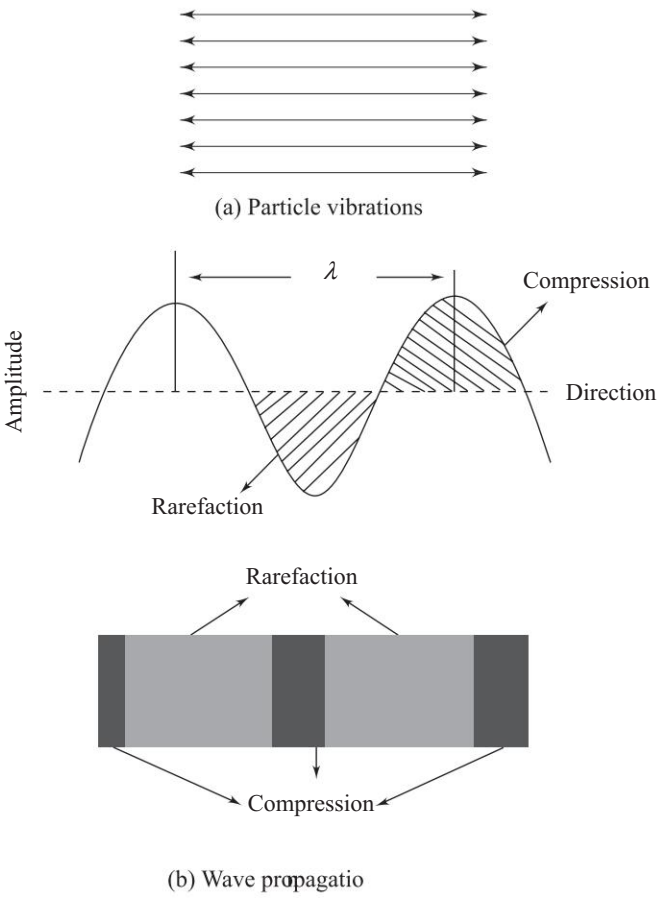


Fig. 5.1 *Longitudinal ultrasonic waves*

The resulting wave is called *longitudinal ultrasonic wave* or *compressional wave*. During the particle motion in the material, compressive and dilative forces are acting and, therefore, these waves are known as *dilational* or *pressure waves*. Longitudinal waves can be propagated in solids, liquids and gases. They are easy to generate, detect and also to convert into other modes of vibrations.

5.2.2 Transverse or Shear Waves

This is another form of propagation of ultrasonic waves and the same get transmitted as shear sound waves. In this type of wave, the vibration of the particle is at right

The McGraw·Hill Companies

angles to the direction of propagation of the wave motion. Such waves are called *transverse waves* or *shear waves*. The schematic representation of a shear wave is shown in Fig. 5.2.

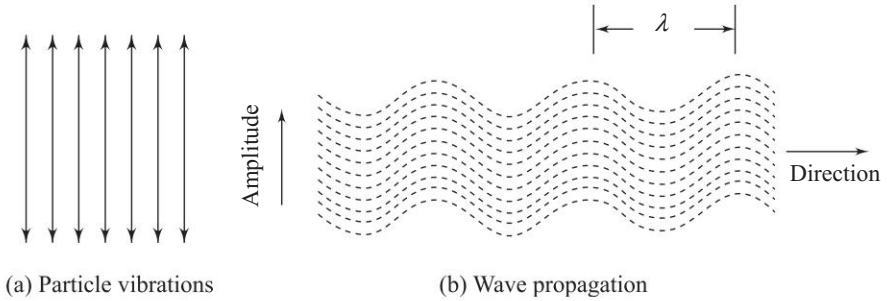


Fig. 5.2 *Shear ultrasonic waves*

The velocity of the shear waves is half that of the longitudinal waves in the same material. When a wave travels through a material, it makes the particles of the material vibrate due to the transfer of energy. The vibrating particles make the neighbouring particles vibrate. This results in reduction of energy of the ultrasonic waves and hence, velocity of the shear wave is half the longitudinal waves. Shear waves can be passed only through solids. They cannot be generated in liquids or gases.

5.2.3 Surface or Rayleigh Waves

Lord Rayleigh, in 1885, demonstrated that waves can be propagated over the plane boundary between an elastic half space and a vacuum or sufficiently rarefied medium (for example, air). The amplitude of the waves decays rapidly with the depth of propagation of the waves in the medium. These waves are called *surface waves* or *Rayleigh waves*. Surface waves are analogous to water waves. In water waves, the motion of the particle is both transverse and longitudinal in a plane containing the direction of propagation and is also normal to the surface. On the other hand, in Rayleigh waves, the particle motion is elliptical as shown in Fig. 5.3.

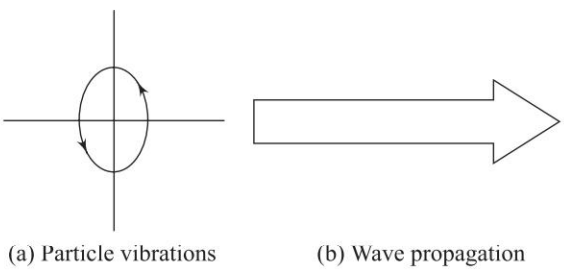


Fig. 5.3 *Surface waves*

Rayleigh waves can travel only on the surface layer of solids because during the existence of Rayleigh waves, there is no particle motion at greater depth. The particle in the medium can vibrate up to one wavelength from the surface of the

medium. The velocity of these waves in the same medium is about 90% of the velocity of longitudinal waves.

When a solid material is immersed in a liquid, surface waves can exist only when the interfacial thickness is very small, i.e., the liquid covers the solid surface in a very thin layer. Under this condition, some of the energy is transferred back into the liquid medium. These waves are called *stonely waves*.

5.2.4 Plate Waves or Lamb Waves

When a surface wave is introduced into a material having thickness equal to three times the wavelength or less, a different kind of wave known as *plate wave* is produced. During the existence of a plate wave, the material begins to vibrate as a plate, i.e., the wave encompasses the entire thickness of the material. This theory was described by Horace Lamb in 1916 and hence, such waves also called *Lamb waves*.

The different modes of plate waves are (i) symmetrical or dilational and (ii) asymmetrical or bending. The existence of Lamb waves in different modes depends on the particle movement.

Symmetrical or Dilational Lamb Waves Here, the displacement of the particles is longitudinal along the neutral axis of the plate and also elliptical on each surface, as shown in Fig. 5.4.

Asymmetrical or Bending Lamb Waves In this mode, the displacement of the particles is shear along the neutral axis of the plate and elliptical on each surface, as show in Fig. 5.5.

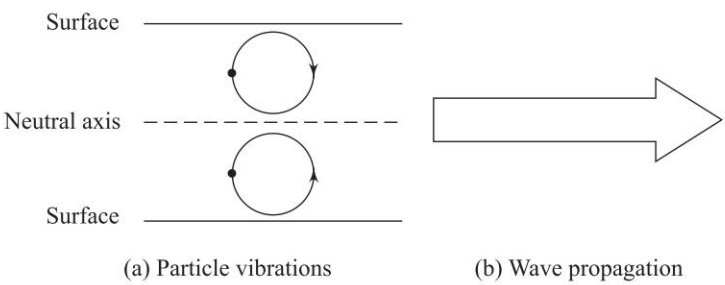


Fig. 5.4 *Symmetrical Lamb wave*

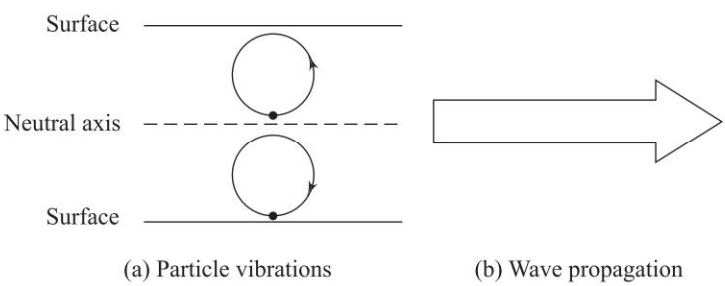


Fig. 5.5 *Asymmetrical Lamb waves*

The McGraw·Hill Companies

In addition to the above, plate waves are sometimes called *rod waves* if the material is circular in nature. Similarly, if the particles vibrate in a direction parallel to the surface, it is known as *Love wave*. Rayleigh, Lamb and Love waves find special applications in industry. The most important types of waves for scientific and industrial applications are longitudinal and shear waves.

Most of the industrial and scientific works are based on longitudinal and shear ultrasonic waves. Therefore, more emphasis is given to these two types of ultrasonic waves in this chapter. In the following sections, the generation and detection of ultrasonic waves are explained.

5.3 PROPERTIES OF ULTRASONIC WAVES

The various applications of ultrasonic waves are due to their following characteristic properties:

(i) The frequency of an ultrasonic wave is greater than 20 kHz.
(ii) It travels a longer distance in the medium without any loss.
(iii) It travels as a well-defined sonic beam.
(iv) It's velocity is constant for a homogeneous medium.
(v) It has many modes of vibrations such as longitudinal, shear and different modes of surfaces vibrations.
(vi) Direct interaction of the wave with the materials takes place. This plays a vital role in the study of molecular interactions, and materials characterisation.
(vii) At high frequencies, the wavelength is shorter and hence, it produces higher resolution in flaw detection.
(viii) It undergoes reflection and refraction at the interface, due to the change in elastic and physical properties of the medium.

5.4 GENERATION OF ULTRASONIC WAVES

In general, ultrasonic waves are produced by the methods given below:

(i) Mechanical generator
(ii) Magnetostriction, and
(iii) Piezoelectric effect

Let us discuss the generation of ultrasonic waves using the above three methods in detail.

5.4.1 Mechanical Generator

The mechanical generator, or Galton Whistle, is a miniature of the organ pipe. The experimental arrangement used in this generator is shown in Fig. 5.6.

It consists of two circular glass cylinders, which are arranged in such a way that the distance between the two ends *B* and *C* can be adjusted by using the two micrometre screws H_1 and H_2. The air blown at the *nozzle A* comes out at the circular end *B*. The outcoming air strikes the sharp tip *C*, which is in the form of a knife edge. By adjusting the distance between *B* and *C,* using the screws H_1 and H_2, the air column in the circular tube is set into vibrations. The frequency can be varied by varying the length of the vibrating column with the help of a piston arrangement.

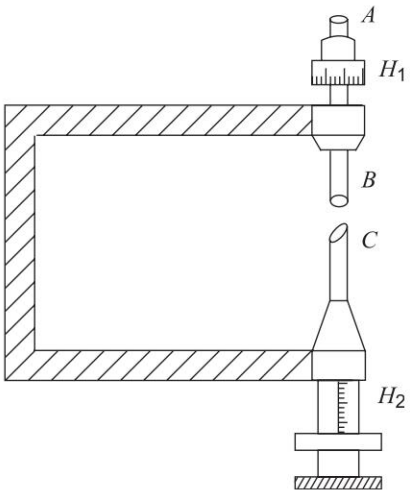


Fig. 5.6 *Mechanical generator*

Let l and x be the length of the vibrating air column and the end correction of the knife edge.

The wavelength of the sound waves is given by,

$$\lambda = 4\,(\,l + x\,) \tag{5.1}$$

The frequency of the ultrasonic waves produced using a Galton Whistle is given by,

$$f = \frac{u}{\lambda} \tag{5.2}$$

where u is the velocity of sound in air.
Hence, the frequency of the ultrasonc waves is

$$f = \frac{u}{4(l + x)}\ \text{Hz} \tag{5.3}$$

Thus, by varying l and x, ultrasonic waves can be generated at any desired frequency.

5.4.2 Magnetostriction Generator

Magnetostriction Effect When a ferromagnetic material in the form of a bar is subjected to an alternating magnetic field as shown in Fig. 5.7, the bar undergoes alternate contractions and expansions at a frequency equal to the frequency of the applied magnetic field. This phenomena is known as magnetostriction effect. As a result, the bar is thrown into longitudinal vibrations, thereby producing ultrasonic waves in the surrounding medium. Such ferromagnetic materials which are used for the production of ultrasonic waves are called *magnetostriction materials*.

The McGraw·Hill Companies

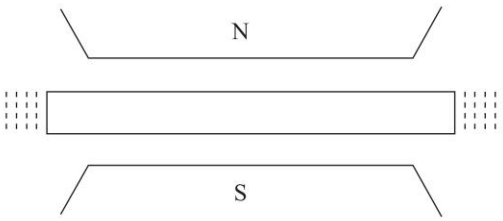


Fig. 5.7 *Magnetostriction effect in ferromagnetic material*

Magnetostriction Oscillator In this method, the ultrasonic waves are produced by using the magnetostriction effect. The experimental arrangement is shown in Fig. 5.8.

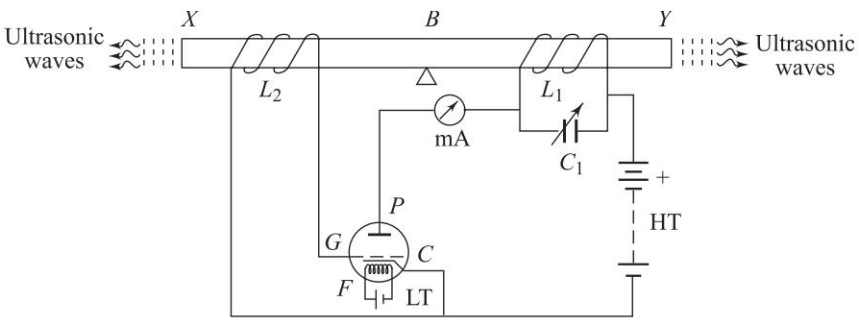


Fig. 5.8 *Magnetostriction oscillator*

The ferromagnetic bar *XY* is clamped at the middle *B*. The coils L_1 and L_2 are wound at the ends of the bar. The coil L_1 and the condenser C_1 are connected in parallel and the combination is connected between the anode and cathode circuit, through a milli-ammeter, while L_2 is connected between the grid and the cathode.

The bar gets magnetised by the plate current passing through the coil L_1. Any change in the plate current brings about a change in magnetisation and consequently the length of the rod. This gives rise to change in the flux in coil L_2 in the grid circuit, thereby inducing an emf in the coil L_2. This varying emf is applied to the triode valve and is fed back to the coil L_1, thereby maintaining the oscillations. By adjusting the condenser C_1, the frequency of oscillation of the tank circuit is varied. If the frequency of the tank circuit matches the natural frequency of the material due to the resonance, the rod vibrates vigorously and hence, produces ultrasonic waves. The milli-ammeter reading gives maximum value of the resonance conditions.

The frequency of ultrasonic waves produced by this method depends upon the length *l*, density *r* and Young's Modulus *Y* of the bar,

i.e.
$$f = \frac{1}{2l}\sqrt{\frac{Y}{\rho}} \tag{5.4}$$

Thus, by varying *l* and *Y* of the bar, ultrasonic waves can be generated at any desired frequency.

5.4.3 Piezoelectric Method

Piezoelectric Effect In this effect, when the opposite faces of a crystal C, like quartz, tourmaline, rochelle salts, etc., are subjected to squeezing (crush), twisting or bending, a potential difference is developed across the perpendicular opposite faces. The magnitude of potential difference developed across the crystal is proportional to the extent of deformation produced. This effect is known as direct *piezoelectric effect*. The converse of piezoelectric effect is also true. According to this effect, if an ac voltage is applied to one pair of faces of the crystals, alternatively mechanical contractions and expansions are produced and hence, the crystal starts vibrating. If the frequency of the applied ac voltage is equal to the vibrating frequency of the crystal, then the crystal will be thrown into resonant vibrations and hence produce ultrasonic waves.

Piezoelectric Crystal When the two opposite faces of a quartz crystal are cut perpendicular to the electric axis (x-axis), it is known as X-cut plate or X-cut crystal. Similarly, when the two opposite faces of a quartz crystal (Fig. 5.9a) a are cut perpendicular to the mechanical axis (y-axis), it is known as Y-cut plate or Y-cut crystal. Examples for X-cut and Y-cut crystals are shown respectivelyin Figs 5.9b and c.

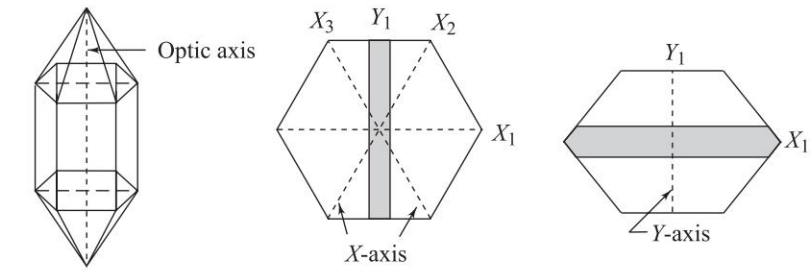


Fig. 5.9 *Piezoelectric crystal*

Piezoelectric Oscillator The experimental arrangement of the piezoelectric oscillator is shown in Fig. 5.10.

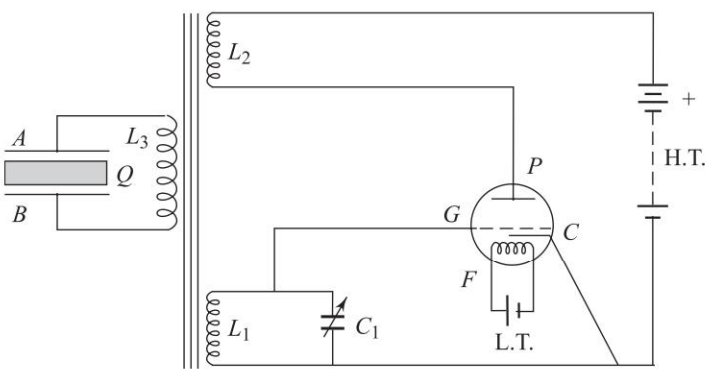


Fig. 5.10 *Piezoelectric oscillator*

The McGraw·Hill Companies

A thin slice of quartz crystal Q is placed in between two metal plates A and B. The plates A and B are connected to the coil L_3. The coils L_1, L_2 and L_3 are inductively coupled. The coil L_2 is connected in the plate circuit, while L_1 is connected between the grid and cathode by a tank capacitor C_1.

When the switch S is closed, the oscillator produces oscillations with a frequency

$$f = \frac{1}{2\pi\sqrt{L_1 C_1}} \tag{5.5}$$

The frequency of oscillation can be controlled by the variable capacitor C_1. Due to the transformer action, an emf is induced in the secondary coil L_3. This emf excites the quartz crystal into vibrations. By adjusting the capacitor C_1, the crystal is set into one of the modes of resonant condition. Thus, the vibrating crystal produces longitudinal ultrasonic waves in the surrounding air. The frequency of vibration is given as,

$$f = \frac{P}{2l}\sqrt{\frac{Y}{\rho}} \tag{5.6}$$

where Y is the Young's modulus, ρ the density of the material and P = 1, 2, 3, etc.

The velocity of longitudinal waves in the crystal is given by the relation,

$$v_l = \sqrt{\frac{Y}{\rho}} \tag{5.7}$$

Thus, using Eq. (5.6), the ultrasonic waves at any frequency can be produced.

5.5 ULTRASONIC VELOCITY MEASUREMENTS

Let us study some of the experimental methods used for the measurement of the velocity of propagation of ultrasonic waves through a medium.

5.5.1 Kundt's Tube

The Kundt's tube apparatus consists of a long circular glass tube of 1–m length and a diameter of 0.03 to 0.04 m. One end of the tube is fitted with an adjustable piston rod with a cork. The quartz crystal placed in between the two metal plates is placed at the mouth of the other end of the tube. Lycopodium powder is spread uniformly inside the tube. The arrangement of the apparatus is shown in Fig. 5.11.

As in the case of sound waves, the stationary supersonic waves are produced in the air contained in a long glass tube which is placed horizontally. The lycopodium powder gets collected in the form of heaps at the nodes and is blown off at the antinodes. By measuring the average distance between the adjacent heaps, the wavelength and hence, the ultrasonic velocity can be calculated using the following relations:

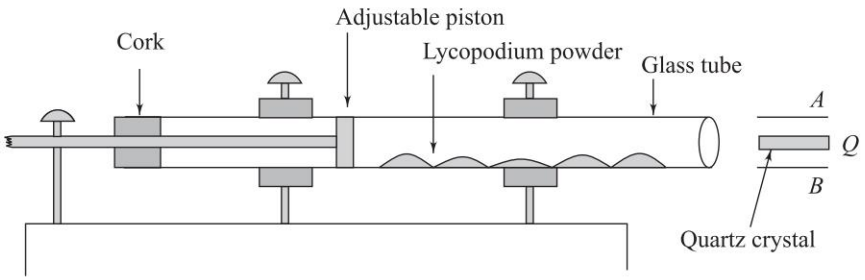


Fig. 5.11 *Kundt's tube*

The wavelength of the ultrasonic waves is given by $\lambda = 2d$
where d is the distance between the successive nodes.
The ultrasonic velocity in the medium is

$$u = f\lambda = 2fd \text{ m s}^{-1} \tag{5.8}$$

This method cannot be employed when the wavelength of the ultrasonic waves in the medium is very small.

5.5.2 A Coustical Diffraction

Principle The ultrasonic waves are produced by the quartz crystal Q which is placed in between the metal plates A and B. The audio–frequency signal is applied between the two plates. At the resonant conditions, the crystal generates ultrasonic waves due to the longitudinal compressions and expansions. These waves are passed through the liquid medium contained in a glass tube and are reflected back by the reflector R. These reflected waves are called *echos*. Thus, longitudinal stationary waves are produced in the liquid medium due to the superposition of the direct and reflected waves. These waves give rise to a fixed pattern of nodal and antinodal planes. At the nodal planes, the density and hence, the refractive index is maximum, while at the antinodal planes, the density and the refractive index are minimum. This arrangement behaves like a diffraction grating and hence the experiment is called acoustic grating experiment.

Working The experimental arrangement used for this study is shown in Fig. 5.12. A parallel beam of monochromatic light from the collimator is made incident normally on the acoustic grating. A diffraction pattern consisting of a central maxima of the direct beam and first–order maxima on either side is produced. This pattern can be viewed through the telescope.

Let d be the distance between the two adjacent nodal or antinodal planes.
The wavelength of the light is given by,

$$2d \sin \theta = n\lambda \tag{5.9}$$

where θ and n are the angles of diffraction for the n^{th} order and the order of the maxima. In addition, the wavelength (λ_0) of the ultrasonic waves can be calculated by using Eq. (5.10) as,

The McGraw·Hill Companies

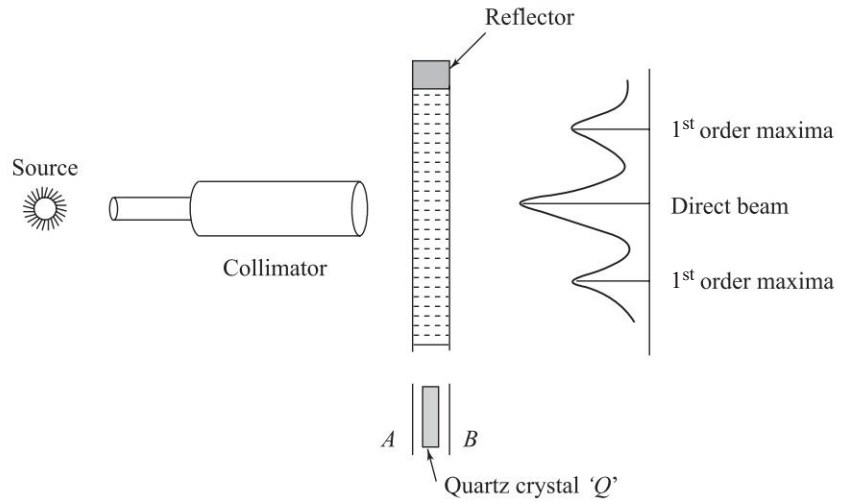


Fig. 5.12 *Arrangement of acoustic diffaction grating*

$$d = \frac{\lambda_0}{2}$$

or
$$\lambda_0 = 2d \tag{5.10}$$

Let f be the resonant frequency of the piezoelectric oscillation. Then, the velocity of ultrasonic waves,

$$u = f\lambda_0$$

$$= 2fd \tag{5.11}$$

This method is useful for the measurement of wavelength and velocity of ultrasonic waves through liquids and gases at various temperatures.

5.5.3 Thermal Detection

During the passage of ultrasonic waves through a medium, the waves undergo compression and rarefaction alternatively. Heat is developed as a result of compression, thereby increasing the temperature at those places and at the places where rarefaction takes place, the temperature decreases. If a platinum wire is placed in the region of ultrasonic waves, due to compression and rarefaction, the platinum wire is alternatively heated and cooled throughout its length. This change in temperature changes the resistance of the platinum wire. Thus, by measuring the resistance of the platinum wire, the ultrasonic waves can be detected.

5.5.4 Sensitive Flame

When ultrasonic waves are passed through a sensitive flame, the intensity of the flame gets changed due to the high frequency of the ultrasonic waves. By measuring the change in intensity of the flame, the ultrasonic waves can be detected.

5.5.5 Piezoelectric detection

The ultrasonic waves are detected by employing the principle of piezoelectric, electrostatic and magnetostriction effects. In piezoelectric method, the ultrasonic waves (longitudinal or shear), after passing through the material, are received by the quartz crystal. The quartz crystal may be *X*- or Y-cut depending on the type of ultrasonic waves, namely, longitudinal or shear. When the quartz crystal receives the ultrasonic waves at one face, it produces a potential difference at the opposite faces of the crystal. Thus, the quartz crystal produces an ultrasonic rf signal. The received signal is amplified by an amplifier circuit and compared with the built-in internal rf signals from the oscillator. By comparing the frequency of the ultrasonic rf signal with the internal oscillator, the frequency of the ultrasonic waves is determined. By knowing the frequency, one can also determine the transit time required for the propagation of the ultrasonic waves in the material.

5.6 ABSORPTION AND DISPERSION OF ULTRASONIC WAVES

When ultrasonic waves pass through the medium, the intensity of sound gets reduced due to the conversion of sound energy into heat. This energy loss in the medium is due to the alternative compression and rarefaction of sound waves. The compression produces heat, thereby increasing the temperature, while the rarefaction reduces the temperature. This leads to absorption of sound in the medium leading to sound attenuation.

If the ultrasonic waves pass through a medium, the measured intensity per unit time per unit area perpendicular to the direction of propagation decreases with distance from the source. This phenomenon is known as *attenuation*. Let I_0 be the intensity of the sound at unit distance from the source.

Then, the intensity at a distance x from the source is given by Eq. (5.12) as,

$$I = I_0 e^{-ax} \tag{5.12}$$

where a is called the coefficient of absorption and it is given by

$$\alpha = \frac{1}{x} \times \ln\left(\frac{I}{I_0}\right) \tag{5.13}$$

The absorption coefficient can be expressed in decibels per unit length or dB/ decibels.

The change in velocity of the ultrasonic waves with frequency is known as *dispersion* of ultrasonic waves. In viscous liquids like glycerine and castor oils, the change in velocity with frequency cannot be observed in the frequency region of ultrasonic waves. But, dispersion was observed indirectly by measuring the change in wavelength.

The McGraw·Hill Companies

5.7 SOURCE OF SOUND ABSORPTION AND DISPERSION

In a medium, the dissipation of energy depends on viscosity, heat conduction and heat radiation.

Viscosity When ultrasonic waves propagate through a fluid, there is relative motion between different parts of the medium, which is opposed by the viscous forces in the fluid. These viscous forces are opposed by the force due to incident ultrasonic waves. Therefore, a part of the energy of ultrasonic waves are converted into mechanical energy. This results in increase in temperature of the fluid and decrease in intensity of the ultrasonic waves. The viscosity of a liquid decreases as the temperature rises, whereas that of gases increases with temperature.

On account of the change in viscosity, the absorption coefficint is given by,

$$\alpha_v = \frac{16\pi^2 f^2}{3\rho_0 \eta c^3} \tag{5.14}$$

where η is the viscosity, f the frequency of the sound wave, c the velocity and r_0 the equilibrium density of the fluid. Equation (5.14) is valid only for low frequencies.

Heat conduction This is the second cause for the dissipation of sound energy. The compression associated with the propagation of sound wave in a fluid produces a local increase in temperature, while the temperature of the neighbouring layers which are in a state of rarefaction is decreased. Consequently, heat flows from the compressed layers to the rarefied layers, and thus a portion of energy is converted into heat.

The sound absorption due to heat conduction a_h is given by,

$$\alpha_h = \left(\frac{\gamma - 1}{\eta}\right)\left(\frac{4\pi^2 f^2 MK}{\rho_0 c^3 C_v}\right) \tag{5.15}$$

where $\gamma = C_p/C_v$ is the ratio of the molar specific heat at constant pressure C_p to that at constant volume C_v, M the mass per mole or the molecular weight, ρ_0 the equilibrium density and K the thermal conductivity. The sum of equations (5.14) and (5.15) gives the total absorption due to viscosity and heat conduction process.

Heat Radiation Whenever one portion of the fluid acquires a temperature different from that of an adjacent portion, thermal radiation occurs, thereby resulting in sound absorption. The attenuation due to viscosity and heat conduction is a function of frequency, whereas attenuation due to heat radiation is independent of frequency.

5.8 APPLICATIONS—INDUSTRY

The ultrasonic waves find extensive application in scientific research, industry, medical field, etc.

5.8.1 Sound Navigation and Ranging (SONAR)

In this method, high-frequency ultrasonic waves are used for finding the distance and direction of submarine, depth of sea, depth of rocks in the sea and the shoal of fish in the sea, etc.

The ultrasonic waves are generated by the piezoelectric method using a quartz crystal placed in between two metal plates. The same quartz crystal is used for generating and detecting the ultrasonic waves as shown in Fig. 5.13. These waves are transmitted towards the bed of the sea and get reflected back from the bed in the form of echos. These reflected echos are received by the quartz crystals and they are amplified and given to a cathode ray oscilloscope. By using the oscilloscope, the time taken by the ultrasonic waves for the to-and-fro distance is measured.

By knowing the velocity of sound in sea water and the measured time, the depth of the sea can be calculated using the relation,

$$\frac{h}{t/2}$$

or

$$\frac{ut}{2} \tag{5.16}$$

where t is the time taken by the waves for the to-and-fro motion and v the velocity of sound in sea water. The *echometer* is calibrated to measure the depth of the sea directly.

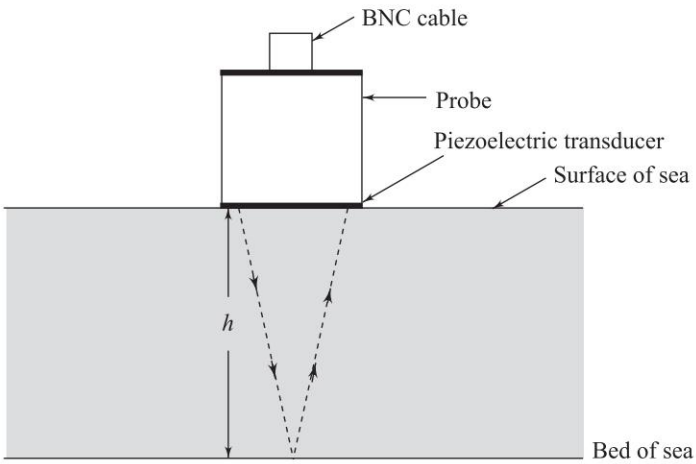


Fig. 5.13 *Determination of depth of sea*

5.8.2 Non-Destructive Testing (NDT)

Detection of Flaws in Metals Non-destructive testing is an experimental method used to study the quality of materials under operating conditions or in the manufacturing process using ultrasonic waves.

The McGraw·Hill Companies

The block diagram of the pulse echo method used to determine the various defects, (flaws) like cracks, holes, air bubbles, laminations, etc., in the specmen is shown in Fig. 5.14.

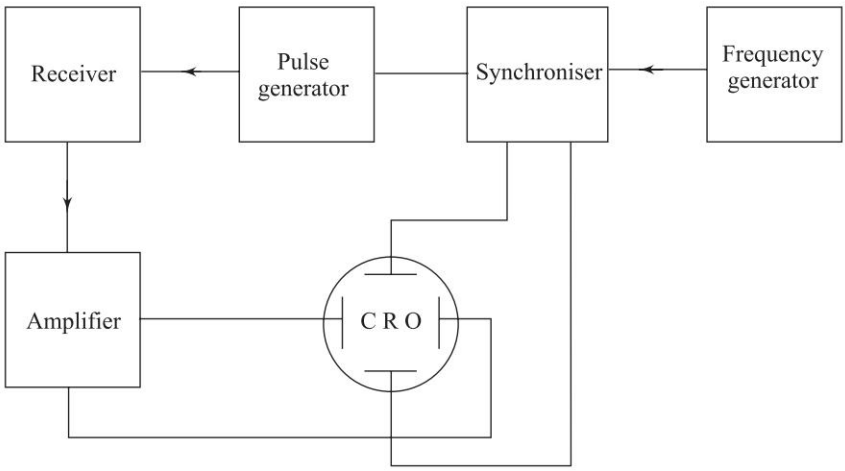


Fig. 5.14 *Block diagram of pulse echo method*

It consists of an ultrasonic frequency generator and receiver, and a cathode ray oscilloscope. A high potential difference from the continuous wave oscillator is applied to the piezoelectric transducer and hence, the transducer generates ultrasonic waves.

The piezoelectric transducer resonating at a desired frequency is used for both the generation and detection of ultrasonic waves. The vertical transducer is coupled to the upper surface of the specimen (metal) as shown in Fig. 5.15a, with a good acoustic couplant to avoid any air gap between the transducer and the specimen.

The ultrasonic waves (signal *A*) are transmitted through the specimen by the vertical transducer. These waves travel to the other end of the material and get reflected at the boundary. The reflected signal (*B*) is received by the same transducer. The total distance travelled by the wave is 2*d,* i.e., twice the thickness of the specimen. The reflected signals are amplified and are found to be almost as that of the transmitted signals, as shown in Fig. 5.15b. Figure 5.15b shows that the specimen is free from defect.

On the other hand, if there are some defects on the specimen, like a small hole as shown in Fig. 5.15c, the ultrasonic waves are reflected back depending upon the depth of the defect, i.e. some of the waves get reflected at the surface of the hole, while the others from the boundary of the specimen. This results in a time delay between the waves (signal *B*), those reflected at the surface of the hole and at the boundary of the specimen (signal *C*). The resultant signals *B* and *C* are as shown in Fig. 5.15c. Thus, the time delay between the transmitted (*A*) and the reflected signals (*B*) at the surface of the holes gives the depth of the defect from the upper surface of the specimen.

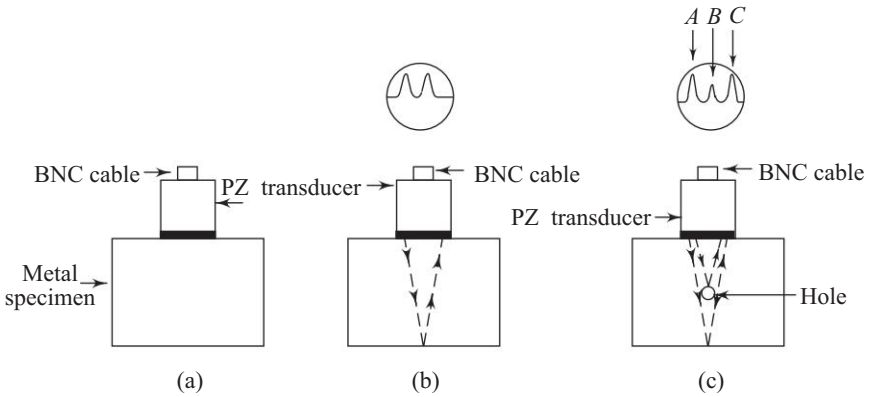


Fig. 5.15 *Determination of flaw in metals*

This technique is generally used in the testing of boiler drum plate, castings, forging, bar stock, etc. This type of test can be done safely under manufacturing or operating conditions.

5.9 APPLICATIONS—MEDICINE

Ultrasounds find widespread applications in the field of medicine for both diagnostic and therapeutic applications. In the following sub-sections, some of the applications, both in diagnostic and therapeutic fields, are discussed briefly:

5.9.1 Blood Flow

The principle involved in this measurement is the Doppler technique. The ultrasonic flowmeter measures the instantaneous flow of blood. The important flow meters which are available in the practical uses are listed below:

(i) Transit time
(ii) Continuous wave Doppler
(iii) Pulsed Doppler, and
(iv) Duplex scanner flowmeters

Let us discuss in detail the principle and working of Doppler shift flowmeter. The Doppler shift technique is used to determine the velocity of blood flow in veins and arteries. In order to obtain better resolution, the frequency of the ultrasound used in this technique ranges from *5* to10 MHz.

The principle of blood-flow measurement is shown in Fig. 5.16. The ultrasound transducer is positioned on the skin as shown in Fig. 5.16. The ultrasound produced by the transducer strikes the blood vessels at an angle q as shown in Fig. 5.16. The incident beam gets reflected back by the blood particles, which are moving with a velocity v relative to the vessel. The horizontal component of the velocity $v \cos q$ is passed to the direction of the beam.

The McGraw·Hill Companies

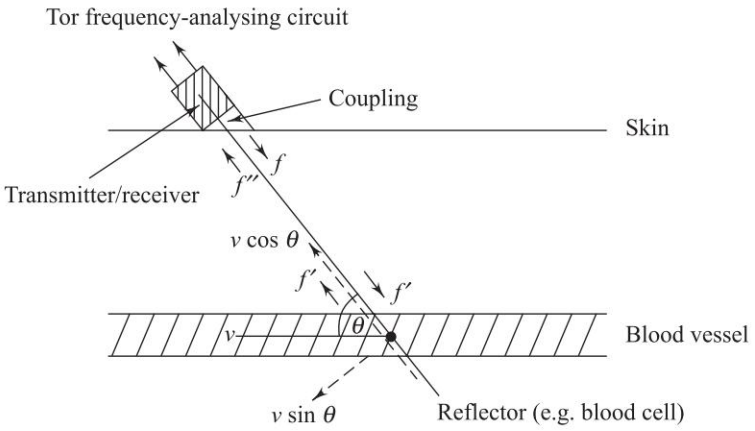


Fig. 5.16 *Blood-flow measurement*

Thus, the blood moving with velocity $v \cos \theta$ towards the receivers (reflectors) produces a Doppler shift and is given by Eq. (5.17),

$$\Delta f_3 \frac{2 f v \cos \theta}{u} \tag{5.17}$$

From Eq. (5.17), by knowing the values of v, u, f and q, one can determine the velocity of the blood motion.

The block diagram of the experimental set-up used to measure blood low is shown in Fig. 5.17.

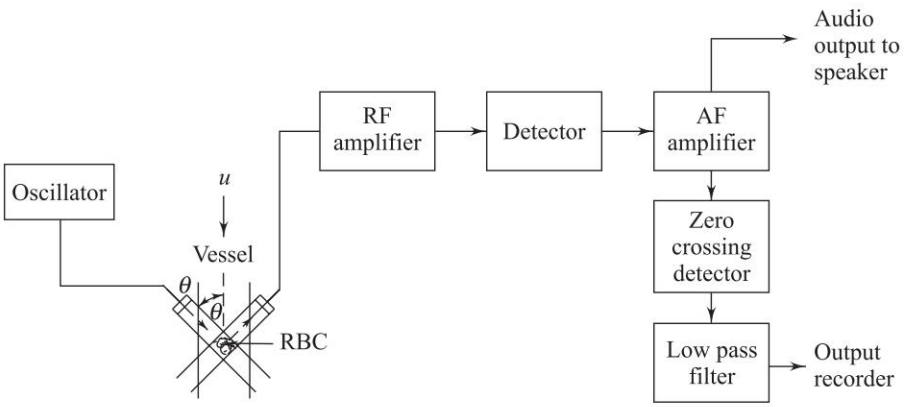


Fig. 5.17 *Block diagram of the experimental set-up to measure the blood flow*

It is the simplest instrument used for blood-flow measurement. The ultrasound beam is generated by the transducer and is passed into the blood-vessel walls. The ultrasonic wave gets scattered back by the blood cells. The same transducer receives back the scattered signals. The receiver and transmitting transducers are identical. The electrical signal received by the RF amplifier consists of two components. The first component is of larger amplitude excitation frequency, which is directly coupled from transmitter to the receiver, while the second component is a very small amplitude Doppler shift, which is scattered

from blood cells. A simple frequency to voltage converter provides a quantitative output to a recorder. The zero crossing detector emits a fixed area pulse each time the audio signal cross the zero axis. The low pass filter selects the difference in frequencies and produces an output proportional to the velocity of the blood cells.

The main advantages of this system are that the surgical penetration of the vessel is unnecessary. Continuous flow measurements can be made without any risk or discomfort to the patient. The main disadvantage of this method is that it is not possible to differentiate the movement of the blood, whether it is towards or away from the transducer. The recently developed Doppler system incorporates colour display, in which a range of colours are used to represent the different blood-flow velocities. Thus, the new colour Doppler imaging system is successful identifying any abnormal holes, blockages or valve defects, etc., from the blood-flow pattern.

5.9.2 Foetal Heart Movements and Blood Flow Measurements

The principle used in this method is a continuous-wave Doppler system. The essential components used in this system are shown in Fig. 5.18. The transducer used in this method has both a transmitting and receiving crystal in a single transducer. The transducer is mounted on the body surface, i.e., on the skin using a couplant in the mother's abdomen as shown in Fig. 5.18. The ultrasound beam generated by the transducer is transmitted into the reference medium (reflector) and gets reflected from the reflector. The same transducer receives the reflected signal. The output signal of the receiver consists of two frequencies, namely, the same frequency as that of the transmitter due to the stationary reflectors and the Doppler shifted frequency due to the moving reflectors. Thus, the mixer produces beat frequencies. The beat frequency is equal to the difference in the frequency between the two input frequencies, which is equal to Doppler shift. The beat frequency is amplified and connected with loudspeakers. The frequency of ultrasound beam used in this method is around 2 to 3 MHz. Certain Doppler shifts produced characteristic sounds.

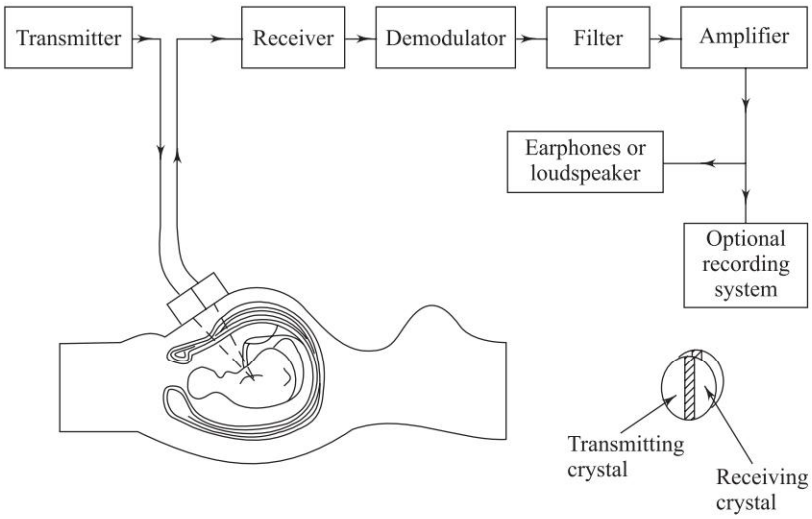


Fig. 5.18 *Block diagram of the experimental set-up for foetal heart movements and blood circulation*

The McGraw·Hill Companies

This method is useful in determining the presence of pulsating heart and blood flow in the foetus. The difference in the pulse rate is used to differentiate the blood flow in the mother's blood vessels from the foetus. Thus, the variation in blood stream velocity is used to identify and examine different parts of the foetus.

5.9.3 Therapeutic Application

In general, most of the physical therapy methods produce symptomatic, palliative supporting or prophylactic effects. Different methods are available to develop heat in tissues. Ultrasonic therapy is one such method and is similar as short-wave diatherm medicine. The ultrasonic therapy is more effective in cases where short wave treatment has failed and where a localisation of heat effect to the tissue is required.

In ultrasonic therapy, a low-frequency and high-power ultrasound beam is penetrated to the tissues. Thus, the propagation of ultrasound into the tissues produces heating effect and a high-speed vibration of tissues. The heating effect in this case is due to the absorption of ultrasonic energy by the tissues. As a result, ultrasonic energy produces a massage in tissues, which is a normal treatment of soft tissue in physical medicine. The ultrasonic medicine can be applied to a greater depth which is not possible manually and in some cases where the pressure cannot be exerted by hand due to intolerable pain to the patient. The most common therapeutic applications of ultrasonic are the treatment of diseases of peripheral nervous system (neuritis), skeletal muscle system (arthritis) and skin (ulcers, etc.). The power required in most of the ultrasonic therapies is less than 3 W cm^{-3}.

In ultrasonic therapy, the transducer is excited by giving the energy from conventional oscillators. The ultrasonic waves are emitted from the front surface of the transducer. The transducer is fixed to the place where the ultrasonic therapy is required using a couplant to provide contact between the transducer and body. The ultrasonic dosage can be controlled by any one of the three variables namely, frequency, intensity and duration of exposures.

Another method of application of ultrasonic therapy is also in use. If the treatment parts are not uniform or are wounds, the treatment will be carried out in a water bath. In this method, one can avoid the mechanical contact of tissue, which is already under damaged condition. In this case, one can also slightly increase the temperature of the water bath. This method is not generally preferred since the dosage of the ultrasonic frequency is not controllable.

5.9.4 Biological Effect

The biological effect of ultrasounds is either due to thermal or non-thermal effect. In thermal effect, when a beam of ultrasound is passed into the medium, a heating effect is produced within the medium. In non-thermal effect, it is due to the formation of cavitations. When an ultrasonic beam passes through the liquid medium, the ultrasonic energy creates small microbubbles.

In high-intensity ultrasonic therapy, a careful control of the beam is more essential due to higher absorption coefficient of bones. Similarly, one also has to consider the gas bubbles produced by the ultrasonic beam when it passes through the body. The bubbles produced in ultrasonic therapy will distrupt the cells and tissues.

5.9.5 Phonocardiography

It is an instrument used for recording the sounds connected with the pumping actions of the heart. These sounds are used to explore information about the heart rate, rhythmicity,

and effectiveness of blood pumping and valve action. The sounds produced by healthy hearts are remarkably identical while the abnormal sounds are due to specific physical abnormalities. An instrument known as *phonocardiograph* is very important and it diagnostically records the waveform of the heart sounds. The recorded graphs of the heart sounds are called *phonocardiogram*.

The sounds are produced by the mechanical events that occur during the heat cycle. There sounds are due to the movement to heart walls, closure of walls and turbulanece and leakage of blood. The frequency of the sounds are generally in the range of 30 to 100 Hz and the duration is about 50 to 100 ms. The second sound is higher in pitch than the first with frequencies above 100 Hz with a deviation 25 to 50 ms. The third and fourth sounds produced by the heart are much lower in intensity and hence inaudible. The sounds are produced principally by closure of the valves between the upper and lower chambers of the heart.

These sounds are normally recorded by using any one of the two types of microphones, namely, contact or air-coupled microphones for recording phonocardiogram. There are two types of microphones, namely, contact or dynamic type based on the operation principle. These microphones generate potentials when mechanical heart sound is applied to them. The amplifier amplifies these signals with a frequency range from 20 to 2000 Hz. The low pass filters are used to control the noise. The readout of a phonocardiograph is either a high-frequency chart recorder or an oscilloscope. A typical operation of heart sound is shown in Fig. 5.19. A possible heart disorder is indicated by the presence of high frequecy in the phonocardiogram.

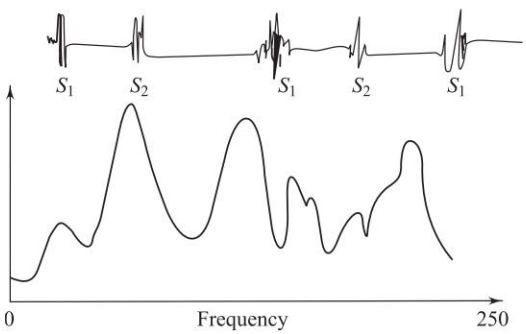


Fig. 5.19 *Frequency spectrum of heart sound*

5.10 ACOUSTIC HOLOGRAPHY

The principle, production and the application of optical holography are discussed in this section. It is a well-known method for storing three-dimensional images on a two-dimensional recording surface. The essential requirements for holography are monochromatic and coherent sources.

An ultrasonic wave is a coherent radiation and the same can be used similar to an optical source for the production of acoustical holography.

5.10.1 Basic Principle

The principle of acoustic holography is similar to that of optical holography. Optical holography is based on interference phenomena obtained on a photographic plate due

The McGraw·Hill Companies

to the superposition of two waves, namely, the reflected wave from the object which contains details about the object, and the reference wave. The recorded image on the photographic plate is called *hologram*.

In acoustic holography, different techniques are employed, such as transmission technique, pulse reflection technique, etc. The coherence conditions can be easily achieved, since an ultrasonic transducer produces waves in the narrow frequency band.

5.10.2 Transmission Acoustic Holography

The experimental arrangement used in transmission acoustic holography is shown in Fig. 5.20. The common generator G is used to excite both transducers T which generate a continuous wave oscillation with certain modulations (This results in reduction of standing waves). The reference wave and the wave passed through the specimen form an interference pattern on the liquid surface. This pattern on the liquid surface is stationary for continuous oscillation. The observed interference pattern gives rise to a hologram.

The hologram can be reconstructed by projecting a laser beam L through a semitransparent mirror M on the holographic plate. The laser beam is reflected through an optical system and a three-dimensional image is formed on the screen. The aperture C serves as a spatial filter to eliminate multiple images.

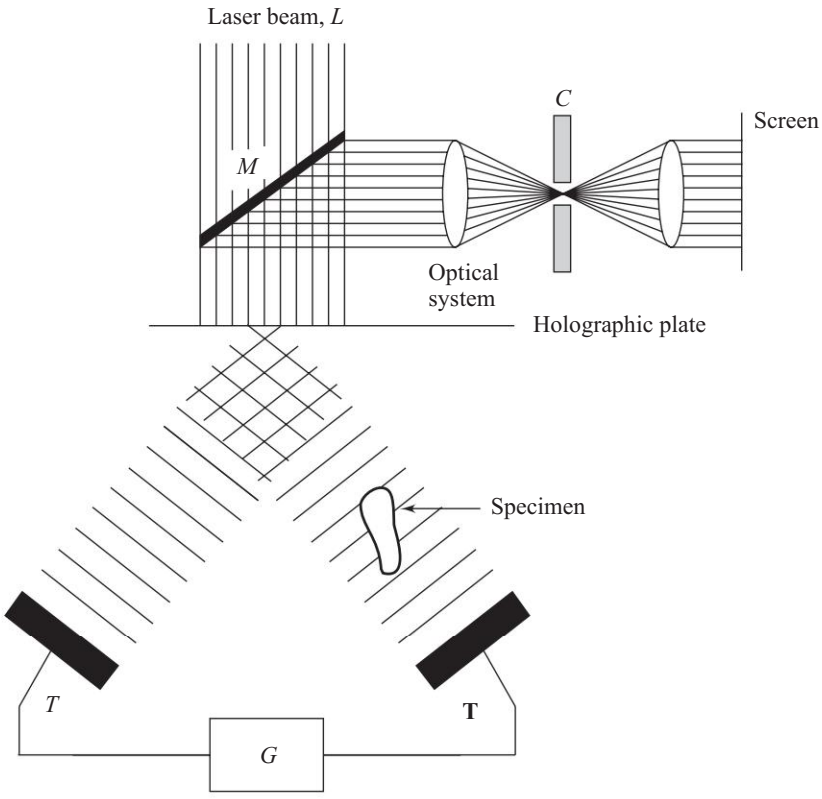


Fig. 5.20 *Acoustic holography*

The above principle is satisfactory at the laboratory level only and is based on the transmission method. The transmission method is not suitable for ultrasonic NDT applications. Therefore, ultrasonic pulse method is used for NDT, since the major application is to detect the defect in the specimens, such as flaw, hole, crack, discontinuity, etc. Therefore, a new principle called *linear acoustic holography* is used for ultrasonic NDT applications.

5.10.3 Pulse Reflection Acoustic Holography

The experimental arrangement used for pulse reflection acoustic holography is shown in Fig. 5.21. The specimen *S* under study is immersed in a water bath. The ultrasonic transducer probe *T* is just above the specimen and can move from one end of the specimen to the other end. The transducer is excited by a generator *G* and this results in high frequency ultrasonic pulses.

The ultrasonic pulses are passed through the specimen. During the propagation of the ultrasonic waves into the specimen, they get reflected at the bottom surface due to change in acoustic impedance if there is no defect in the specimen. On the other hand, if there is a defect in the path of the waves, they get reflected due to the change in acoustic impedance. The reflected pulses are received by the same transducer and are passed into the scanner *C*. The received pulses are amplified using the amplifier *A* and then passed to the mixer *M* along with reference signals from the source *R*. The output of the mixer is modulated suitably and is connected to the CRT tube as shown in Fig. 5.21. Thus, the hologram is viewed on the CRT tube and is recorded using a suitable arrangement.

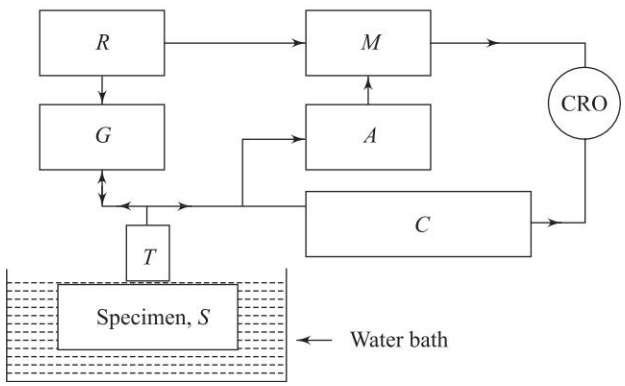


Fig. 5.21 *Schematic diagram for pulse reflection holography*

The experimental arrangement used for the reconstruction of the holoram is shown in Fig. 5.22.

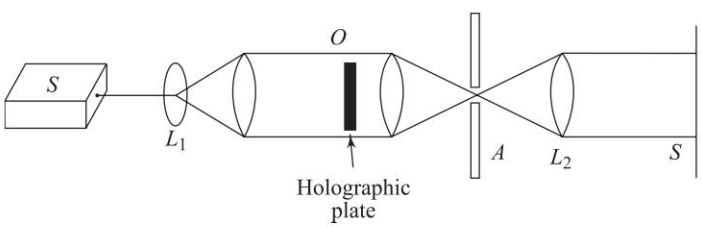


Fig. 5.22 *Experimental arrangement for the reconstruction of hologram*

The McGraw·Hill Companies

The reconstruction of the hologram is made using a laser beam. The laser beam is generated by the source *S,* and its wavelength is in the order of 0.5 mm. The laser beam is made incident on the lens L_1. The scattered laser beam from the lens L_1 is made to fall on the optical system *O*. The optical system consists of two lenses into which the holographic plate is inserted. When the scattered laser beam is incident on the optical system, some of it will be captured in the hologram. The rays of the laser beam for which no diffraction was captured by the hologram are rejected by the aperature *A*. The actual hologram is viewed on the screen *S* after passing through the lens L_2.

A recent development in acoustic holography is *linear acoustic holography*.

5.10.4 Linear Acoustic Holography

Principle Linear acoustic holography is based on pulse technique. The ultrasonic transducer is scanned along a line and produces a wide sound field in the plane of scane line. The echoes received from the specimen are multiplied with the reference signals at two different angles, namely, 0° and 90°. As a result, real and imaginary parts of the different parts of the points which are scanned on the specimen are obtained. These data are stored in a computer using quadratic phase function. Thus, the computer defines the reconstruction of plane, observation angle and also collects the ultrasonic field intensity at the defect point. Thus, the intensity of the ultrasonic field is used to find the defect size on the specimen. In acoustic holography, the resolution is very poor along the direction of ultrasound field. The resolution can be increased by means of using a focusing probe or multifrequency holography. The same can also be achieved by using a new technique known as Synthetic Aperture Focusing Techniques (SAFT). In this technique, the resolution is increased using a computer programme.

5.10.5 Applications

Some of the applications of ultrasonic holograms are given below:

1. It is used in ultrasonic NDT to detect defects such as holes, cracks, flaws, etc.
2. One can detect the size of the defect even of the order of a mm using SAFT technique.
3. They are used in medicine for imaging.

5.11 GENERAL APPLICATIONS

Soldering Aluminium cannot be soldered by ordinary soldering method. To solder aluminium, ultrasonic waves are used in addition to the electrical soldering iron. The ultrasonic waves remove the oxide film and facilitate soldering.

Biology Ultrasonic waves are used to kill small animals like rats, fish, frogs, etc. Microorganisms like bacteria are torn into pieces or burnt and yeast loses its power of reproduction.

Medicine Ultrasonic waves are used

(i) to detect tumours in human body
(ii) for bloodless brain surgery as well as painless extraction of broken teeth

(iii) to look at the development of a child in a mother's womb without risk to the mother or the child; this field is known as *Sonography*

Chemistry When ultrasonic waves come in contact with substances, they set up high frequency vibrations in them. These vibrations bring about an intimate mixing of particles of matter and cause or accelerate molecular reactions. Their applications are listed below:

(i) They accelerate the process of crystallisation and enable smaller and uniform crystals to be obtained.
(ii) They form stable emulsions of even immiscible liquids like water and oil, water and mercury, etc.
(iii) They coagulate, i.e., disperse fine particles of solid or liquid into a gas, e.g., dust, smoke, mist, etc.

Industry

(i) They are used to form alloys of uniform compositions.
(ii) They are used for cleaning clothes and tiny objects like needles and parts of watches.
(iii) They are used for drilling holes in glass or steel plates by using a glass rod oscillating at ultrasonic frequency.

Keypoints to Remember

- The sound waves whose frequencies lie above the audible frequency of 20,000 Hz are called ultrasonic waves.
- Different modes of ultrasonic waves are longitudinal, transverse or shear, surface or Rayleigh and plate or Lamb waves.
- In longitudinal waves particles vibrate parallel to the direction of wave propagation.
- In shear waves, particles vibrate perpendicular to the direction of wave propagation.
- The velocity of shear waves is half of longitudinal waves.
- Surface waves are the waves whose amplitude decays rapidly with the depth of propagation of waves in the medium.
- Plate waves are produced when a surface wave is introduced into a material having thickness equal to three times the wavelength or less.
- Dilational and bonding Lamb waves are the two modes of plate waves.
- When the particle vibrates in a direction parallel to the surface, it produces a new kind of wave known as Love wave.
- Different methods of generation of ultrasonic waves are mechanical generators, magnetostriction and piezoelectric effect.

The McGraw·Hill Companies

- When a ferromagnetic material in the form of a bar is subjected to an alternate magnetic field, it undergoes alternate contraction and expansion at a frequency equal to the frequency of the applied magnetic field. This phenomenon is known as magnetostriction method.
- When the opposite faces of a crystal are subjected to squeezing, twisting or bending, a potential difference is developed across the perpendicular opposite faces. This effect is known as piezoelectric effect.
- The change in velocity of ultrasonic waves with frequency is known as dispersion of ultrasonic waves.
- The intensity of the ultrasonic waves measured per unit time per unit area perpendicular to the direction of propagation decreases with distance from the source. This phenomenon is known as attenuation.
- SONAR is the acronym for sound navigation and ranging.
- Echo meter is used to measure the depth of a sea directly.
- NDT is the acronym for non-destructive testing.
- The velocity of blood flow in veins and arteries is determined using the Doppler shift technique.
- The frequency of ultrasound beam used for foetal heart movements is 2 to 3 MHz.
- The power required for ultrasonic therapy is 3 W cm^{-3}.
- In linear acoustic holography, the ultrasonic waves are passed into the specimen whose image is to be obtained and the image patterns are constructed employing the received ultrasonic waves.
- In pulse reflection holography, the ultrasonic waves are incident on the specimen whose image is to be obtained, and then form the reflected waves, the images are constructed.
- The image resolution is increased by means of using a focusing probe or multi frequency holography, known as Synthetic Aperture Focusing Technique (SAFT)
- Sonography is the technique used to obtain the image of a child in a mother's womb without risk to the mother's child.

Solved Problems

Example 5. 1 Calculate the frequency to which a piezoelectric oscillator circuit should be tuned so that a piezoelectric crystal of 0.1 cm thickness vibrates in its fundamental mode to generate ultrasonic waves. (Young's modulus and density of the materials of the crystal are 80 GPa and 2654 kg m^{-3}).

Given Data

Thickness	$t = 1 \text{ mm} = 1 \times 10^{-3}$ m
Density	$r = 2654$ kg m^{-3}
Young's modulus	$Y = 8.0 \times 10^{10}$ N m^{-2}

Solution

The fundamental frequency for electronic oscillator circuit is,

$$f = \frac{p}{2t}\sqrt{\frac{Y}{\rho}}$$

where $p = 1$ for fundaental vibration $f = \frac{1}{2\times 0.001}\sqrt{\frac{8.0\times 10^{10}}{2654}}$

$$= 2.7451 \times 10^6 \text{ Hz.}$$

The fundamental frequency of the piezoelectric oscillator is 2.74 MHz.

Example 5.2 Calculate the fundamental frequency of vibration when a quartz crystal of 0.15 cm thickness is vibrating at resonance. Given that Young's modulus of quartz is 7.9×10^{10} N m^{-2} .

Given Data

Thickness of the crystal $t = 0.15 \times 10^{-2}$

Density of quartz $\rho = 2650$ kg m^{-3}

Young's modulus of quartz $Y = 7.9 \times 10^{10}$

Solution

The fundamental frequency of a qurtz crystal is given by

$$f = \frac{p}{2t}\sqrt{\frac{Y}{\rho}}$$

For fundamental vibraion, $p = 1$. Therefore,

$$f = \frac{1}{2t}\sqrt{\frac{y}{f}}$$

$$f = \frac{1}{2\times 0.5\times 10^{-2}}\sqrt{\frac{7.9\times 10^{10}}{2650}}$$

$$= 0.546 \times 10^6$$

The fundamental frequency of a quartz 0.546×10^6 Hz.

Example 5.3 A ultrasonic generator consists of a quartz plate of 0.7 mm thickness and 2800 kg m^{-3} density. Find the fundamental frequency of ultrasonic waves if the Young's modulus of quartz is 8.8×10^{10} N m^{-2} .

Given Data

Thickness of the quartz $t = 0.7$ mm $= 0.7 \times 10^{-3}$ m

Density of quartz $\rho = 2800$ kg m^{-3}

The McGraw·Hill Companies

Young's modulus $Y = 8.8 \times 10^{10}$ N m^{-2}

Solution

The fundamental frequency of quartz piece,

$$f = \frac{p}{2t}\sqrt{\frac{Y}{\rho}}$$

where $p = 1$ for fundamental frequency

$$f = \frac{1}{2\times 0.7\times 10^{-3}}\sqrt{\frac{8.8\times 10^{10}}{2800}}$$

$$= 4.04 \times 10^{6} \text{ Hz.}$$

The fundamental frequency is 4 MHz.

Example 5.4 A nickel of 10 cm length with a density of 8.1×10^{3} kg m^{-3} and a Young's modulus of 8.2×10^{11} N m^{-2} is used in a magnetostriction oscillator. Determine the fundamental frequency of the ultrasonic wave generated .

Given Data

Length of the nickel $t = 10 \times 10^{-2}$

The density of nickel $\rho = 8.1 \times 10^{3}$ kg m^{-3}

Young's modulus of nickel $Y = 8.2 \times 10^{11}$

Solution

The frequency of ultrasonic waves produced by the magnetostrition oscillator is

$$f = \frac{1}{2l}\sqrt{\frac{Y}{\rho}}$$

$$f = \frac{1}{2\times 10\times 10^{-2}}\sqrt{\frac{8.2\times 10^{11}}{8.1 \times 10^{3}}}$$

$$= 0.05 \times 10^{6}$$

The frequency of produced ultrasonic waves by nickel is 0.05x10^{6} Hz.

Example 5.5 An ultrasonic beam is used to determine the thickness of a steel plate. It is noticed that the difference between two adjacent harmonic frequencies is 50 kHz. The velocity of sound in steel is 5000 m s^{-1}. Determine the thickness of the steel plate.

Given Data

Velocity of sound in steel $u = 5000$ m s^{-1}

The difference between two adjacent frequencies $f = 50$ kHz

$= 50 \times 10^{3}$ Hz

Solution

The frequency of the ultrasonic wave $f = \dfrac{u}{2d}$

Let f_n and f_{n-1} be the two adjacent harmonic frequencies.

The difference between the two harmonic frequencies is equal to the frequency of the fundamental note.

$$f = f_n - f_{n-1} = 50 \text{ kHz}$$
$$= 50 \times 10^3 \text{ Hz}$$

ie.,
$$f = f_n - f_{n-1} = \frac{u}{2d}$$
$$d = \frac{u}{2(f_n - f_{n-1})}$$

Substituting the values of u, f_n and f_{n-1} in the above equation, we get,

$$d = \frac{5000}{2 \times 50 \times 10^3}$$
$$= 0.05$$

Therefore, the thickness of the plate is 0.05 m.

Example ***5.6*** Longitudinal standing waves are set up in a quartz plate with antinodes at opposite faces. The fundamental frequency of vibration is given by the relation, $f = 2.87 \times 10^3/t$, where f is the thickness of the plate in metre. Compute (i) Young's modulus of the quartz plate, (ii) the thickness of the plate required for a frequency of 1200 kHz. The density of quartz is 2660 kg m^{-3}.

Given Data

The fundamental frequency $f = 2.87 \times 10^3/t$ Hz m^{-1}

Density of quartz $\rho = 2660$ kg m^{-3}

Solution (i) The frequeny of vibration is given by

$$f = \frac{p}{2t}\sqrt{\frac{Y}{\rho}}$$

where p = 1, 2, 3, etc.

For fundamental vibrations p = 1

Thereore,

$$f = \frac{1}{2t}\sqrt{\frac{Y}{\rho}}$$

$$\frac{2.87 \times 10^3}{t} = \frac{1}{2t}\sqrt{\frac{Y}{2660}}$$

or
$$Y = 4 \times (2.87 \times 10^3)^2 \times 2660$$
$$= 8.764 \times 10^{10} \text{ N m}^{-2}$$

The McGraw·Hill Companies

The Young's modulus of the quartz crystal is 8.76×10^{10} N m^{-2}.

(ii) The frequency of vibration = 1200 kHz

We know that,

$$f = \frac{1}{2t}\sqrt{\frac{Y}{\rho}}$$

or the thickness

$$t = \frac{1}{2f}\sqrt{\frac{Y}{\rho}}$$

$$= \frac{1}{2\times1200\times10^3}\sqrt{\frac{8.764\times10^{10}}{2660}}$$

$$= 2.39165 \times 10^{-3}$$

The thickness of the crystal is 2.39×10^{-3}m.

Example 5.7 Determine the maximum audio frequency of a Doppler ultrasonic blood flowmeter that has a carrier frequency of 7 MHz, a transducer angle of 45^0, a blood velocity of 150 cm s^{-1} and an acoustic velocity of 1500 m s^{-1}.

Given Data

The ultrasonic frequency $f = 7$ MHz

The transducer angle $\theta = 45^0$

The blood velocity $v = 1.5$ ms^{-1}

The ultrasonic velocity $U = 1500$ ms^{-1}

Solution We now that the Doppler shift $\Delta f = \frac{2 f v \cos\theta}{U}$

Substituting the values of f, U, θ and v in the above equation, we get

$$\Delta f = \frac{2\times7\times10^6\times1.5\times\cos 45^0}{1500}$$

$$= 10 \text{ kHz}$$

The audio frequency of the Doppler flowmeter is 10 kHz.

Objectives Questions

5.1 The frequency of ultrasonic wave is _____.

5.2 Ultrasonic waves are classified based on the mode of vibration of _____.

5.3 During the transmission of longitudinal waves, the particles in the medium vibrate ___to the wave propagation.

The McGraw·Hill Companies

5.4 Longitudinal waves propagate in _____, _____ and _____ medium.
5.5 The particles vibrate at right angles to the direction of propagation during the propagation of _____ waves.
5.6 _____ waves can pass only through solids.
5.7 _____ waves amplitude decays rapidly with depth of propagation of the waves in the medium.
5.8 In Rayleigh waves, the particle vibration is _____
5.9 _____ and _____ are the modes of plate waves.
5.10 In Love waves, the particle vibrates in a direction _____ to the surface.
5.11 The frequency of ultrasonic waves generated in a mechanical generator method is

a) $\frac{u}{4(l+x)}$ b) $\frac{u^2}{4(l-x)}$

c) $\frac{u}{4(l+x)}$ d) $\frac{u}{4(l+x^2)}$

5.12 Magnetostriction effect is obtained from _____ materials.
5.13 The frequency of ultrasonic waves produced by magnetostriction method is

a) $\frac{1}{2l}\sqrt{\frac{Y^2}{P}}$ b) $\frac{1}{2l}\sqrt{\frac{Y}{P}}$

c) $\frac{1}{4l\sqrt{\frac{Y}{P}}}$ d) $\frac{1}{2l\sqrt{\frac{2Y}{P}}}$

5.14 _____ crystals is used to produce longitudinal ultrasonic waves.
5.15 Shear ultrasonic waves are produced employing _____crystal.
5.16 _____ a powder is used to obtain the nodes and antinodes in Kundt's tube.
5.17 The wavelength of light λ = _____ for the first order.
5.18 During the propagation of ultrasonic waves, the compression effect will produce _____.
5.19 When the ultrasonic waves propagate through the medium, a locking effect is produced due to _____
5.20 The velocity of longitudinal waves in crystal is equal to

a) $\sqrt{\frac{Y}{P}}$ b) $\sqrt{\frac{Y^2}{P^2}}$

c) $\sqrt{\frac{Y}{2P}}$ d) $\sqrt{\frac{Y}{P}}l$

5.21 The frequency of piezoelectric oscillator is equal to

a) $\frac{1}{2\pi}\sqrt{L^2C^2}$ b) $2\pi\sqrt{LC}$

c) $\frac{1}{2\pi}\sqrt{LC}$ d) $\frac{1}{2\pi\sqrt{LC}}$

5.22 The intensity of ultrasonic waves at a distance x from the source is

a) $I_o e^{-\alpha x}$ b) $I_o e^{-\alpha xt}$

c) $I_o e^{-2\alpha x}$ d) $I_o e^{-2\alpha x}$

The McGraw·Hill Companies

5.23 The coefficient of ultrasonic of ultrasonic waves is _____

a) $\frac{1}{x}\ln\left(\frac{I}{I_o}\right)$ (b) $\frac{1}{2x}\ln\left(\frac{I}{I_o}\right)$

c) $\frac{1}{x}\ln\left(\frac{I+I_1}{I_0}\right)$ d) $\frac{1}{2x}\ln\left(\frac{I^2}{I_o{}^2}\right)$

5.24 Dispersion of ultrasonic waves means the change in velocity of the waves with _____.

5.25 The attenuation of ultrasonic waves due to heat radiation is _____ of frequency.

5.26 The depth of sea is obtained using the relation _____.

5.27 _____ is used to measure the depth of sea directly.

5.28 The frequency of ultrasound is used in blood flow measurement is from _____ to _____ MHz.

5.29 The Doppler shift is _____

a) $\frac{2f^2V^2\cos\theta}{u}=\Delta f$ b) $\frac{2fV\cos\theta}{u}=\Delta f$

c) $\frac{2fV\cos\theta}{u^2}=\Delta f$ d) $\frac{V\cos\theta}{u^2}=\Delta f$

5.30 Doppler imaging systems are used to identify if the abnormalities from the blood flow pattern are _____, _____ and _____.

5.31 _____ MHz frequency of ultrasound beam is used to study the foetal heart movements.

5.32 The power required in ultrasonic therapy is _____ W m cm^{-3}.

5.33 _____ and _____ sources are essential requirements of holography.

5.34 The image pattern is obtained in pulse reflection holography using the principle of

a) rarefaction and compression b) absorption

c) reflection d) transmission

5.35 Linear acoustic holography is based on _____ technique.

5.36 The technique used to look at the development of a child in a mother's womb is _____.

5.37 The ultrasonic waves are generated by the _____.

5.38 _____ receives the ultrasonic waves.

5.39 Heart disorder is indicated by the presence of _____ in phonocardiogram.

KEY

5.1	Greater than 20 kHz	5.2	Particles
5.3	Parallel	5.4	Solid, liquid, gas
5.5	Shear	5.6	Shear
5.7	Surface or Rayleigh	5.8	Elliptical
5.9	Symmetrical and asymmetrical	5.10	Parallel
5.11	a)	5.12	ferromagnetic material
5.13	a)	5.14	X-cut
5.15	Y-cut	5.16	Lycopodium
5.17	$2d \sin \theta$	5.18	Heat

5.19 rarefaction
5.20 a)
5.21 d)
5.22 $I_o e^{-2x}$
5.23 a)
5.24 frequency
5.25 Independent
5.26 $h = (ut/2)$
5.27 Echo meter
5.28 5, 10
5.29 a)
5.30 holes, blockages, valve defects
5.31 2 to 3
5.32 3
5.33 Monochromatic and coherent
5.34 Reflection
5.35 Pulse
5.36 Reflection
5.37 Transmitter
5.38 Receiver
5.39 High frequency

Exercises

5.1 An ultrasonic wave is being propagated through a liquid which is illuminated by radiation of wave length 5893 Å. If the frequency of the ultrasonic waves is 25×10^3 Hz and the velocity is 800 m s^{-1}, calculate the angle of first order of diffraction

5.2 In an ultrasonic interferometer, an ultrasonic wave having a frequency of 30×10^3 Hz is propagating. If the interferometer is filled with a gas having the velocity of sound as 450 m s^{-1}, calculate the distance at which the recording meter will show maximum deflection

5.3 Calculate the attenuation coefficient of a sound wave of frequency 50 kHz, when propagating through a liquid dioxane having a coefficient of viscosity of 1.068×10^3 N s m^{-2}. (Velocity of sound in the liquid is 1,320 m s^{-1}. The equilibrium density = 1018 m^{-3}.)

5.4 A 2.5 MHz ultrasound beam travels at 1.5 km s^{-1} through soft tissue before being reflected normally from a moving interface of a Doppler shift of 500 Hz is detected in the reflected beam. Determine the velocity of the moving surface.

Short Answer Questions

5.1 What are ultrasonic waves?

5.2 What is the main difference in the quality of ultrasonic waves produced by piezoelectric and magnetostriction oscillator?

5.3 What are longitudinal ultrasonic waves?

5.4 What is transverse ultrasonic wave?

5.5 What are Rayleigh wave?

The McGraw·Hill Companies

5.6 What is a stonely wave?
5.7 What is a plate wave?
5.8 What is an asymmetrical Lamb wave?
5.9 What is a Love wave?
5.10 What is the basic principle of ultrasonic holography?
5.11 Mention different types of ultrasonic holography.
5.12 What is acousstical grating?
5.13 State any two methods of the piezoelectric oscillator.
5.14 What is a hologram?
5.15 What is the reason for using ultrasonic waves to produce holograms?
5.16 What is meant by piezoelectric effect?
5.17 What is magnetostriction effect?
5.18 What is ment by inverse piezoelectric effect?
5.19 What are the different methods used for the production of ultrasonic waves.
5.20 Mention the properties of the ultrasonic waves?
5.21 Explain the production of ultrasonic waves using Galton's Whistle.
5.22 What is meant by attenuation of ultrasonic waves?
5.23 What is meant by dispersion of ultrasonic waves?
5.24 Mention any two methods of detecting ultrasonic waves?
5.25 Mension some of the uses of ultrasonic waves.
5.26 What is SONAR? Explain how ultrasonic waves are used in SONAR.
5.27 Write a few applications of ultrasonics?
5.28 Explain how ultrasonic waves are used in non-destructive testing.
5.29 Mention the different applications of ultrasonic in medicine.
5.30 What is ment by phonocardiography?

Descriptive Questions

5.1 Explain different types of ultrasonic waves with a neat sketch.
5.2 Explain with a neat sketch the construction and production of ultrasonic waves using piezoelectric oscillator.
5.3 What is meant by magnetostriction effect? Explain how are the ultrasonic waves are produced using a magnetostriction oscillator.
5.4 Write an essay about the absorption and dispersion of ultrasonic waves.
5.5 Write the applications of ultrasonic waves in science, medicine and industry.
5.6 Explain the process of non-destructive testing of materials using ultrasonic waves by pulse echo overlap method.
5.7 Write an essay about the applications of ultrasonic in medicine.
5.8 What is Doppler effect? Explain how the Doppler effect is used in ultrosonics to study foetal heart movement and blood flow measurements.
5.9 Explain the method of production and reconstruction of an ultrasonic hologram.
5.10 Explain how ultrasonic waves are produced. Explain in detail how the ultrasonic pulse techniques used in SONAR.

Chapter

6

Non-Destructive Testing

OBJECTIVES

- To study the importance of non-destructive testing methods in industry and medicine
- To study the various non-destructive testing techniques
- To gain the practical knowledge and explore the limitations of differet NDT methods
- To discuss the possible applications of NDT in industries

6.1 INTRODUCTION

With the invent of modern technologies, we are using a variety of machines and systems in different fields, such as power plants, industrial plants, transport machines, chemical industries, etc. The machines and systems consist of a number of individual components. The smooth performance of the machines and systems depend on the reliability of the operation and performance. One can increase the reliability of the components by giving better quality components, which in turn depend on the factors such as design, raw materials, properties and fabrication techniques. Different types of defects are generated during the operations. Therefore, the nature of the defect depends on various factors such as design, material properties, process technology and service condition. One has to increase the product quality during the processing stage and also increase the life of the components by examining the defects without affecting the performance.

A technique that is used to characterise the individual components, systems, machines and materials by examining the defects and inhomogeneity during the process by operating both off-line and on-line is known as Non-Destructive Testing (NDT). Now a days, a newer term known as *Non-Destructive Evaluation* (NDE) is coming up in this field, which includes the acceptance and rejection criteria of the non-destructive testing process.

The McGraw·Hill Companies

The role of NDT in everyday life is crucial and is necessary to assure the safety and reliability. The following are some of the examples where NDT plays a key role, namely, aircraft, motor vehicles, pipelines, trains, power stations, refineries, etc. In order to avoid expensive repairs or early replacements, and to ensure that the material products and equipment should possess the required design or reach the estimated life, the NDT techniques can be applied at each and every stage. One can successfully apply the NDT techniques to monitor the integrity of the components or structure throughout their life. In this chapter, an attempt has been made to discuss some of the available non-destructive testing methods in brief, along with suitable illustrations.

6.2 CLASSIFICATION OF TESTING METHODS

There are different methods of non-destructive testing, which start from the simple to the complicated ones. The methods which are used for the NDT depend on the physical properties of the materials. Even though there are different NDT techniques available, some of the universally accepted common methods are given below:

(i) Visual inspection
(ii) Liquid penetration inspection
(iii) Acoustic emission
(iv) Magnetic particle inspection
(v) Radiography
(vi) Ultrasonic inspection
(vii) Eddy current methods, and
(viii) Thermography

Let us discuss the above methods in brief in the following sections:

6.3 VISUAL INSPECTION

Visual inspection (VI) is a surface examination method, and it is the most widely employed NDT method among all other testings. It is very simple, easy, low in cost and is commonly applied before using any other testing, such as ultrasonic, magnetic particle, etc. This method provides more precise surface defects. This test can be carried out with or without optical instruments. In view of the rapid developments in the field of computer technology, the automated method provides high speed surface inspection, which is helpful for the subsequent inspection by other techniques.

6.3.1 Basic Principle

The principle of VI is the illumination of the specimen by high intensity light beam, and then the examination of the specimen surface through optical aids or through eye. In most of the testings, human eye is amused as a tool for VI. However, the same cannot be applied to distinguish the difference in the intensities. Therefore, visual inspection requires highly sensitive optical aids. The most commonly used optical aids for visual inspection are microscope, borescope or introscope, endoscopic, flexible fibre optic borescope or flexiscope, photography, electronic speckle pattern interferometry, etc.

The McGraw·Hill Companies

6.3.2 Identification of the Defects

Some of the general information about the components that can be obtained by VI are given below.

(i) The overall general condition of the component
(ii) The presence or absence of surface breaking from the strength point of view
(iii) The presence or absence of the weld defects
(iv) The shape and roughness of the surface

The above information about the components, obtained by VI, is used to proceed towards the other NDT testing methods and to suppliment the findings.

6.3.3 Optical Aids

Generally, the use of optical instruments in visual inspection is recommended rather than the human eye. In most of the VI methods, the area to be examined has to be well cleaned before testing. The optical instruments, for example, horoscopes or introscopes, which are normally known as *industrial telescopes*, are used to examine the internal surface of the tubes and piping, and also the holes in machines. The borescope, along with the fibre optical cables, is used to illuminate and retrieve images up to several metres. The ordinary microscope with higher resolving power is used to identify the minute defects and the details of the structure on the surface of the components. The endoscope is superior than the borescopes. The main advantage of an endoscope over a borescope is that the contact focus of object ranges from a millimetre to infinite distance. Holography is yet another technique used in VI. The advantage of this method is the three dimensional recording of the image of the entire surface of the components.

6.3.4 Applications

Some of important applications of visual inspection are given as follows:

(i) It is used to test the leakage/abnormal operations, etc., of the plant systems and components.
(ii) It is used to identify the defects such as cracks, etc.
(iii) It is used to identify the misalignments of the equipment and discontinuities, etc.

6.4 LIQUID PENETRANT TESTING

Liquid Penetrant Testing (LPT) is a method used to explore the surface breaking flaws by the natural accumulation of the fluid around the discontinuity. This is due the action of the surface tension effect and negligibly due to the viscosity effect of the liquid. The material whose surface is to be tested has to be cleaned completely by means of suitable procedure. This will facilitate the penetrant to get into the defects/cracks. This method is applicable to all types of material.

6.4.1 Basic Principle

The surface crack/defect in the workpiece/component is determined with very high precision by the liquid penetrant testing method. The basic principle behind this

The McGraw·Hill Companies

testing is the capillary action. The capillary action is due to the existence of surface tension of the liquid. The existence of cohesive forces in molecules is responsible for the surface tension. As a result of the capillary action, the liquid may rise in a tube or can be drawn into the crack. In most of the cases, the low viscosity liquids are used as penetrants, since the rate of flow of liquid within the crack is very high, rather than the highly viscous liquids. The penetrant liquid should also have good dissolving power so that a suitable dye or fluorescent compound can be added. The addition of a dye or a suitable fluorescent compound to the liquid penetrant will help in identifying the defects/cracks more precisely in the work piece when it is illuminated by visible or ultraviolet light. In addition, it is very effective in identifying the defects under dull light during the inspection.

Some important points to be considered to increase the resolution of the LP testing are given below:

(i) The surface of the workpiece/component should be cleaned using a suitable liquid before LP testing.
(ii) A proper selection of the penetrant liquid based on the viscosity and capillary action of the liquid is required.
(iii) A suitable dye or compound has to be selected to dissolve in with the liquid penetrant.

6.4.2 Experimental Procedure

The different stages of LP testing which will be carried out during the inspection are given below:

(i) Surface Preparation The cleaning of the surface of the workpiece/component is one of the first and most important steps to be carried out in LP testing. The presence of flakes, dirt, grease, chemicals, etc., on the surface of the test piece will prevent the penetrant to be drawn into the cracks/defects. Sometimes, the impurities get accumulated on the penetrant and give wrong information about the presence of defects, where there is no such actual defect/crack. Therefore, the preparation of a good surface, which is free from oil, dirt, grease, etc., is more essential. Thus, by employing a suitable solvent, brush, etc., a clean surface is obtained. In order to avoid the interaction of the cleaning solvent with the penetrant, the same is to be dried suitably after the cleaning. The existence of the crack after the surface cleaning is shown in Fig.6.1.

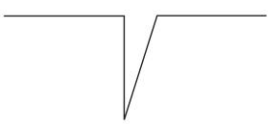


Fig. 6.1 *Workpiece-surface cleaning*

(ii) Penetrant Application The second stage of the LP testing is the penetrant application. The suitable fluorescent component is added with a low viscosity penetrant fluid and then applied on the surface of the test piece very evenly. The time required by the penetrant to be drawn into the crack depends on the crack size, shape and environmental conditions, such as temperature and surface inclination.

The McGraw·Hill Companies

Generally, the minimum time required for filling of the penetrant is in the order of 20–30 minutes. The workpiece after the application of the penetrant is shown in Fig.6.2.

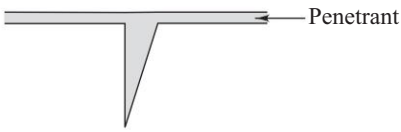


Fig. 6.2 *Workpiece-penetrant application*

(iii) Removal of Excess Penetrant The removal of the excess penetrant from the surface of the workpiece is the next step in LP testing. Care must be taken to remove the penetrant which is deposited at the surface of the workpiece without affecting the upper region of the defect. A rough cleaning of the surface results in the failure of the distribution from the defect and normal surface. Therefore, some expertise is required to clean the surface for better resolution. The existence of the penetrant at the crack after the removal of the excess penetrant from the surface is shown in Fig.6.3.

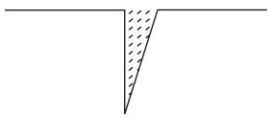


Fig. 6.3 *Workpiece-removal of Excess Penetrant*

(iv) Application of Developer It is one of the important steps to identify very clearly the defect on the workpiece. A thin layer of developer is applied over the surface after the removal of excess penetrant. The developer covers the surface with a colour which provides a good visual contrast to the penetrant. The developer without penetrant at the surface and the penetrant with developer at the defect site can be easily visualized based on the colour contrast. Therefore, the developer explores the opening of the crack edges very clearly, and hence, provides a good visibility of the crack for visual inspection of the defect. The enlarged width of the penetrant at the defect site due to the application of the developer is shown in Fig.6.4.

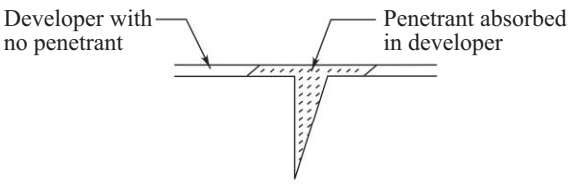


Fig. 6.4 *Workpiece-application of developer*

(v) Inspection and Evaluation The last step in LP testing is the inspection and evaluation. After allowing a minimum time to the developer for developing, the

The McGraw·Hill Companies

surface of the workpiece is scanned by means of white light or ultraviolet or laser light depending on the type of developer used. A white light is sufficient to examine the defect in case of visible dye-penetrant, while an ultraviolet light is required for a fluorescent penetrant. In order to increase the accuracy and reliability, the inspection process is automated by employing laser light and high quality optical sensors. The indication of the defect part in the workpiece after the absorption of the absorber into the penetrant is shown in Fig.6.5.

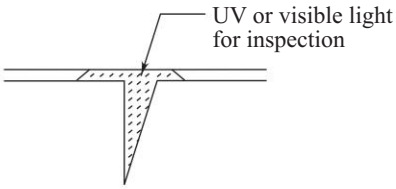


Fig. 6.5 *Workpiece-exploring the defect using light*

Care must be taken in examining the defect in the workpiece, since the unremoved drop of penetrant at the surface gives false information. Based on the inspection results, the acceptance and rejection of the workpiece/component can be made in comparison with the specifications and standards applicable for the LP testing.

6.4.3 Characteristics of the LP Testing Materials

In LP testing, the selection of various materials used for the testing is more essential. The materials such as penetrant, solvent, developer, etc. can be selected based on the characteristic features of the materials. The characteristic features of the different materials used in LP testing are listed below:

(i) Penetrant Some of the characteristic features required for the selection of penetrant for the LP testing are as follows:

(a) It should be chemically stable and inert.
(b) It should have low viscosity and high sensitivity to light.
(c) It should have higher wettability.
(d) It should be easily removable.
(e) It should be of low cost.

(ii) Developer For the proper identification of the defects, the selection of developer to the penetrant is more essential. Some of the important factors to be considered for the selection of developer are listed below.

(a) The developer must clearly differentiate the penetrant which is drawn into the defect when it is exposed to illumination.
(b) It should have high absorbing capacity.
(c) It may be a dry powder, aqueous (wet) powder suspension or water soluble, etc., with a low grain size.
(d) By employing several techniques such as spray, immersion, etc., it should be possible to form a thin uniform coating over the surface of the components.
(e) It should be easily wettable when added with penetrant.

The McGraw·Hill Companies

(f) It should give a high sensitivity when techniques like solvent spray, plastic film spray, etc., are used.
(g) It should give a low sensitivity when techniques such as dry immersion are used.

6.4.4 Different Testing Methods

One can select the appropriate method for the penetrant inspection based on the number of factors, such as testing condition, type of materials, etc. Generally, water-washable, post-emulsifiable and solvent-removable methods are the common methods employed. In case of water-washable method, the materials used, such as penetrant, developer, cleaner, etc., all are water soluble. The second method, namely, post-emulsifiable method is a combination of solvent- and water-based inspection. In this method, the solvent-based penetrant used during the process stage requires a water-washable path, which is different when compared with the first method. The solvent-removable method is an oil-based method. This technique is usually applied to small parts, pipelines, pressure vessel, etc. The dwell time for the penetrant as well as the developer depends on the type of materials and discontinuity. In general, the selection of the methods for a particular inspection is based on the sensitivity of that method. The sensitivity is the ability of the penetrant to be drawn into the defects and the removal of the penetrant from the surface based on the above discussion. In general, the order of inspection methods followed are post-emulsifiable fluorescent, solvent-removable fluorescent and water-washable fluorescent method.

6.4.5 Applications and Limitations

The applications of LP testing is wide in range. Some of the applications are given below:

(i) It is used to explore the defects, such as cracks, in products like pressure vessels, pipes, weld joints, turbine rotor discs, blades, etc.
(ii) It is used to explore the defects in automotive parts such as aluminium castings and forgings, piston heads, etc.
(iii) It is used for the detection of fatigue cracks which occur during service life of the materials.
(iv) It is used to test the bogey frames of railway locomotives, rolling stock, etc.
(v) It is used to test the electrical ceramic parts such as spark plug insulators, glass to metal seal, etc.
(vi) It is used to bend the plastic moulds.

The above applications are possible in view of the salient advantages of the LPT technique over the other available techniques, as listed below.

(i) It requires only a short time, irrespective of the shape, size and orientation of the defects, to inspect and evaluate the defects.
(ii) It requires only a penetrant, a developer and a light source, and does not require any electronic equipment.
(iii) It is applicable to test the components of any size, shape and materials.
(iv) It is a suitable technique for in-situ inspections.

The limitation of the LP testing is the applicability of this technique to nonporous material surfaces and metals.

6.5 RADIOGRAPHY

X-ray was discovered by W C Roentgen in the year 1895. Immediately after its discovery, the X-rays were used for industrial applications to examine objects/materials. The X-ray has emerged as one of the most widely used NDT techniques in industry as radiography. The radiography testing (RT) technique is used to detect the internal defects such as porosity and voids in materials. This technique is applicable over a wide range of components to test, ranging from micro size to a macro size such as plant components. The major advantage of this technique is that it does not require any surface preparation work as that required for other techniques. In the following section, the principle, operation, testing procedure, applications and limitations of the RT are discussed in brief.

6.5.1 Basic Principle

When X-rays or ionizing radiations are passed into the object or material, the penetrated rays form the image of the object on the other side. The image of the object is photographed by employing the conventional method. The process of obtaining the photography of the object is known as *radiograph*, and the study of the defect using radiograph is called as *radiography*. The principle of radiography is shown in Fg. 6.6.

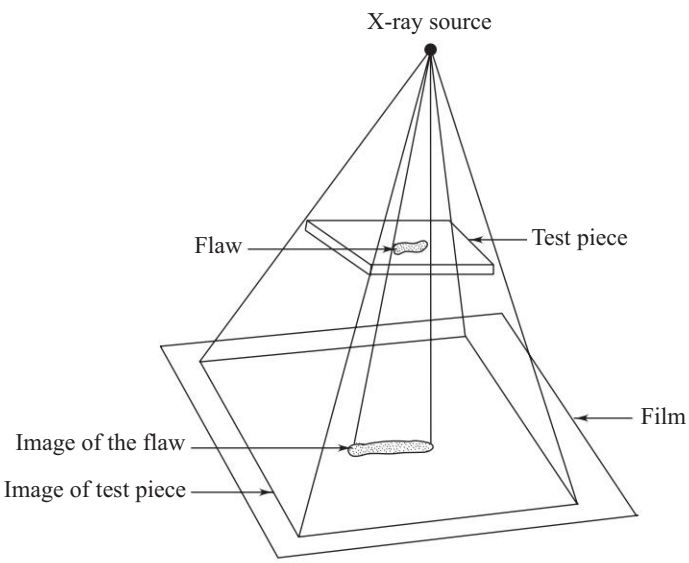


Fig. 6.6 *Radiography—Principle*

The intensity of the radiation after penetrating through the material varies due to the nature of the material. The presence of defects, such as voids or porosity within the material, absorbs more energy and hence, the intensity of the output radiation is reduced. The absence of any defects inside the material results in less absorption of energy and hence, the radiation emerges out with high intensity. Thus, the varying intensity of the output radiation is recorded on the photographic plate or film. The varying radiation intensity is recorded in the plate as dark and bright spots. The dark spot region corresponds to the highest intensity, while the bright spot region refers to the low or attenuated radiation.

The McGraw·Hill Companies

The spatial variations in the intensity of the recorded film or plates indicates the presence of defect. The observed image on the film is the shadow of the object and is known as *radiograph*. The basic experimental set up required for the radiography is shown in Fig. 6.6. It consists of an X-ray source, object/material and the photographic plate/film.

Characterisation of Defect The characterisation of flaw such as location, shape, size, etc. cannot be determined in a single exposure of X-rays on a film. One can determine the location, shape, etc. of the flaw by employing the *tube shift (TS) method* as shown in ig. 6.7.

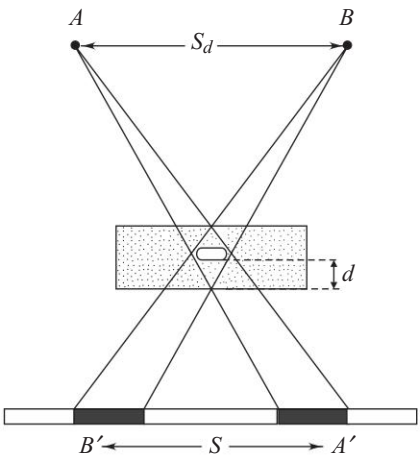


Fig. 6.7 *Tube shift method*

The X-ray source is first placed at the position A as shown in Fig. 6.7. The X-rays are passed into the object and the penetrated rays form the image at A' on the X-ray film. The exposure time is half the value of the normal exposure time. Then, the X-ray source is moved to the next position B which is at a distance S_d from the previous position A. As earlier, the object is exposed to half the exposure time. The image is recorded on the film at B'. Let the S be the distance between the images A' and B'. If f_d is the Film Focus Distance (FFD) which is the distance between the fixed spot of the X-ray film and the source, then the location of the flaw from the bottom surface of the object or material is givenas,

$$d = \frac{f_d S_d}{(S + S_d)} \tag{6.1}$$

The location of the flaw from the top surface can be determined by reversing the object/material and repeating the above procedure one more time. Thus, one can characterise the flaw by employing the tube shift method.

6.5.2 Types of Radiography

Radiography is classified into different types based on the type of ionisation radiation used for the identification of defects in the materials. The following are the some of the frequently used methods:

The McGraw·Hill Companies

(i) X-ray Radiography X-rays are electromagnetic waves which have the same physical properties as that of the visible light, radio waves, etc. The X-rays are produced by the bombardment of fast-moving cathode rays like electrons on a heavy target such as tungsten or molybdenum. The accelerated electrons strike the target material, about 99.8% of the electrons are wasted in heating the material. The remaining 0.2% of the electrons penetrate through the target material. The penetrating electrons produce two types of X-rays, namely, soft and hard X-rays. The soft X-rays will have only low penetrating power in materials, about 10^{-4} mm. The hard X-rays will have high penetrating power in materials, up to 500 mm in steel.

The X-rays are produced by a number of ways. A simple way of producing the X-rays is the Coolidge tube method. The important features of X-rays are intensity and quality. The intensity refers to the radiation power, while the quality refers to the penetrating power. In general, the energy of the X-rays used in radiography is 1 MeV or above. The various high energy X-ray sources are used in industries are, namely, resonance transformer, Van de Graff generators, betatrons, and linear electron accelerators.

(ii) Gamma Ray Source The physical properties of the gamma rays are similar to that of X-rays. The gamma rays are also electromagnetic waves and hence, the X-rays and gamma rays are indistinguishable from one another. However, the gamma rays differ only by the way of generation. The gamma rays are emitted by radioactive sources. These rays are known to exist in the similar wavelength range as that of the X-rays. Therefore, gamma rays are also used in industry for defect identification and are known as *gamma rays radiography*. The gamma rays are emitted as a part of the decay processes of the radioactive substances. The most popular nuclear radioactive substances which emit gamma rays are Cobalt 60 (Co-60), Iridium 192 (Ir-192), Cesium 137 (Cs-137) and Thulium 170 (Th-170).

(iii) Neutron Ray Source Generally, the energy of the neutrons ranges from a fraction of an eV to several MeV. The properties of the thermal neutrons and cold neutrons suit the requirements of radiography. The energy of the thermal neutrons lies between 0.01 and 0.3 eV, while that for the cold neutrons lies between 0 and 0.01 eV. In neutron radiography, the absorption is due to the interaction with the atomic nucleus and not with the electronic structure of the atoms.

6.5.3 Exposure Factor

In X-ray radiography, a standard rule known as the *exposure rule,* based on the reciprocity law, is adopted to determine the exposing time. According to this law, the exposure time is inversely proportional to the X-ray tube current for a given source to film distance. Let I be the X-ray tube current and t be the exposure time, then for a given source to film distance (L),

$$It = \text{constant} \tag{6.2}$$

One can obtain a satisfactory image on the X-ray film by adjusting the exposure time, the source to film distance and the tube current as a constant value known as the *exposure factor* for a given object/material, i.e.

$$\text{Exposure factor} = \frac{It}{L^2} = \text{constant} \tag{6.3}$$

The McGraw·Hill Companies

Exposure Time in Gamma Ray Radiography The experimental procedure for the Gamma ray radiography is same as that of the X-ray radiography. However, a different procedure is adopted for the determination of exposure time in gamma ray radiography. The exposure time depends on the factors, such as age of the source, thickness of the material and the distance between the sample and the film.

$$\text{Exposure time} = \frac{\text{Exposure factor}}{\text{source strength at present}} \tag{6.4}$$

The gamma rays are obtained from the radioactive product and their radioactivity or strength of the source decreases exponentially. Therefore, the age of the source has its role in determining the exposure time. Similarly, the thickness of the material and distance between the film and the source is also important.

6.5.4 Attenuation of Radiation in the Specimen

When X-rays or gamma rays or ionisation radiations are passed into the materials, it gets attenuated and hence, the intensity of the radiation decreases. The attenuation of the radiation in the material at a distance x from the surface is

$$I = I_0 \, e^{-\mu x} \tag{6.5}$$

where I_0 is the intensity of the incident radiation at $x = 0$, I the intensity of the emerging radiation out of the specimen and μ the linear absorption coefficient. The *half valve thickness* (HVT) is the thickness of the material which reduces the intensity of the X-ray into half of its initial value.

The observed linear absorption in the material can be explained based on the interactions of the photons with the atoms in the material. The interactions of the photons are due to the X-ray or gamma radiation. The important modes of the above interactions are, namely, *photoelectric effect*, *Rayleigh scattering*, *Compton scattering* and *pair production*. In photoelectric absorption, the electrons in the outermost shell of the atom absorb the energies from the incident X-ray photon. During Rayleigh scattering, the incident photon is scattered without any change in the internal energy of the scattering atom. During the above process, there is no release of the electrons. In *Compton scattering*, the recoil electron is produced by the absorption of atom due to the non-elastic scattering of X-ray photon. In *pair production*, the incident photon is disappeared due to the formation of electron–positron pair production. Therefore, the observed total *absorption* in the material is the sum of the above four processes.

6.5.5 Effect of Radiation on Film

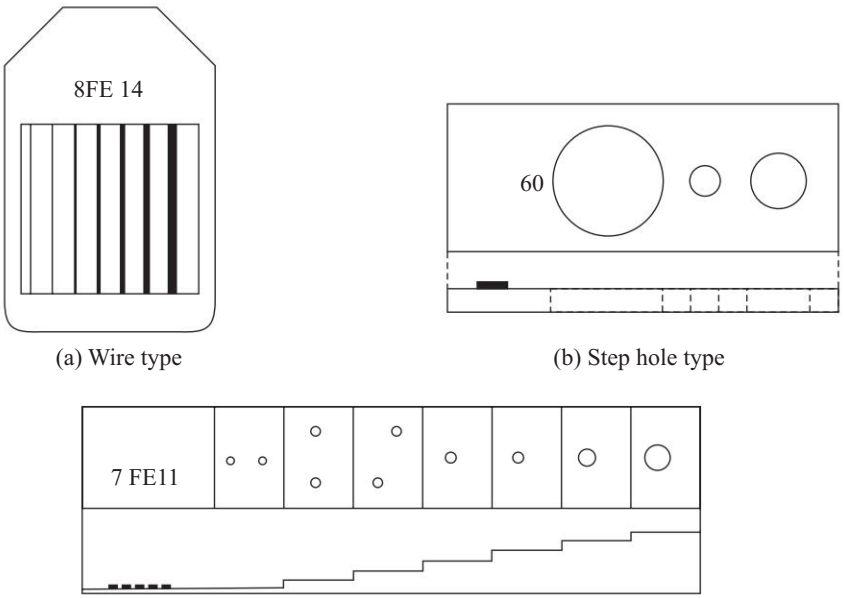
The ionisation of the film depends on the energy of the radiation incident on the film. Thus, a film with different sensitive powers to observe both low and high energy is required. The silver halide coated on the film will absorb some energy from the radiation incident on the film. Therefore, a part of the energy is released. These electrons will interact with the nearby silver halides and hence, give rise to an unsharpness at the boundaries. The unsharpness of the image is known as *inherent unsharpness*.

In view of the above reasons, the radiographic films are different from the conventional photographic films. They are different in film thickness, coating of emulsion on both sides, and sensitive to both low and high intensity radiations.

6.5.6 Image Quality Indicator and Sensitivity

In radiography, due to the large number of parameters such as film type, source to film distance, film density, etc., the wide range of sensitivity is determined by a device known as *penetrameter*. Now a days, it is called *Image Quality Indicator* (IQI), a measure of the sensitivity of radiography. The IQI is designed to explore the radiography quality in terms of sensitivity.

According to the British standard, the IQI are of two types, namely, wire IQI and step hole IQI, as shown in Fig. 6.8. In wire IQI, wires of the same materials with diameters ranging from 0.032 to 3.20 mm, respectively, as wire no. 1 to 21, with 30 mm length have been selected. A typical wire IQI which contains seven numbers of the above wire is shown in Fig. 6.8 a. The wire with 3.20 mm diameter is taken as the first wire and these wires are placed at an interval of 5 mm. Similar to the above set, there are four sets which suit materials such as steel, aluminium, copper, zinc and their alloys. Similarly, the step hole IQI and plug type are shown in Figs 6.8 b and c.



(a) Wire type (b) Step hole type

(c) Plaque type

Fig. 6.8 *Image standard indicator*

In the step hole IQI, the hole diameter is equal to the step thickness. The diameter increases from no. 1 (0.032 mm) to no. 18 (1.6 mm) as shown in Fig. 6.8 b.

$$\text{Wire IQI sensitivity} = \frac{\text{Diameter of the smallest discernible wire}}{\text{Thickness of the specimen under wire}} \times 100\% \qquad (6.6)$$

The McGraw·Hill Companies

Generally, the sensitivity is quantified by employing IQI. The percentage of sensitivity for different thickness are given in Table 6.1.

Table 6.1 *Sensitivity for different metal thickness*

Test metal thickness	*Percentage of Sensitivity*	
(mm)	*Wire IQI (%)*	*Step hole IQI (%)*
12	1.4	3
50	1.0	1.8
150	0.7	1.3

6.5.7 Radiographic Imaging

The analysis of the radiography image is more important to evaluate the defects in the materials. The observed image difference in intensities reflects the nature of these materials. The following are some of the points to be considered while doing the RT imaging:

(i) The radiation should pass through the object.
(ii) The ideal source is a point source.
(iii) The distance between the source and the object should be very large.
(iv) The film should be in close contact with the object.
(v) The film which contains extra fine grains, high contrast with slow or medium or high speed depending on the requirements.
(vi) Heavier metal screens are ideal to produce high intensification.
(vii) The fluorescent screens are not suitable as it produces unsharpness.
(viii) The film density should be in the range of 1.5–3.3.
(ix) The sensitivity of the film should reveal the smallest discontinuity.
(x) The IQI or penetrameter is used to establish the quality level of the RT inspection, by employing standard materials.

6.5.8 Techniques

RT has different testing techniques; one has to select the appropriate technique based on the requirements such as geometry, size, sensitivity requirements, in-situ space availability, etc. The following are some of the general techniques available for the RT inspection in industries:

(i) Single wall single image
(ii) Double wall penetration
(iii) Double film, and
(iv) Special technique

Let us discuss the above techniques in brief.

(i) Single Wall-Single Image Technique This technique is applicable in case when the material to be inspected is in the form of pipes. There are two ways to do the inspection in this case. One way is to keep the source(S) outside the pipe and the film (F) inside, as shown in Fig. 6.9. The other way is that to keep the source inside the pipe and the film around the weld, as shown in Fig. 6.10. Let t be the thinkness of the pipe to be inspected.

The density of the film at the centre is bright since the thickness of the weld penetrated at the centre is less. The film density decreases while measuring from the centre to edges due to the increase in the thicknes of the weld penetrated.

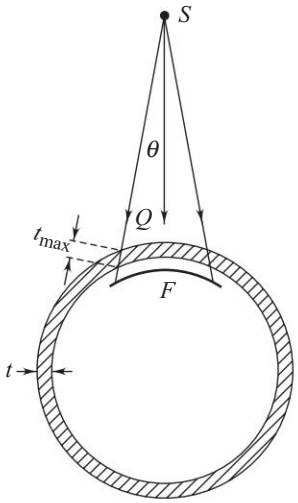
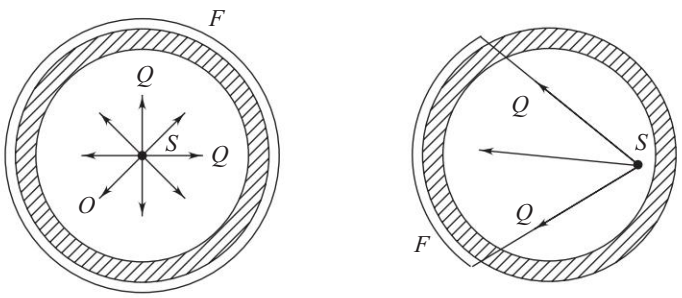


Fig. 6.9 *Single wall-single image technique—Source outside*



(a) Centre (b) Off centre

Fig. 6.10 *Single wall—Single-image technique—Source inside*

In the second method, the source can be fixed in two ways, namely, at the centre and off centre, as shown in Fig. 6.10. When the source is at the centre and if it emits radiation in all directions, the total circumferential weld is exposed at one exposure (Fig. 6.10 a). On the other hand, if the source is off centre (Fig. 6.10 b), it requires more energy to penetrate the weld thickness and hence, it requires two or three exposures.

(ii) Double-Wall Penetration Technique In this method, both the sources and the film are kept outside the pipe. There are three ways of doing inspection based on the prevailing pipe diameter and site restrictions double wall-single image, double wall-double image and double wall superimposing image.

(a) Double-Wall single Image In this method, the source (*S*) is kept very close to the outside surface of the pipe, as shown in Fig. 6.11. The film (*F*) is kept on the other outer

The McGraw·Hill Companies

side of the pipe. The radiation penetrates into the walls, and then the image of the weld is formed on the film if width NN^1. In this technique, only the film side weld is exposed. The technique requires a number of exposures to cover the whole length of the weld. Further, the source is at the point *M* which is far away from the film and hence, the image is extremely poor in sharpness. In this technique, due to the extra absorption of the radiation in comparison with the single-wall method, the ensitivity is very poor.

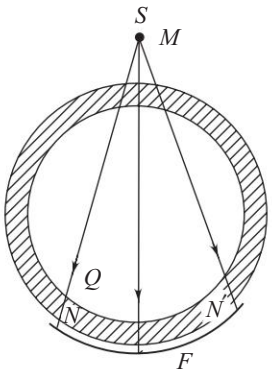


Fig. 6.11 *Double-wall single image technique*

(b) Double-Wall Double Image This technique is applicable for only small outer diameter pipes, up to 90 mm. The diameter of the source is very small. The source is kept at a distance larger than the Source to Film Distance (SFD). The above arrangement removes the superimposing of the source side weld over the film side weld and forms the elliptical image on the film. The portion of the source and the film in the double wall-double image is shown in Fig. 6.12. One can image both the film and the source side weld. It requires minimum of two exposures each perpendicular to obtain the image of the weld on the entire circumstances.

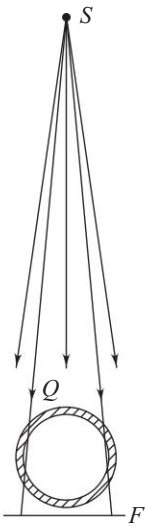


Fig. 6.12 *Double wall-double image technique*

The McGraw·Hill Companies

(c) Superimposing In this method, the source (*S*) is kept at the SFD distance without doing the offset. The film (*F*) is placed on the other outerside of the pipe as shown in Fig. 6.13. Therefore, the source side weld is superimposed on the film side weld. This technique is applicable only in case when the double image is not applicable. To cover the entire weld, it requires a minimum of three exposures each at an angle of 120º. The arrangement of the source and the film in the superimposing technique is shown in Fig. 6.13.

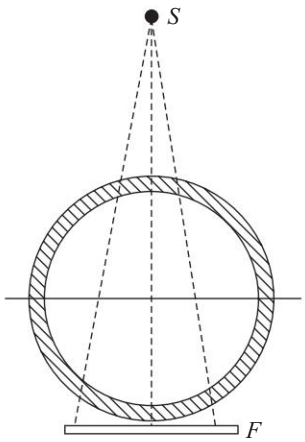


Fig. 6.13 *Superimposing technique*

(iii) Double-film Technique Due to the high and low contrast of the film, there is a restriction in the material thickness. At the same time, it is better to have less number of exposure time. One way of studying both increase in specimen thickness and reducing the number of exposure is to use two films with different speeds. The application of two different films and one exposure time results in the summing of the latitudes of the image. This technique is known as *double-film technique*. The arrangement of the films in the double film-technique is shown in Fig. 6.14. One can also obtain images using this technique without employing the lead screen.

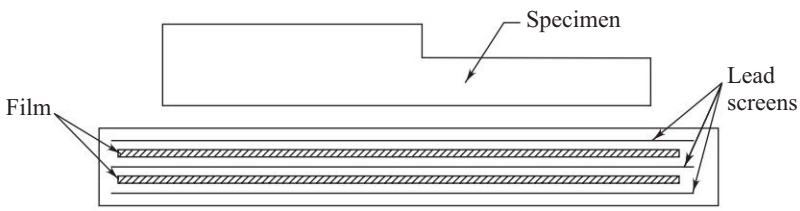


Fig. 6.14 *Double film technique*

(iv) Special Technique The special technique used to image the complex point is known as *penetration technique*. This technique is further classified into multi-wall single and double-image techniques. The multi wall single image technique is used to image the defects when the outer diameter of the double wall envelope pipe is more than 90 mm. This technique requires a number of exposures to cover

The McGraw·Hill Companies

the overall weld length. The experimental arrangement used in this technque is shown in Fig. 6.15.

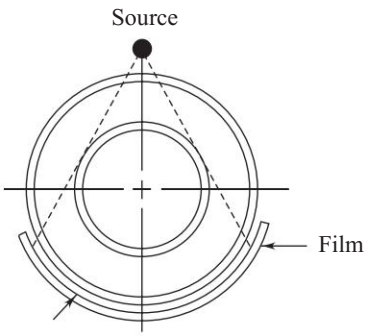


Fig. 6.15 *Multiwall-single-image technique*

When the outer diameter of the double wall envelope pipe is less than 90 mm, the multiwall-double image technique is used to image the defects. It requires a minimum of three exposures, each at an angle of 120º. The experimental arrangement used in this mehod is shown in Fig. 6.16.

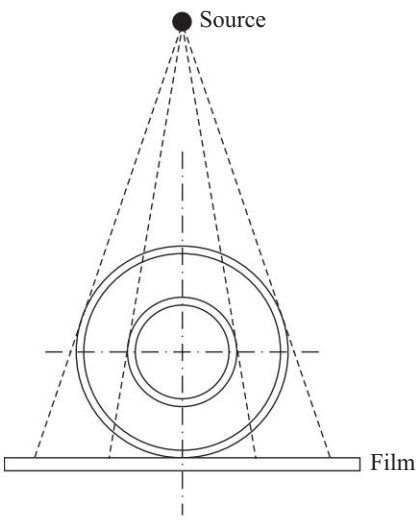


Fig. 6.16 *Multi-wall-double image technique*

6.5.9 Fluoroscopy or Real-Time Radiography

In X-ray radiography, the radiation after passing through the material is recorded in the photographic film. The X-ray radiation is converted into light by using a screen in the place of the film, known as *fluorescent screen*. The process of obtaining the image of the object by converting the X-rays into light on the fluorescent screen is known as *fluoroscopy*. Now-a-days, the above technique is renamed as *Real Time Radiography* (RTR).

The McGraw·Hill Companies

The image obtained using the fluoroscopy method is usually faint and hence, the sensitivity of the flaw characterisation is very poor. In order to improve the brightness of the image so as to enable to view the image under normal light and increase the sensitivity of the flaw characterisation, a Closed Circuit Television (CCTV) camera i used as shown in Fig. 6.17.

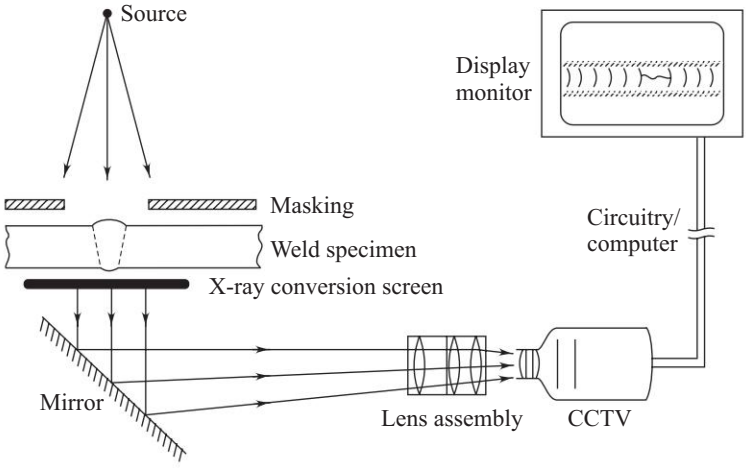


Fig. 6.17 *Group-I—Television fluoroscopic system*

The arrangement of the X-ray source, materials under test, fluorescent screen, CCTV and other computer supports used in the group-I television fluoroscopic system is shown in Fig. 6.17. The image is received by the CCTV and is then amplified using the amplification circuit associated with camera. The amplified signal is then processed by the computer. The processed image is then displayed on the monitor. The display monitor is operated using a remote from the X-ray source and hence, the radiation hazards could be eliminated. However, the sensitivity of the light of the image is very small. Therefore, it requires an image intenifier as shown in Fig. 6.18.

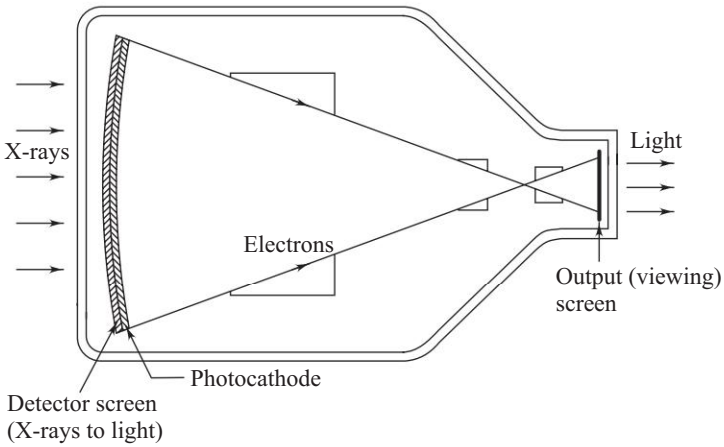


Fig. 6.18 *X-ray image intensifier*

The McGraw·Hill Companies

One can obtain the high intensity image, by employing the image intensifier. The image intensifier consists of a double-layered screen with one side to receive the X-rays, and then convert them into electrons. The image intensifier is arranged in such a way that the detector is facing the materials. The electrons emerging from the double layer screen is accelerated and then focused to fall on the output screen, namely, the viewing screen. The acceleration of the electrons leads to an image intensification of about 100 to 1000 times. The image of the object is viewed through the viewing screen. In order to avoid the radiation hazards, i.e., for radiation-free remote operation, the image is viewed by employing a CCTV camera as shown in Fig. 6.19.

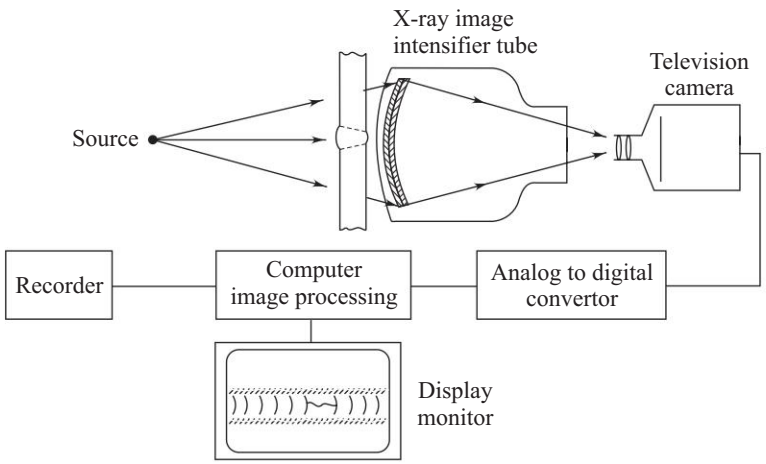


Fig. 6.19 *Group-II—Television fluoroscopic system*

The sensitivity of the image obtained on the monitor after the image processing, shown in Fig. 6.19, is very high. The increased density is adequate for any type of defect identification. The output is available either as display or recorder. Generally, the group-II fluoroscopic method is used for weld inspections (light alloy castings), composite studies, and for some medium speed-line radiographic applications. The limitations in the industrial applications are due to the unsharpness in the image observed, which is due to the polycrystalline nature of the sources. In recent years, due to the technological development such as digital imaging, increased pixels (a very small picture elements) and projection magnifications, this technique has been widely employed in industries.

The slightly modified version of the group I and group II X-ray fluoroscopy is the group III television fluoroscopic system. The schematic representation of the system is shown in Fig. 6.20.

In this system, a small group of array of detectors are used as the primary X-ray absorber/converter. A total of 600 X-ray sensitive elements are present in the 3 mm × 1 mm array. Each array is made up of photodiode which is covered by a fluorescent material. There are two ways of operating this equipment. In the first method, the material is fixed, the detector array is moved along the material. While in the other method, the detector array is fixed, the material to be tested is moved

on a travelling belt. The second type of testing is employed in aiports for luggage searching.

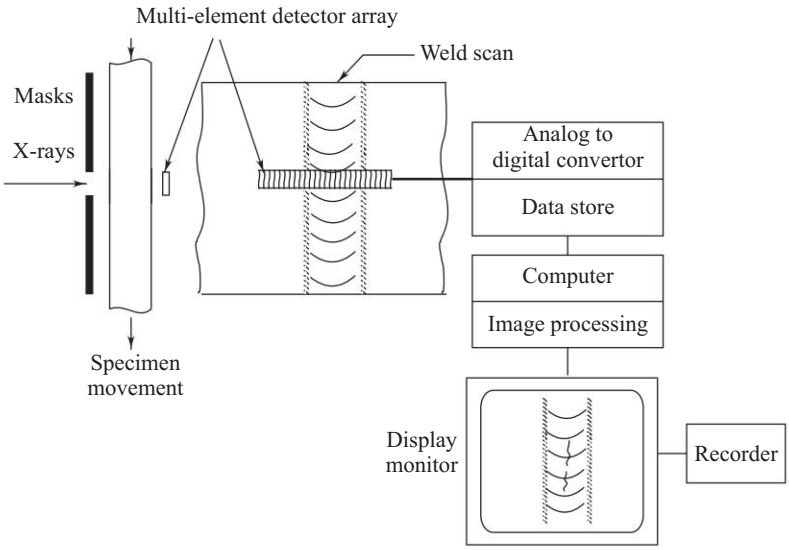


Fig. 6.20 *Group-III—Television fluoroscopic system*

6.5.10 Comparison of X-ray and Fluoroscopy Radiography

A comparison between the X-ray and fluoroscopy radiography is given in Table 6.2.

Table 6.2 *Comparison between X-ray and fluoroscopy radiography*

	X-ray	*Fluoroscopy*
1.	X-rays are recorded with X-ray film.	X-ray energy is converted into visible light and hence, displayed on the screen.
2.	The intensity and sensitivity are high.	The intensity and sensitivity are moderate.
3.	Objects/materials of any size with regular shape can be tested.	Objects/materials of small size cannot be tested.
4.	Radiation hazards will exist.	It is free from radiation hazards
5.	Remote operation is not possible.	Remote operation is possible.
6.	It is a slow processing technique, and hence, immediate inspection is not possible.	It is a faster processing technique, and hence, immediate inspection is possible.

6.5.11 Applications

The applications of the X-rays in industries are enormous. Some of the important applications are given below:

(i) It is applicable for a wide range of materials such as ferrous, non-ferrous alloys, non-metallic materials and composites.

The McGraw·Hill Companies

(ii) It is used for the inspection of pipe welding.
(iii) It is used for the inspection of casting and forgings.
(iv) It is used to detect the cracks, broken wires, unsoldered connections, foreign materials, etc. in semiconductor devices.

6.5.12 Limitations

The following are the some of the limitations of radiographic techniques:

(i) It fails to detect the flaw, when it is not parallel to the direction of the radiation beam.
(ii) It fails to detect the minute cracks such as inclusion in wrought material, flakes, microporosity, etc.
(iii) It is not possible to inspect the irregular geometry materials.
(iv) It requires safety from hazards both for the testing persons and the vicinity area of the inspection site.
(v) The cost of the equipment is very high.

6.6 ULTRASONIC TESTING

Ultrasonic Testing (UT) is one of the versatile NDT methods. UT is used to inspect and characterise the defects/flaws in metallic and non-metallic materials. UT inspection uses the short wavelengths and high frequency ultrasonic sound waves. One can test the surface defects and sub-surface defects, such as discontinuity, using this method. This method utilises the one side accessibility for the accurate evaluation and characterisation of flaws. The possible solution for a wide range of frequencies and different modes of ultrasound waves extended the wide spread applications in the range of materials, plant components and systems. Generally, in industries, the UT is used to detect the flaws and the thickness measurements.

In the following section, the basic principle, different modes of ultrasonic sound waves, different types of testing techniques, simple flaw detectors and different modes of displays are given in brief:

6.6.1 Basic Principle

The principle behind the UT inspection is the transmission of ultrasonic sound waves with the medium and the reflection or scattering at any surface or internal discontinuity in the medium. An ultrasonic wave is a sound wave whose frequency is above 20 kHz. The ultrasound waves or mechanical vibrations are propagated into the materials (solids or liquids or gases). During the propagation of the ultrasound, the particle in the medium vibrates and hence, the energy is transferred from one atom to another atom in the specimen. When these sound waves are propagated into the specimen, they are reflected back or scattered at any surface or internal discontinuity in the materials due to the change in the acoustic impedance.

The discontinuity means the existence of flaw or defect or crack or hole, etc. in the material. The reflected or scattered sound waves are received and amplified and hence, the defects in the specimen are suitably characterised.

The ultrasonic waves have all the properties as that of the sound waves such as highly directional, undergoes reflection, scattering, etc. The ultrasound waves are

transmitted to any medium such as solid, liquid and gas. The following are the different modes of the propagation of the ultrasonic waves:

(i) Longitudinal or transverse waves
(ii) Shear waves or compressional waves
(iii) Surface or Rayleigh waves

In longitudinal ultrasonic waves, the particle in the medium vibrates parallel to the direction of the propagation. In shear ultrasonic waves, the particle in the medium vibrates perpendicular to the direction of the propagation. The velocity of the shear ultrasonic wave is nearly 50% as that of the longitudinal wave. The waves which can travel only at the surface of the solid medium are known as *surface waves or Rayleigh waves*. The surface waves can travel only at the surface in the order of a millimeter depth.

6.6.2 Transducer, Couplant and Inspection Standards

The ultrasonic waves are normally generated and received for inspection applications by the transducer which is working based on the *piezoelectric effect*. The transducer consists of a quartz crystal, which generates mechanical vibrations and hence produces the ultrasonic waves, when a potential difference is applied across the crystal. The crystal is mounted on a metal case, where the front surface is open for the generation of ultrasonic waves with protection layer, while the back surface has absorbing materials which absorb the ultrasonic waves generated at the back side. The cross sectional view of the transducer is shown in Fig. 6.21.

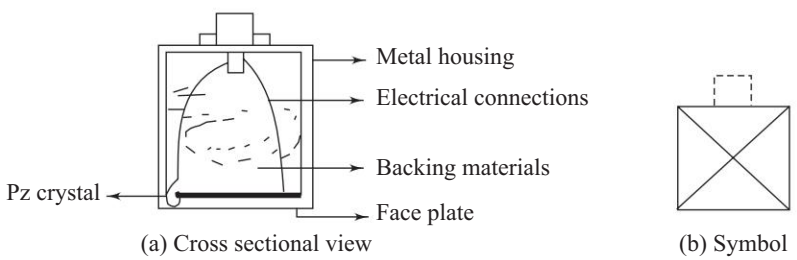


Fig. 6.21 *Transducer*

In order to provide a good acoustic coupling between the transducer and the test specimen, UT requires good couplant. The couplant also eliminates the air gap between the transducer and the test specimen. The couplants which are generally used for UT are of two types, namely, high viscosity and low viscosity. Some of the examples are oil, water, silicone grease, etc.

The ultrasonic equipment, and the probes are normally calibrated before any test using reference and calibration blocks. As per the recommendation of the International Institute of Welding (IIW), three different types of reference blocks are used for the calibration. They are area amplitude blocks, distance amplitude blocks and test blocks.

6.6.3 Inspection Methods

The UT methods are broadly classified based on the nature of propagation of the ultrasonic waves into the medium. When an ultrasonic wave propagates between

The McGraw·Hill Companies

two media, one part gets reflected and another part gets transmitted into the medium. Similarly, when the ultrasonic waves propagate in a medium, whenever there is a change in the acoustic impedance, a part of the wave gets reflected completely or transmitted into the medium, depending on the nature of the defects.

In UT, three methods, namely, through transmission, pulse echo and resonance methods are used. In pulse echo method, the reflected ultrasonic waves are used to identify and characterise the flaw. On the other hand, in the through transmission method, the transmitted beam will be used to identify and charactrise the flaw. The above three methods are normally employed in industries for UT studies. Let us discuss the above methods in brief.

(i) Through Transmission Method The through transmission method consists of two probes. The first probe generates the ultrasonic signals and is known as *transmitter*. The second probe receives the signal and is known as *receiver*. The position of the transmitter and the receiver probes in through transmission method is shown in Fig. 6.22. Let *d* be the thickness of the material. Consider a defect in the material,at a depth of *h* from the top surface.

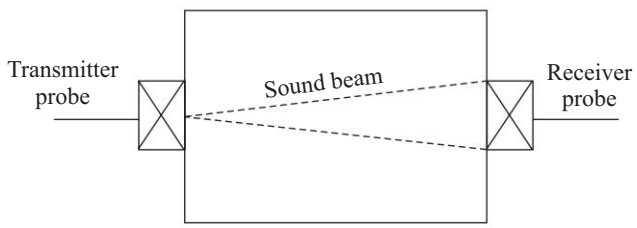


Fig. 6.22 *Through transmission method*

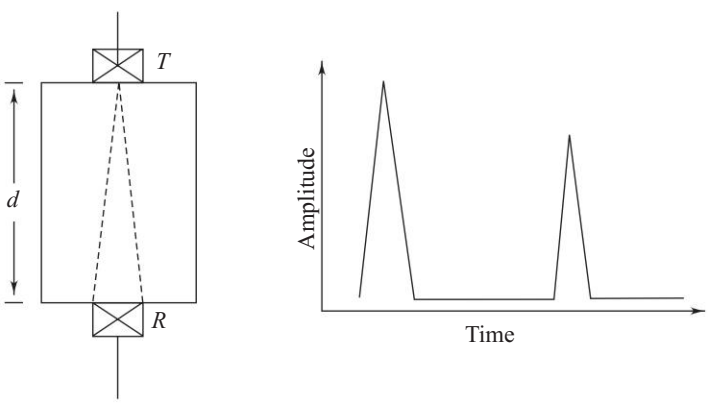


Fig. 6.23 *Through transmission—CRT screen appearance for defect free specimen*

When there is no defect in the specimen, the time domain trace obtained in the CRT is shown in Fig. 6.23. On the other hand, if there is a defect at a depth *h*, the time domain trace obtained in the CRT screen is shown in Fig. 6.24. The amplitude of the second echo reflects the nature of the defect. When the ultrasonic waves are completely reflected by the defect (large defect), the transmitted signal is completely

lost. Therefore, the receiver does not receive any signal and the corresponding tie domain trace is shown in Fig. 6.25.

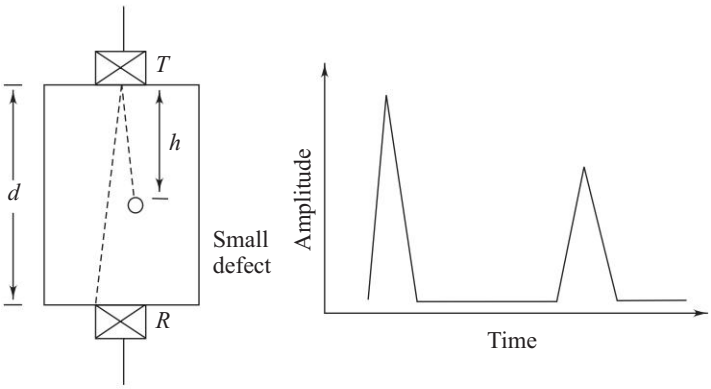


Fig. 6.24 *Through transmission—CRT scren appearance for small defect sample*

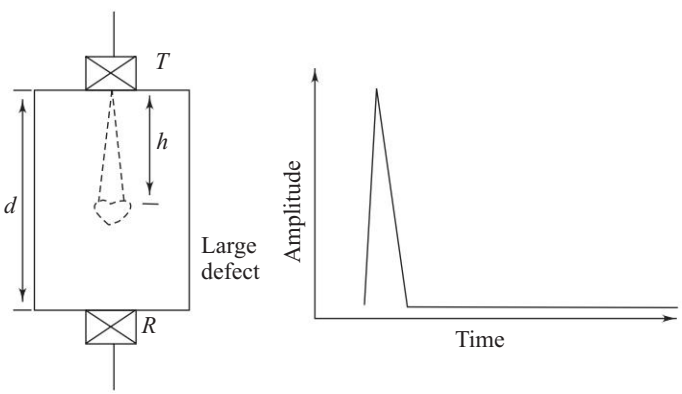


Fig. 6.25 *Through transmission—CRT screen appearance for large defect sample*

The main advantage of this method is its ability to inspect the defects such as large ingots and castings in high attenuating materials. This method provides only qualitative assessment of the defects, and not the actual size and location. This method requires a good couplant to provide good coupling between transducer and specimen, and hence good time domain traces.

(ii) Pulse Echo Method In UT, the Pulse Echo (PE) method is the most popular and commonly employed method. A single probe is used for the transmission and reception of the ultrasonic signal. A schematic representation of the PE method is shown in Fig. 6.26. In order to identify the defects, the CRT screen is first calibrated between the initial signal and the time of arrival of the first back wall signal. When there is no defect in the specimen, the transmitted and back wall echo will appear as shown in Fig. 6.26 a. On the other hand, if there is defect at a depth *h* from the top surface of the specimen, a part of the ultrasonic wave gets reflected from the defect

The McGraw·Hill Companies

and a part at the back of the specimen. The time domain traces obtained in this case is shown in Fig. 6.26 b. The reflected signals from the defect appear in between the transmitted pulse and back wall echo. By measuring the distance of the defect echo from the initial pulse, the depth of the defect is determined. When the defect is larger, the entire signal is reflected by the defect, and hence the back wall echo disappears. The observed time domain traces is shown in Fig. 6.26 c.

The PE method provides better results than other testing methods. This method is used to identify and characterise the flaw. In this method, a single transducer is used for the transmission and reception of the signal. A single side accessibility is sufficint for the identification of the flaw.

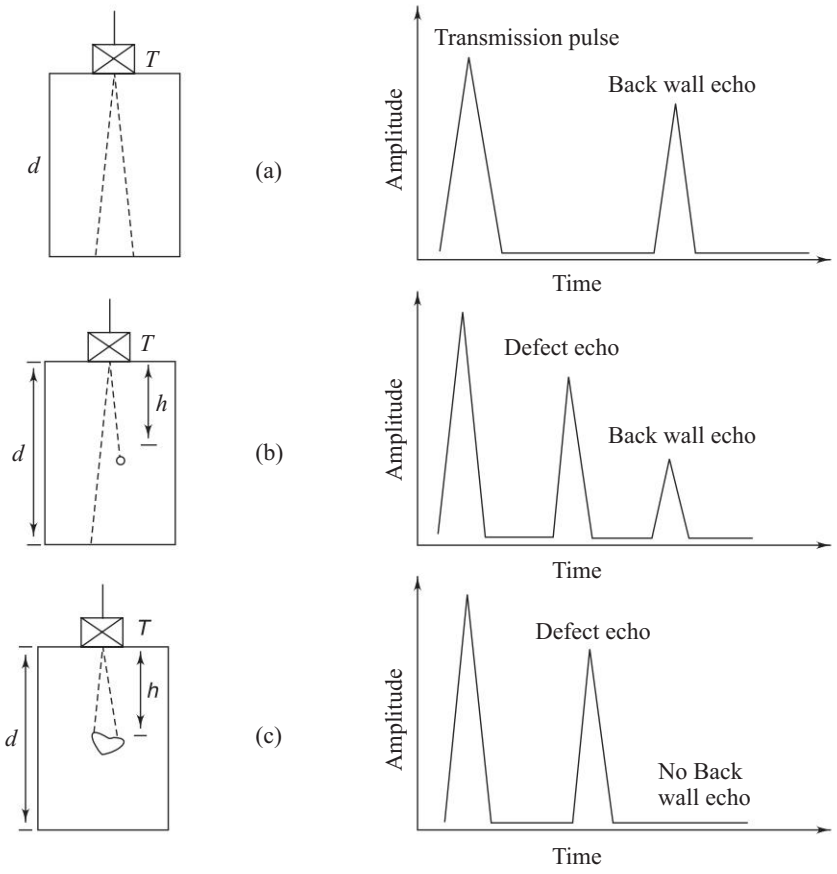


Fig. 6.26 *Principle of pulse echo: (a) Defect-free specimen, (b) Specimen with small defect, and (c) Specimen with large defect*

(iii) Resonance Method Generally, in any materials one can obtain the resonance condition when the thickness of a material is equal to half of the wavelength of sound or any multiple thereof. We know that the wavelength of the ultrasonic wave depends on the frequency. Therefore, one can create the condition of resonance for the thickness of the plate under test by varying the frequency of the ultrasonic waves. This condition

of resonance is easily recognised by the increase of received pulse amplitude. The thicknes can be obtained from the relation

$$t = \frac{v}{2f} \tag{6.7}$$

where v is the ultrasonic velocity in the specimen and f the fundamental frequency of the resonance.

6.6.4 Pulse Echo Testing Technique

Generally, the UT are classified into two types, namely, contact and immersion type. In contact type testing, the probe is placed in direct contact with the test specimen. In this method, a couplant is used between the probe and the test specimen for better transmission of ultrasonic waves into the test specimen. On the other hand, in case of immersion type testing, a waterproof probe is used. The probe is immersed in water at a distance from the test specimen. Thus, the ultrasonic beam is transmitted into the test material through a water path or water column.

(i) Contact Type Technique The contact type technique is further divided into three types: (a) normal beam, (b) angle beam and (c) surface wave technique. Let us discuss the principle and operation of the above techniques as follows.

(a) Normal Beam Technique In normal beam technique, the ultrasonic beam is passed into the test specimen in a perpendicular direction. One can use either a normal beam single or double TR probe for testing. In single probe method, the transmission of the ultrasonic signals and reflection of the signals from the defect in the test specimen is as shown in Fig. 6.27. This technique is not effective for the detection of near-surface flaws and for thin wall thickness measurements due to the occurrence of a large transmission pulse.

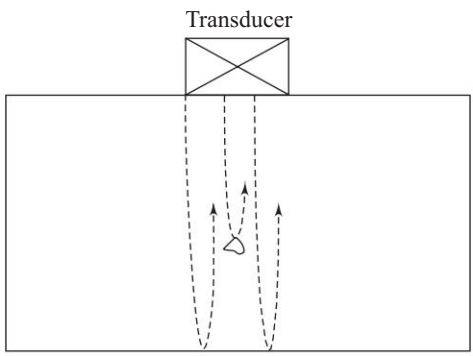


Fig. 6.27 *Pulse echo technique—Single transducer*

In dual or double normal beam technique, one probe is used for the transmission of the signal into the test specimen, while the other receives the echoes from the flaws and back wall as shown in Fig. 6.28. This technique is more useful for specimens with irregular shapes and a non-parallelism between the front and back surface of the specimen.

The McGraw·Hill Companies

In order to measure the thin wall thickness and near-surface flaw detection, one has to eliminate the transmission pulse. It is achieved by casing the two transducers in a same case known as *transmission-reception (TR) probes*. This technque is known as TR normal beam technique.

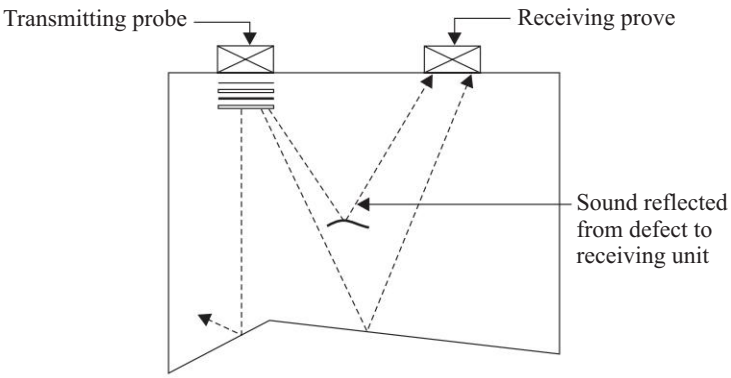


Fig. 6.28 *Pulse echo technique—Double transducer*

(b) Angle Beam Technique In angle beam testing, the ultrasonic waves are transmitted to the test specimen at a predetermined angle from the test surface. Generally, the transverse wave probes are used in angle beam testing. However, the modes of waves transmitted to the test specimen are either mixed longitudinal and transverse or transverse waves only. The defects which are not detected using normal beam technique are detected employing transverse waves with angle of refraction between 35 and 80º.

(ii) Immersion Testing Technique Immersion testing technique is further divided into three types, namely, immersion, bubbler or squirter and wheel transducer technique. These techniques are normally used in the laboratory level of testing. In this technique, a uniform couplant condition can be achieved easily, which is an added advantage. One can generate longitudinal or transverse waves at different angles by changing the angle of the incident beam. Let us discss the working of the immersion technique.

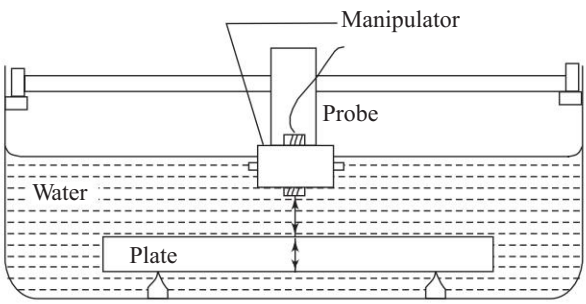


Fig. 6.29 *Immersion technique—Longitudinal wave generation by a normal beam technique*

In the immersion technique, both the probe and the test specimen are immersed in water. The ultrasonic beam, generated by the probe, is directed through the

water into the test specimen. The probe can be fixed at a normal beam technique (Fig. 6.29), for generating longitudinal waves. Similarly, for generating transverse waves, the probe will be fixed at an angle as shown in Fig. 6.30.

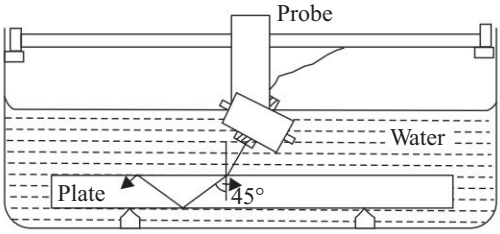


Fig. 6.30 *Immersion technique—Transverse wave generation by an angle beam technique*

6.6.5 Flaw Detector

Flaw detector is an electronically controlled unit. The operation of the flaw detectors are controlled by the testing person. In ultrasonic testing, the ultrasonic signals are generated by the means of burst of energy with time intervals. The generated ultrasonic pulse is passed into the specimen and gets reflected either at the flaw or at the bottom/other end of the specimen. The reflected signal is received and amplified, and then connected into the CRT screen. The CRT screen displays the time domain traces of the signal or other modes of presentation, namely, B-scans or C-scans. The block diagram of a simple pulse echo flaw detector is shown in Fig. 6.31.

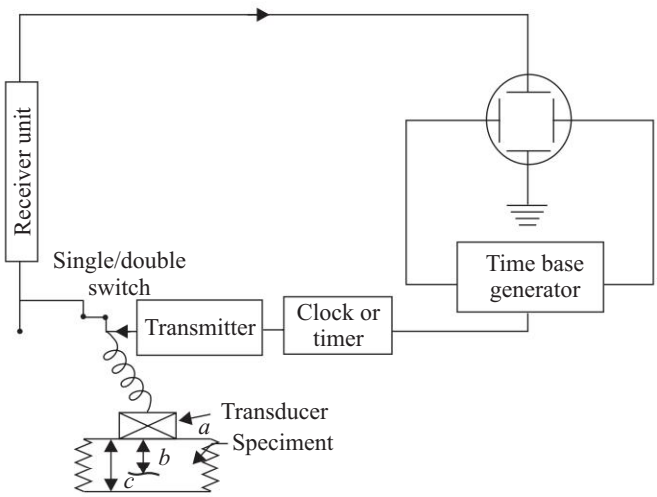


Fig. 6.31 *Block diagram of flaw detector*

It consists of an ultrasonic receiver, time base generator, transmitter, testing probe and CRT screen. A short electrical voltage is applied to the transducer from the transmitter. The transducer generates the ultrasonic signals and then, it is passed into

The McGraw·Hill Companies

the specimen. A part of the signal is reflected at the surface of the specimen *a*, a part of the signal gets reflected at the flaw, if there is a flaw b, and a part at the other end of the specimen *b* as shown in Fig. 6.31.

The received signal by the transducer is sent to the receiver unit which consists of an amplifier, a rectifier and an attenuator. The amplifier is used to amplify the signal with controlled pulse width. The a.c. voltage signal is rectified by the rectifier. The amplitude of the signal is varied using the attenuator and is controlled in dB. The clock or timer triggers the time base generator and the transmitter simultaneously. This facilitates the ultrasonic pulse from the transducer and an electronic spot on the CRT screen. The amplified signal by the receiver unit is fed to the CRT screen and it displays the front surface also as echo *a*, the reflected waves from the flaw as echo *b* and the back surface reflection as echo *c*. The time domain trace obtained in the abve process is represented in the Fig. 6.32.

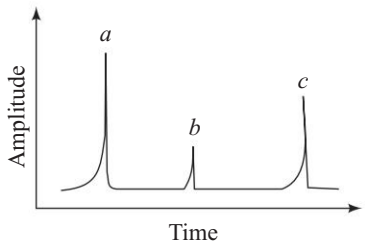


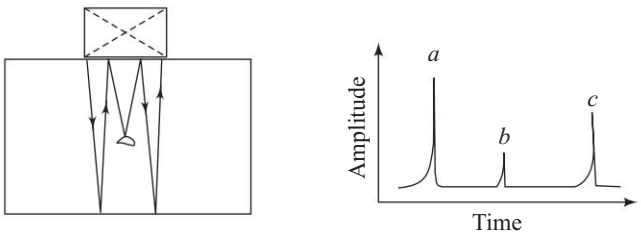
Fig. 6.32 *Time domain trace*

By measuring the amplitude of the echo *b* and the distance of the echo *b* from the end *a*, one can determine the depth of the flaw from the surface *a* of the specimen. The flaw detector also has the facilities to vary the frequency of generation of ultrasonic waves, pulse repetition rate, etc., depending on the type of specimen to be tested. One can also use the through transmission technique for the flaw characterisation using the flaw detector by selecting the appropriate mode in the flaw detector.

6.6.6 Different Types of Scans

In UT, the ultrasonic signals are transferred into the CRT screen to obtain the image of the defects. The presentation of the image of defects in the CRT screen are of different types, namely, *A*-, *B*- and *C*-scan.

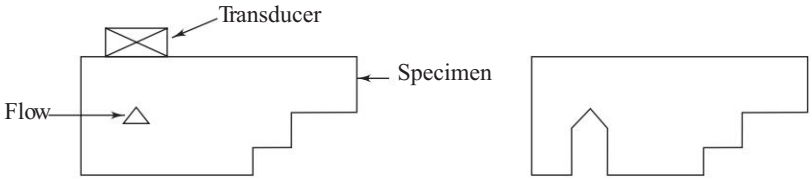
(i) A-scan The most commonly used presentation in UT is the *A*-scan. The time of flight of the pulse is represented in the *x*-axis and the amplitude of the echoes are represented in the *y*-axis. The schematic representation of the *A*-scan presentation of the test specimen is shown in Fig. 6.33. When the defect is absent, there is no flaw in the specimen. Therefore, the distance between the first echo and back wall echo gives the thickness of the specimen. The distance between the first echo and echo due to the defect, locates the depth of the flaw. The information obtained from *A*-scan on flaw is only one-dimensional, and hence it requires some mathematical calculation for the characterisation of the flaw. The *A*-scan display is used in a number of applications such as flaw characterisation, thicknes gauging, sound velocity measurements, etc.



(a) Probe and sample (b) *A*-scan display

Fig. 6.33 *Typical A-scan display*

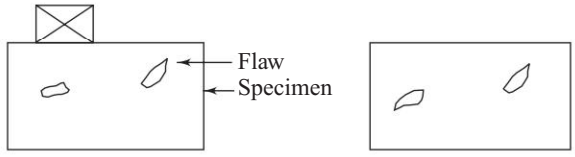
(ii) B-scan In *B*-scan presentation, the dimension of the specimen is displayed in the *x*-axis, while the travel of the ultrasonic pulse in the specimen is displayed in the *y*-axis. The amplitude of the ultrasonic pulse is obtained using the modulation. Thus, the *B*-scan presentation gives the cross-sectional view of the specimen. Therefore, one can determine the length and depth of the flaw in the specimen. A typical *B*-scan display of the specimen with defect is shown in Fig. 6.34. The probe is placed on the test specimen and the reflected ultrasonic pulse is received, and then it is moved over the surface of the specimen. The portion of the probe on the test specimen is shown in Fig. 6.34 a. The received pulse on the CRT screen display shows the cros-sectional view of the specimen Fig. 6.34 b.



(a) Probe and sample (b) *B*-scan display

Fig. 6.34 *Typical B-scan display*

(iii) C-scan *C*-scan display is the best method to display the flaw depth. In order to determine the accurate depth of the flaw, the probe is moved on the test specimen. The echo of the scan output produces a shaded line with varying intensity which indicates the absence of flaw. The echo of the scan output without any shading indicates the flaw shape. Typical *C*-scan display is shown in Fig. 6.35.



(a) Probe and sample (b) *C*-scan display

Fig. 6.35 *Typical C-scan display*

The McGraw·Hill Companies

6.6.7 Applications

Ultrasonic testing is a versatile technology for the NDT and NDE of defect identifications. The ultrasonic testing has been applied in most stages for quality control and material inspection. The most important application of UT is the weld inspection. The UT can be carried out over wide range of components/systems starting from simple shape little bars/rods to very large objects such as forgings. The general on-line inspection of UT in industries are production of steel, non-ferrous metal, structure of the airframes, pressure vessels, ships, bridges, motor vehicles, jet engines, etc. UT is used for on-line inspection in industries such as rail rods, fatigue cracks and effects of corrosion equipment, etc. The UT is also used to monitor the fatigue cracks in welds and corrosions of the power plants, boiler components, etc.

Some of the advantages of UT compared to the other techniques are as follows:

(i) Extremely small size flaws are detected due to the high sensitivity and reliability.
(ii) Extended to all types of materials due to its wide range of frequency and mode selections.
(iii) One side accessibility is sufficient for accurate evaluation of the defects.
(iv) Test can be done on large thickness samples/materials.
(v) Automatic and semi-automatic techniques provide accurate evaluation of the defects.
(vi) The observed results are instantaneous and hence the acceptance/rejection can be made on the spot.
(vii) It is harmless and hence, operated under any circumstances.
(viii) It is highly portable which leads to on-line inspection.
(ix) It uses low cost components as compared to other techniques.
(x) It can also be used to measure thickness.

6.6.8 Limitations

Some of the limitations of the UT are as follows:

(i) It requires a trained person for operating the manual equipments.
(ii) It requires extensive technical skill for the development of inspection procedure.
(iii) It requires couplant for the effective transfer of ultrasonic signal to the materials to be tested.
(iv) Calibrating blocks reference and standards are required for standardizing the equipment and characterizing the defects.

6.7 THERMOGRAPHY

All materials emit electromagnetic radiations when they are at a temperature greater than 0 K. These radiations are predominantly infrared (IR) radiations, whose frequency occupies between visible light and radio waves in the electromagnetic spectrum. The IR rays can be converted into a suitable visual image by detecting them with the help of a detector. Thermography is the process of producing the thermal image of the object surface by detecting and analysing the infrared radiations emitted from

the object. Most of the commercial thermogram lies in between 0.75 and 15 mm. It is an important tool in all the fields, such as industry and medical, in view of their non-destructive, non-invasive, comfortable, easy operations and the interpretation of the observed results. In most of the thermography instruments, the photovoltaic and photoconductive infrared radiation detectors are universally employed.

In view of the immense applications of thermography in all the fields, the general principle, the different types of recording thermogram and their applications in industry and medicine are given in detail below.

6.7.1 Basic Principle

The principle behind thermography is the emission properties of the object. The radiation emitted by the object is given by the formulae, namely, Kirchhoff's law (the coefficient of absorption will be equal to the coefficient of emission under thermal equilibrium), Stephan–Boltzmann's law (any high temperature body will emit more IR radiation), Wien's displacement law (the wavelength corresponding to the maximum energy is inversely proportional to the temperature) and Planck's equation (the relationship between the spectral emissivity, temperature and radiation energy). Therefore, the main factors which affect the thermal measurements are emissivity, surroundings and atmosphere.

The emissivity of a material depends on the nature of the material and condition which is a critical parameter for the quantitative measurements of temperature. It is important that the object should be free from radiation sources, otherwise there will be an enormous increase in the temperature of the radiation from the objects. The effect of atmosphere is an important effect to be considered. Due to the change in atmosphere, there is a possibility of loss of radiation from the object and hence, changes in the spectral characteristics. Thus, it requires a suitable correction such as high temperature filters and also, a correction factor to take into the ambient temperature and attenuation due to these filters.

Thermogram is the thermal image of the scanned object surface that can be displayed as light, dark, grey and hot spots in the CRT screen. The warm areas of the object appear as light, the cool area as dark and the intermediate temperature areas as grey on the thermogram. On the other hand, the local disorders (for example, in the medical field) infection, burns, superficial turmours and general inflammation produce a hot spot on the thermogram due to the increased object surface temperature as the result of the increase in blood supply.

The infrared detector used in a photoconductor is made up of a single crystal of indium–antimony alloy (InSb). The resistance of the crystal changes when a light radiation is incident on the crystal, depending on the intensity of the incident light. This change in resistance is translated into voltage variations for further amplifications and display. The observed infrared radiations from the surface of the object are of low intensity and hence, the noise or background radiations will disturb the amplified voltage. The amplified variations in voltage explore the original voltage monitored on the surface. One can easily reduce the thermal noises by cooling the detector by liquid nitrogen. The sensitivity of InSb, Pb_xSn_{1-x} and $SnHg_{1-x}$ Cd_x detectors at liquid nitrogen temperature (77 K) ranges, respectively, from 3 to 6 μm, 8 to 13 μm and 7 to 13, μm. In recent years, detectors like pyroelectric ($BaTiO_3$), with high sensitivity and stability, can be operated at room temperature without liquid nitrogen.

The McGraw·Hill Companies

6.7.2 Detectors and Equipment

The thermographic equipment are of two different types, namely, contact and non-contact telesystem such as thermographic camera. The surface systems are simple and less costly, but have high resolution. The non-contact system, i.e., tele camera system has wider engineering and medical applications. The block diagram of the non-contact system is shown in Fig. 6.36. It consists of an infrared detector, optical systems, scanning mechanism and associate accessories for recording, storing and processing the observed images. There are two different types of thermographic techniques, namely, active and passive. In the active technique, a part/complete object is subjected to heating/cooling externally and hence, the movement and redistribution of the temperature profile across the surface is measured. In case of passive technique, the natural distribution of heat is measured.

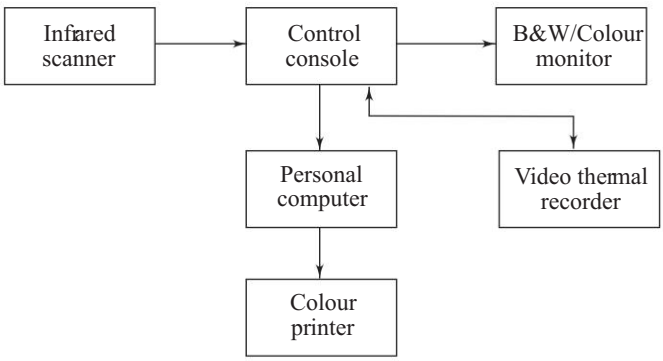


Fig. 6.36 *Block diagram of the thermographic system*

The thermographic camera and its working in the field of medicine is described in the section given below.

6.7.3 Thermographic Camera

The block diagram of the scanning thermographic camera is shown in Fig. 6.37. The thermographic camera is fixed just above the table where the object lies. The plane image scanning mirror reflects the infrared radiation received from the object skin to the concave focusing mirror to a small infrared detector, through the beam chopper as shown in Fig. 6.37. The detector converts the incoming infrared signal into an electrical signal. In order to provide a good reference temperature as that of the skin, and also to compensate the electronic drift, the incident radiation is chopped by a polished rotating blade. Thus, the detector will present the image of the skin and the detector alternatively.

The converted electrical signal is then passed into the CRT after amplification by the amplifier. A spot of light appears on the CRT whose intensity is a measure of the temperature of the monitored point on the skin. For obtaining the image of the infrared scanning of the object, as a first step, the object is scanned by oscillating the image scanning mirror. Thus, by changing the angle of the scanning mirror, different strips of the object are scanned. By transforming the movement of the

mirror through the synchronizer and the image of each strip to the CRT, the display of spot on the CRT is exactly synchronised as that of the corresponding spot on the skin. Thus, the image displayed on the CRT screen gives the thermal image, i.e., temperature distribution of the object. To get the complete thermal image of the object, the body is scanned both horizontally and vertically. In recent years, the scanning time ranges from 1 to 2 seconds for approximately 500 – 600 strips. The resolution of the temperature is better than ± 0.07 ºC. In order to avoid the short wavelength radiations emitted from the human body, the germanium and silicon lens focusing elements are employed. The whole set-up is kept under controlled temperature to increase the speed, accuracy and reliability.

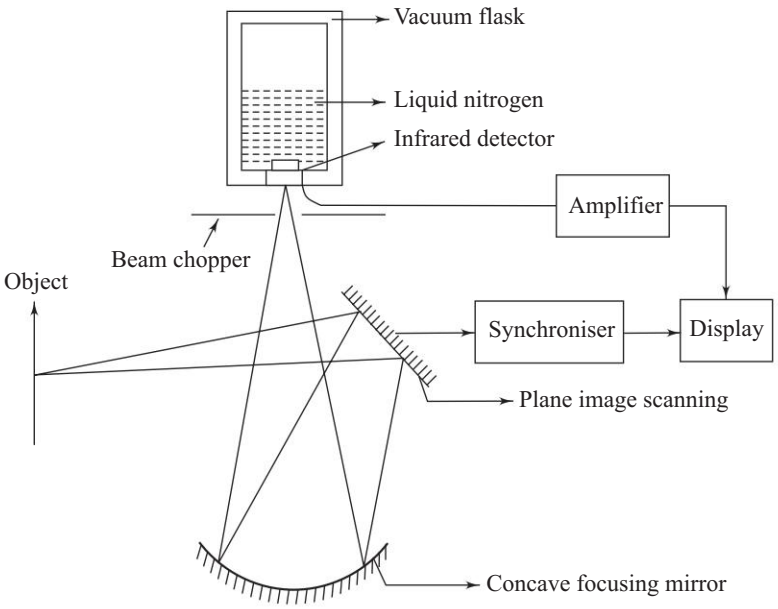


Fig. 6.37 *Thermographic camera*

6.7.4 Non-camera Detectors

Non-camera detectors are also used to detect the defects in materials. In this method, the change in temperature on the surface is determined by monitoring the change in the colour of the cholesteric liquid that is coated on the surface of the material. The cholesteric liquid is a chemical compound which changes its colour with change in temperature. It covers the whole range of colour starting from red to violet, respectively, in the range of temperature 4 – 37 ºC. The material surface is first coated with a thin layer of black paint, and then the cholesteric liquid is coated above it. In order to avoid the degradation of the cholesteric liquid, the measurement has to be carried out immediately after spraying the liquid. The observed images can be taken as colour photographs, at regular intervals to identify the defect in the materials.

The McGraw·Hill Companies

6.7.5 Advantages of Thermographic Studies

Some of the advantages of the thermographic studies compared to other studies are given below:

1. It is possible to assess the problem online.
2. Identification of premature failure leads to extension of components and plan components life.
3. It is a non-invasive, comfortable and easy operation technique.
4. Easy interpretation of the observed results is possible.
5. It shows high resolution and reproducibility.

6.7.6 Applications

Some of the applications of thermography in the field of industry and medicine are listed as follows:

In Industry

Petroleum Industry Thermography is used to monitor the stack temperature, maintenance of plant equipment such as reaction towers, refining furnaces, ducts, etc. It is also used to detect the corrosion in oil tank shell and measurements of oil level.

Condition Monitoring It is an ideal technique for condition monitoring of furnace tubes, gas and fluid transfer lines, etc.

Electrical Power Systems Thermographic inspection can be made on electrical systems such as core insulation, slip ring temperature, etc., of the turbo generator.

Transmission lines The inspection of transmission lines, sub-stations and distribution systems are some of the salient features of the thermography in most of the country.

Insulation Checking It is used to explore the insufficient insulation area that exists in wall by inspection, concrete integrity inspection, etc.

Mechanical System It is used to explore the defects in steam pipes, leakage in pipe line, heat exchanger quality, efficiency evaluation, flame propagation, explosion analysis, furnace refractory inspections, heat ventilation, air-conditioning equipment evaluation, mechanical bearing inspection, power plant, etc. during the working condition.

In Medicine

The applications of thermography in medicine for diagnosing have been used for many years and are vital for different studies.

Detection of Tumours One can easily identify the breast and thyroid tumours by comparing the thermographs of the image of the breast/thyroid. The growth of any tumour tends to draw more blood than surrounding normal tissues. As a result, the flow of extra blood leads to a local hot spot that appears as a light area in the thermogram. This light spot on the image will help to identify the tumours.

Mapping of Blood Vessels The blood vessels are normally used to locate the constrictions or abnormal dilations, and also to monitor the conditions of blood vessel as

the effects of drugs or diet. For example, in order to explore the constrictions or abnormal dilation, the skin of the body to be tested is first cooled below normal low temperature.The cooling is removed and at the same time the thermogram is taken. The thermal images show light areas on the screen, which is due to the rise in temperature of blood vessels, since the blood vessels heat up faster than the surrounding tissues. In a similar way, the effects of drugs or diet on the blood vessels are monitored by the thermograph.

Investigation of Bone Fracture Thermograms are used to detect hairline fractures on bones which are not detectable using X-rays. This is due to the effect of irritations on the surrounding tissues and muscles as an effect of fracture. The obtained thermogram explores the hairline fracture, by observing a hot spot on the screen due to the irritation caused.

Placental Localisation The thermography is a more ideal method to determine the locations of the placenta during the pregnancy without any danger to the mother and child. The increase in placenta leads to an increase in blood flow which results in an increase in temperature that can be easily trapped from the thermogram. The knowledge of the same is more essential for the safe delivery of the child.

Burns and Frostbite In case of burns, the first degree of burns show a high temperature due to the reactive hyperaemia, while the second degree burns shows a lower temperature than the normal temperature of the blood flow. Thus, the changes in temperature are mapped more precisely using the thermogram, which results in the identifications of the depth of tissue destructions, for taking necessary further action.

Keypoints to Remember

- NDT is an acronym of Non-destructive testing.
- Non-destructive Evaluation (NDE) includes the acceptance and rejection criteria of the non-destructive process.
- The different testing methods are visual inspection, liquid penetration inspection, acoustic emission, magnetic particle inspection, radiography, ultrasonic inspection, eddy current and thermography.
- In visual inspection, the human eye is used as a tool and it requires high sensitive optical aids.
- Liquid penetrate testing is used to explore the surface-breaking flaws.
- The principle behind LPT is capillary action.
- The experimental procedure used for LPT is surface preparation, penetrant application, and removal of excess penetrant, applications of developer, and inspection and evaluation.
- Radiography testing is used to detect the internal defects, such as porosity and voids, in materials.
- The process of obtaining a photograph of the object is known as radiography.
- The study of defects using a radiograph is known as radiography.

The McGraw·Hill Companies

- In RT, the locations, shapes, etc., of flaws are determined employing the methods of Tube Shift (TS) method.
- The different types of radiography are X-ray radiography, gamma ray source and neutron ray source.
- The exposure factor for a given object in X-ray radiography is

$$\frac{lt}{L^2} = \text{constant}$$

- Exposure time in gamma ray radiography is defined as the ratio between the exposure factor and source strength at present.
- The attenuation of radiation in a materials is $I = I_o e^{-\mu x}$, where I_o is the intensity of incident radiation at $x = 0$, I the intensity of emerging radiation out of the specimen and μ, the linear absorption coefficient.
- The total absorption of radiation in a material is the sum of the process due to the photoelectric effect, Rayleigh scattering, Compton scattering and pair production.
- In radiography, a large number of sensitive parameters like film-type, source to film distance, film density, etc., are determined using penetrameter.
- Different types of testing in RT are single-wall single-image, double-wall penetration, double-film and special technique.
- Fluoroscopy is the process of obtaining the image of an object by converting X-rays into light on the fluorescent screen. It is also called real time-radiography.
- The principle behind ultrasonic testing is the transmission and reflection or scattering at any surface (or) internal discontinuity in the medium.
- Differents model of propagation of ultrasonic waves are longitudinal, shear and surface waves.
- The ultrasonic waves are generated based on the principle of piezoelectric effect. The transducers are used for the generation of waves.
- UT inspection methods are three types, namely, through transmission, pulse echo and resonance method.
- The different methods in pulse echo techniques are contact types, angle beam, and immersion testing techniques.
- Flaw detector is an electronically controlled unit used to characterise the flaw employing the UT.
- Different types of UT scans are A-scan, B-scan and C-scan.
- In thermography, the image of the object is obtained by observing the image of the object surface by detecting and analysing the infrared radiations emitted from the object.
- The principle behind thermography is the emission property of the object.
- Non-camera detectors are used to detect the defects in materials, i.e, the change in the surface temperature is monitored by changing the colour of cholesteric liquid which is coated on the surface of material.
- Thermography imaging is a non-invasive method.

Solved Problems

Example 6.1 Determine the sensitivity of a 50 mm thick steel wire for a 1mm hole diameter discernible on the radiograph.

Given Data

Thickness of the metal = 50 mm
The hole diameter discernible on the radiograh = 1mm

Solution The sensitivity percentage

$$= \frac{\text{Smallest hole discernible on the radiograph}}{\text{Thickness of metal under test}} \times 100$$

$$= \frac{1}{50} \times 100$$

$$= 2\%$$

The percentage sensitivity of the wire is 2%.

Example 6.2 Determine the exposure factor for a radiograph at a distance of 1 m from the source to the film for an exposure time of 60s. Given that the tube current is 20 mA.

Given Data

The tube current, I $= 20 \times 10^{-3}$ A
The source to film distance, L $= 1$ m
The expoure time, t $= 60$ s

Solution Exposure factor $= \frac{It}{L^2} \times 100$

Substituting the values of I, t and L^2 in the above equation, we get,

$$\text{Exposure factor} = \frac{20 \times 10^{-3} \times 60}{1^2}$$

$$= 1.2$$

The exposure factor is 1.2.

Example 6.3 Determine the thickness of the aluminium absorbed to cut down the intensity of the X-ray beam to 1/20 of its initial value. Given that the linear absorption coefficient of aluminium for X-rays having wavelength about 0.32 Å is 1.62 cm^{-1}.

Given Data

The reduction in the intensity, $I/I_0 = 1/20$

The McGraw·Hill Companies

The liner absorption coefficient, $\mu = 162 \text{ cm}^{-1}$

Solution We know that, $I = I_0 e^{-\mu x}$

or $\frac{I}{I_0} = e^{-\mu x}$

Substituting, the values of I, I_0 and μ in the above equation, we get,

$$\frac{1}{20} = e^{-1.62x}$$

or $$20 = e^{1.62x}$$

$$\log_e 20 = 1.62x$$

or $$x = \frac{\log_e 20}{1.62}$$

$$= 1.85 \times 10^{-2} \text{ m}$$

The thickness of the aluminium is 1.85×10^{-2} m.

Example 6.4 For 200 kV X-rays and steel, the absorption coefficient m has a value of 2 cm^{-1}. In case of a 2.5 cm thick steel, what is the percentage of transmitted X-rays? Also, determine the half value thickness of the specimen.

Given Data

The thickness of the material, $x = 2.5 \times 10^{-2}$ m
The linear absorption coefficient, $\mu = 2 \text{ cm}^{-1}$

Solution We know that $I = I_0 e^{-\mu t}$

or $$\frac{I}{I_0} = e^{-\mu t}$$

Substituting the value of μ and x in the above equation, we get,

$$= e^{-2 \times 2.5} = e^{-5}$$

$$= 0.7\%$$

The percentage of transmitted X-rays from 2.5 cm thick steel is 0.7 %.Similarly, for half value thickness (d_h),

$$\frac{I}{I_0} = \frac{1}{2} = e^{-\mu d_h}$$

$$= e^{-2d_h}$$

$$e^{2dh} = 2$$

or $$\log_e 2 = 2d_h$$

The half value thickness,

$$d_h = \frac{\log_e 2}{2}$$

$$= \frac{0.693}{2} = 0.34657 \text{ cm}$$

The half value thickness of the specimen is 3.5×10^{-3} m.

Example 6.5 For 20 mm steel at 300 mm source film distance, the exposure factor is 0.35 curie hour. Initially the iridium 192 source has a strength of 10 curie. Now its age is 50 days. From the decay chart, it is found that after 50 days, its strength is 5 curie only. Determine the necessary exposure time.

Given Data

The exposure factor = 0.35 curie hour
The present sorce srength = 5 curie

Solution Exposure time

$$= \frac{\text{Exposure factor}}{\text{Source srength at present}}$$

$$= \frac{0.35 \text{ curie hour}}{5 \text{ curie}} = 0.07 \text{ h}$$

$$= 4.2 \text{ min}$$

The exposure time is 4.2 min.

Example 6.6 In the tube shift method, film focus distance is 20 cm. The distance between the images of the flaw on the film is 0.5 cm. The displacement of the X-ray tube is 5 cm. Calculate the location of the flaw from the top surface of the steel specimen of 5 cm thickness. Assume the thickness of the flaw is negligible.

Given Data

The film focus distance, $f_d = 20 \times 10^{-2}$ m
The displacement of the X-ray tube, $S + S_d = 5 \times 10^{-2}$ m
The thickness of the specimen $= 5 \times 10^{-2}$ m

Solution The distance of the flaw from the bottom surface of the steel specimen is

$$d = \frac{f_d S_d}{(S + S_d)}$$

Substituting the value of f, S_d and S in the above equation, we get,

$$d = \frac{20 \times 10^{-2} \times 0.5 \times 10^{-2}}{5.5 \times 10^{-2}}$$

$$= 1.8 \times 10^{-2} \text{ m}$$

The location of the flaw from the top surface of the steel specimen

$$= 5 \times 10^{-2} - 1.8 \times 10^{-2}$$

$$= 3.2 \times 10^{-2} \text{ m}$$

The location of the flaw from the top surface is 3.2×10^{-2} m.

The McGraw·Hill Companies

Objectives Questions

6.1 Defects and inhomogenity in materials are studied by _____ technique.
6.2 The human eye is used as tool for _____ inspection.
6.3 Internal surface of tubes and pipes, and the holes in machines are examined using _____.
6.4 The surface crack and defect in the components are determined with very high precision using _____ method.
6.5 The liquid is drawn into the crack in liquid penetrant testing due to
a) capillary action b) surface tension
c) high viscosity d) low viscosity
6.6 _____ forces are responsible for the surface tension.
6.7 In LPT, the following arrangement is not required.
a) Preparation of good surface b) Propagation of penetrant liquid
c) Selection of suitable edge d) Chemically unstable penetrant
6.8 LPT fails in testing _____ materials.
6.9 Radiography testing is used to detect the internal defects such as _____ and _____ in materials.
6.10 In radiography, location of flaw from the bottom surface of an object is obtained using the relation

a) $d = \dfrac{f_d S_d}{(S + S_d)^2}$ b) $d = \dfrac{f_d S_d}{(S - S_d)^2}$

c) $d = \dfrac{f_d S_d}{(S - S_d)}$ d) $d = \dfrac{f_d S_d}{(S + S_d)}$

6.11 Types of X-rays are _____ and _____.
6.12 The penetrating power of soft X-rays in material is _____ mm.
6.13 _____ mm is the penetrating power of hard X-rays in steel.
6.14 The properties of gamma rays are similar to
a) electron b) X-rays
c) Electromagnetic rays d) β-rays
6.15 The energy of thermal neutron lies between _____ and _____.
6.16 _____ and _____ is the energy of the cold neutrons.
6.17 The expsure factor or a given object in X-rays is

a) $\dfrac{I}{L}$ = constant b) $\dfrac{I}{L^2}$ = constant

c) $\dfrac{I}{L} = 0$ d) None of the above

6.18 The attenuation of radiation in material at adistance x from the surface is
a) $I = I_o e^{\mu x}$ b) $I = I_o e^{\mu x 2}$
c) $I = I^2_o e^{\mu x}$ d) $I = I_o e^{-\mu x}$

The McGraw·Hill Companies

6.19 In pair production, the incident _____ disappeares due to electron-position pair production.
6.20 The sensitivity of radiography is measured using _____.
6.21 _____ camera is used in real time radiography to obtain the image.
6.22 The frequency of ultrasonic waves is above _____ kHz.
6.23 _____ is used to generate and receive the atomic waves.
6.24 In ultrasonic technique method, the thickness of materials is obtained using the relation,

a) $t = \frac{v}{2f^2}$ b) $t = \frac{v}{2f}$

c) $t = \frac{v^2}{2f}$ d) $t = \frac{2v}{f}$

6.25 In group–III television fluoroscopic system, number of sensitive X-ray elements present in a 3 mm x 1 mm array is _____.
6.26 The probe which generate the ultrasonic signals is known as _____.
6.27 _____ probe is used to receive the ultrasonic signals.
6.28 _____ is the most commonly used presentation in UT method.
6.29 UT testing does not require the following for flaw characterisation.
a) a trained person b) a couplent
c) a flaw d) a calibrating block
6.30 The frequency of the radiation which lies between visible light and radio waves in electromagnetic spectrum is _____.
6.31 _____radiation emitted from object surface is used to obtain thermography images.
6.32 The frequency of commercial thermographs lies in between ____ and ____ mm.
6.33 The principle behind thermography is the _____.
6.34 According to Stephan-Boltzmann, high temperature body emits more _____ radiations.
6.35 The following are the main factors which affects thermography measurements
i) emissivity, surroundings and atmosphere
ii) emissivity, surface nature of object and temperature
iii) emmisivity, surface area of object and frequency of radiation
iv) emissivity, operating frequency and atmosphere
6.36 The detectors which is used to operate at room temperature in thermography is known as _____.
6.37 In non-camera thermography, change in the surface of object is monitored by coating _____ liquid.

KEY

6.1 NDT
6.2 Visual
6.3 industrial teleshapes
6.4 Liquid penetrant
6.5 a
6.6 cohesive
6.7 d
6.8 highly porous surfaces

The McGraw·Hill Companies

6.9 porosity, voids
6.10 d
6.11 soft and hard
6.12 10^{-4}
6.13 500
6.14 b
6.15 0.01, 0.3ev
6.16 0, 0.01 eV
6.17 b
6.18 d
6.19 Photon
6.20 image quality indicator
6.21 CCTV
6.22 20
6.23 Transducer
6.24 b
6.25 600
6.26 transmitter
6.27 receiver
6.28 A-scan
6.29 d
6.30 infrared radius
6.31 infrared
6.32 0.75, 15
6.33 emission properties
6.34 infrared
6.35 i)
6.36 pyroelectric
6.37 cholestric

Short Answer Questions

6.1 What is meant by NDT?
6.2 What are the classifications of testing methods?
6.6 Define visual inspection.
6.7 Explain the principle behind visual inspection.
6.8 Mention few general information obtained from VI.
6.9 Why is the optical aid required in visual inspection?
6.10 Explain liquid penetrant testing.
6.11 What is the basic principle behind liquid penetrant testing?
6.12 Why is the cleaning of surface essential in LPT?
6.13 What are the different stages of LP testing?
6.14 What is the role of developer in LP testing?
6.15 How is the defect identified in LP testing?
6.16 Mention some of the characteristics of testing materials in LPT.
6.17 What is meant by penetrant?
6.18 Mention few applications of LPT.
6.19 What are the advantages of LPT over other techniques?
6.20 What is meant by limitations? List out a few limitations in case of LPT.
6.21 What is meant by radiography?
6.22 Explain the principle behind radiography.
6.23 What are the different types of radiography?
6.24 Explain how the radiography is classified into different types.
6.25 What is meant by gamma ray radiography?
6.26 Explain neutron radiography.
6.27 Define the attenuation of radiography in material.
6.28 What are the different types of interactions?
6.29 What takes place with atom in the materials during the propagation of X-ray?

6.30 What is meant by Rayleigh scattering?
6.31 Explain how the X-rays interact with film.
6.32 Explain how the radiographic imagining is obtained.
6.33 Explain why the radiographic film is different from conventional film.
6.34 What is meant by inherent unsharpness?
6.35 What are the different techniques used for radiography and explain how one can select a suitable technique?
6.36 What is meant by superimposing?
6.37 Explain double film technique.
6.38 Explain multi wall penetration technique.
6.39 What is meant by real time radiography?
6.40 What is meant by image intensifier?
6.41 Why is the image intensifier used in X-ray radiography?
6.42 Why is the television fluorescence system used in X-ray radiography?
6.43 Mention some of the applications of radiography in industries.
6.44 What are the limitations of X-ray radiography?
6.47 What is meant by ultrasound?
6.48 Explain the principle behind the ultrasonic testing.
6.49 What are the different modes of ultrasonic waves?
6.50 Explain how transverse and shear waves are different from each other.
6.51 What is meant by transducer?
6.52 What are the different methods used in ultrasonic testing in industries?
6.53 Explain the through transmission method.
6.54 What is meant by pulse echo method?
6.55 Explain how the through transmission and pulse echo methods are different from each other.
6.56 What is meant by resonance?
6.57 What are the different types of testing in contact type of testing?
6.58 Differentiate between normal and angle beam ultrasonic testing.
6.59 What is meant by immersion testing?
6.60 Define flaw detector.
6.64 Mention some of the applications of ultrasonic testing in industries.
6.65 What are the limitations of ultrasonic testing in industries?
6.66 What is thermography?
6.67 Explain the principle of thermography.
6.68 Explain why atmospheric change affects the IR radiation from the object.
6.69 Explain the materials used as infrared detectors.
6.70 What is meant by constant telesystem?
6.71 What is meant by active thermography?
6.72 What is meant by passive thermography?
6.73 How is the light radiation converted into electrical signal?
6.74 What is meant by chopping?
6.75 Mention any two applications of thermography in industry.
6.76 Mention any two applications of thermography in medical field.

The McGraw·Hill Companies

Descriptive Questions

6.1 Explain briefly how the surface flaws are evaluated using liquid penetrant testing.
6.2 What is meant by liquid penetrant testing? Explain the different techniques along with their limitations and applications.
6.3 What is meant by magnetic particle testing? Explain the basic principle and the different techniques with suitable illustrations.
6.4 Explain the principle behind radiography. What are the different types of radiography? Explain with suitable sketch.
6.5 Explain with neat sketch, the various techniques used for radiography to characterize the flaw.
6.7 What is meant by fluoroscopy or real time radiography? Explain with neat sketch the working of television fluoroscopic system.
6.8 What is meant by image intensifier? Compare with neat sketch, the performance of the different stages of television fluoroscopic systems.
6.9 What is the basic principle in ultrasonic testing? Explain with neat sketch the different testing methods used in ultrasonic testing.
6.10 Explain with neat sketch how the flaw is characterised using the pulse echo technique.
6.11 What is meant by ultrasonic flaw detector? Explain the different methods of defect display methods with suitable examples.
6.12 Compare the relative merits of various NDT methods which are used for flaw characterization.
6.13 Explain with block diagram the non-contact telesystem thermography along with industrial applications.
6.14 Explain with neat sketch the working of thermography camera and explain how the mapping of the defects in the surface of the object is made.
6.15 Write an essay on the thermography principle, working and applications, both in industry and medicine.

Chapter

Interference

7

OBJECTIVES

- To study the principle of interference, interference pattern and its applications
- The study the mechanism of obtaining interference pattern in a thin film and a wedge-shaped film
- To study the principle, working and applications of a Michelson interferometer
- To study the principle, coating structure and applications of antireflections coating in optical instruments
- To study the principle, working and application of etalon

7.1 INTRODUCTION

Interference is an important phenomenon used for many applications such as determination of wavelength of light source and thickness of transport material film. The principle behind the interference is the superimposition of one light wave over another. As a result of superimposition of two light energies, the intensity of the resultant light is distributed as maximum and minimum intensity. The formation of maximum intensity is known as *constructive interference* which is formed due to the overlapping of the crest of one wave on the crest of another light wave. On the other hand, when the crest of one wave falls on the trough of another wave, it gives the minimum intensity and is known as *destructive interference*. Rainbows observed during rainfall is due to interference pattern produced by sunlight.

In this chapter, the principle, operation and applications of interferometers are discussed along with suitable illustrations in detail.

7.2 INTERFERENCE PRODUCED IN THIN FILMS DUE TO REFLECTED LIGHT

Consider a transport film *XYWZ* of thickness t and refractive index μ as shown in Fig.7.1. *XY* and *WZ* are respectively the upper and lower surfaces of the film. Consider that a ray of light *AB* is incident on the upper surface of film at an angle i. At the point

B, a part of the ray is reflected as *BC*, while a part of the ray is transmitted through the film as *BD*. The angle of the refracted beam is *r*. The transmitted ray is partly reflected as *DF* at the lower surface *WZ* of the film and partly transmitted as *DE*. Similarly, the ray *DF* undergoes reflections and refractions respectively as *FH* and *FG* at the upper surface of the film *XY*.

The path difference between the two rays is determined by considering two rays, namely, *ABC* and *ABDFG*. The distance traveled by ray *ABDFG* is more than that traveled by *ABC*. In order to determine the path difference, draw a perpendicular line from the point *F* to the line *BC*. Hence, the line *JF* is normal to the line *BC*. Therefore, the path difference between the two rays is

$$x = \mu(BD + DF) - BJ \tag{7.1}$$

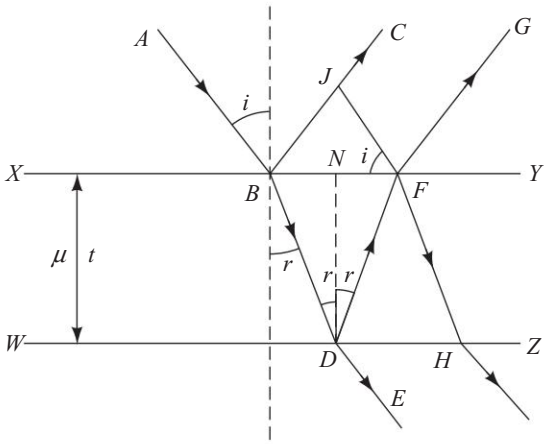


Fig. 7.1 *Interference—thin films*

Consider the triangle *BDN*

$$\cos r = \frac{DN}{BD}$$

$$BD = \frac{DN}{\cos r} = \frac{t}{\cos r}$$

Similarly, from Δle *DNF*

$$DF = \frac{t}{\cos r}$$

Therefore, $$BD + DF = \frac{t}{\cos r} + \frac{t}{\cos r} \tag{7.2}$$

Further, from Δle *BDN*

$$\sin r = \frac{BN}{BD}$$

$$BN = BD \sin r$$

Therefore, $$BF = 2\,BN = 2\,BD \sin r \tag{7.3}$$

Substituting the value of *BD* in eq (7.2), we get,

$$BF = 2\,\frac{t}{\cos r}\sin r \tag{7.4}$$

Consider the Δ^{le} *BFJ*,

$$\sin i = \frac{BJ}{BF}$$

Therefore, $$BJ = BF \sin i \tag{7.5}$$

From Eqs. (7.4) and (7.5),

$$BJ = \frac{2t \sin i \sin r}{\cos r} \tag{7.6}$$

According Snell's law,

$$\mu = \frac{\sin i}{\sin r}$$

$\therefore$ $$\sin i = \mu \sin r \tag{7.7}$$

Substitute the value of sin *i*, in Eq. (7.6), we get,

$$BJ = \frac{2t \sin r . \mu \sin r}{\cos r} \tag{7.8}$$

Substituting the value of *BD* + *DF* and *BJ* from Eqs. (7.2) and (7.8) in Eq. (7.1), we get,

$$x = \frac{\mu\, 2t}{\cos r} - \frac{2\,\mu\, t}{\cos r}\sin^2 r$$

$$= \frac{2\mu\, t}{\cos r}(1 - \sin^2 r)$$

$$= 2\,\mu\, t \cos r$$

We know that *BC* is the reflected ray while *FG* is the transmitted ray. Therefore, there is a phase change π degree or an additional path difference $\lambda/2$ between these two rays. Therefore, the total path difference is

$$x = 2\,\mu\, t \cos r - \frac{\lambda}{2} \tag{7.9}$$

We know that the condition for bright fringe, i.e., path difference is an integral multiple of λ, (i.e., $n = 1,2,3,\ldots$).

Therefore, $$2\,\mu\, t \cos r - \frac{\lambda}{2} = n\lambda$$

or $$2\,\mu\, t \cos r = \left(\frac{2n+1}{2}\right)\lambda \tag{7.10}$$

Similarly, the condition for a dark fringe, the path difference is equal to half integral multiple of λ,

The McGraw·Hill Companies

Therefore, $2\mu t \cos r - \frac{\lambda}{2} = (2n-1)\frac{\lambda}{2}$

or $$2\mu t \cos r = n\lambda \tag{7.11}$$

The conditions for bright and dark fringes are given by Eq. (7.10) and (7.11) respectively.

7.3 AIR WEDGE

An air wedge is a wedge-shaped air film which is obtained in between two glass plates. The experimental arrangement used to produce the air film is shown in Fig. 7.2. The air wedge is used to measure the thickness of a thin sheet of paper and also to test the planeness of a glass plate.

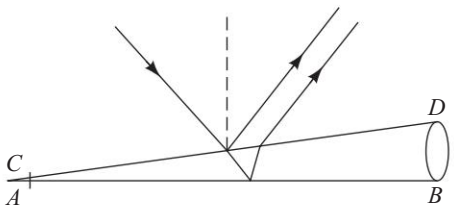


Fig. 7.2 *Air film—wedge shaped*

Consider two glass plates *AB* and *CD*, which are placed one above the other as shown in Fig. 7.2. One end of the glass plates are tied using a rubber band while on the other end, a thin sheet of paper whose thickness is to be measured is introduced. A wedge-shaped air film is formed between the two glass plates. The thickness of the film at the paper is maximum while it is minimum at the point of contact of the two glass plates. When a light ray is incident perpendicular to the top surface of the glass plate, an interference pattern is formed. The interference pattern consists of alternate bright and dark parallel fringes.

The interference pattern consists of alternate fringes. Consider that the n^{th} bright fringe is formed at a distance x_n from the tied end. Then, the condition for bright fringe is,

$$2\mu t_n \cos r = \frac{(2n+1)}{2}\lambda \tag{7.12}$$

where t is the thickness of the film and μ the refractive index of the air film. We know that for normal incidence of light, $i = 0$. Similarly, the angle of refraction is very small for normal incidence and hence $\cos r = 1$.We know that for air, $\mu = 1$. Applying the above conditions, the condition for bright fringe is

$$2t_n = (2n+1)\frac{\lambda}{2} \tag{7.13}$$

Consider that the next bright fringe $(n+1)^{th}$ fringe occurs at x_{n+1}. Let t_{n+1} be the thickness of the air film at x_{n+1}.

Therefore, condition for $(n+1)^{th}$ bright fringe is

$$2t_{n+1} = (2(n+1)+1)\frac{\lambda}{2}$$

or $$2t_{n+1} = (2n + 3)\frac{\lambda}{2} \tag{7.14}$$

Subtracting Eq. (7.13) from Eq. (7.14), we get,

$$2(t_{n+1} - t_n) = \frac{3\lambda}{2} - \frac{\lambda}{2} = \lambda$$

or $$t_{n+1} - t_n = \frac{\lambda}{2} \tag{7.15}$$

Equation (7.15) indicates that the next bright fringe $(n+2)$ takes place at $\lambda/2$, which is equal to the distance between the $(n + 1)^{th}$ and $(n + 2)^{th}$ bright fringe.

Consider that there are n bright fringes between the distance x_o and x_n. These fringes are formed at $x_1, x_2, \ldots x_n$ respectively at the corresponding film thickness $t_1, t_2, \ldots t_n$. Therefore, the distance between x_0 and x_n is

$$t_n - t_o = n\frac{\lambda}{2} \tag{7.16}$$

Let the distance, $x_n x_0 = x$.

Therefore, $$\theta = \frac{t_n - t_o}{x_0 x_n} = \frac{n\lambda}{2x}$$

Rearranging the above equation, we get

$$x = \frac{n\lambda}{2\theta} \tag{7.17}$$

Substituting, the values of n = 1, 2,3,….. The position of the 1st, 2nd, 3rd ..... n^{th} bright fringes are obtained from the above equation by taking n = 1, 2, 3…..

$$n = 1, x_1 = \frac{\lambda}{2\theta}$$

$$n = 2, x_2 = \frac{2\lambda}{2\theta}$$

$$n = 3, x_3 = \frac{3\lambda}{2\theta}$$

$$n = 4, x_4 = \frac{4\lambda}{2\theta}$$

.....

Therefore, when $x = n$,

$$x_n = \frac{n\lambda}{2\theta} \tag{7.18}$$

Let β be the fringe or bandwidth, i.e, the distance between any two successive fringes.

Therefore, $$\beta = x_2 - x_1 = x_3 - x_2, \text{ etc,}$$

or $$\beta = \frac{\lambda}{2\theta} \tag{7.19}$$

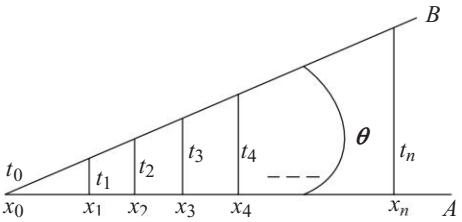


Fig. 7.3 *Air wedge*

Let l be the length of the glass plate and t, the thickness of the paper which is introduced between the glass plates.

Therefore, from Fig. 7.3,

$$\theta = \frac{t}{l} \tag{7.20}$$

Substituting the value of θ in Eq.(7.19), we get

$$\beta = \frac{\lambda}{2}\frac{l}{t}$$

Therefore, the thickness of the paper

or $$t = \frac{\lambda l}{2\beta} \tag{7.21}$$

where l is the length of the glass plate and λ the wavelength of monochromatic light.

Equation (7.21) gives the value of the thickness of the paper introduced in between the two glass plates.

7.3.1 Applications of Air Wedge

The principle of forming a wedge shaped air film between glass plates is used to test the plainness of a glass plate and to measure the thickness of a thin sheet of paper.

(i) Testing of Flat Surfaces The testing of flat surface of a glass plate is carried out by forming the airwegde and measuring the fringe pattern. The glass plate for which the planeness is tested is placed above the optically flat glass plate. Thus, the interference pattern forms an air film in between the glass plate. The observed interference pattern is used to test if the given glass plate is flat or not. When the observed fringes are of equal thickness and straight lines, then it shows that the given glass plate is flat. On the other hand, when the observed fringes are straight and coloured, it shows that the glass plate is not flat. The same can be explained by considering the necessary condition for a bright fringe.

We know that the condition for bright fringes are

$$2\,\mu t \cos r = (2n + 1)\frac{\lambda}{2} \tag{7.22}$$

The McGraw·Hill Companies

When the incident beam is normal ($\cos r = 1$) and the medium is air ($\mu = 1$), the above equation can be written as,

$$t = (2n + 1)\ \frac{\lambda}{2} \tag{7.23}$$

It is clear that the value of n determines the interference pattern. When straight-line fringes are obtained, it means that the thickness of the air film and flatness of the given glass plate are constant. On the other hand, if the glass plate is not flat, it gives non-uniform air-film thickness and hence, coloured fringes. The observed coloured pattern indicates that the glass plate is not flat. The accuracy of the measurement of flatness depends on the wavelength of light used.

(ii) Determination of Thickness of a Thin Sheet of Paper The thin sheet of paper whose thickness is to be measured is kept at one end of the two glass plates *AB* and *CD*. The air wedge is formed in between the glass plates. The monochromatic light from the source *S* is made incident on the glass plate which is kept at an angle of 45o to the direction of light after passing through the lens *L*. The incident rays are reflected at the glass plate *G* towards the air wedge as shown in Fig. 7.4.

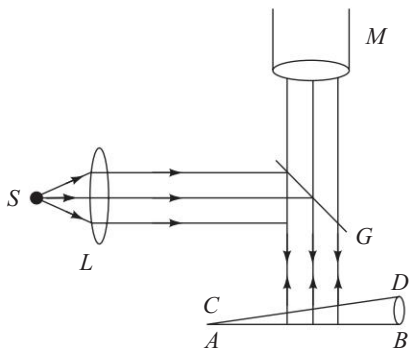


Fig. 7.4 Air wedge—experimental arrangement

A part of the incident rays get reflected from the upper surface of the glass plate AB while the remaining part gets reflected from the bottom surface of the glass plate CD, passing through the first glass plate. There is a path difference between the rays reflected from the top surface of the glass plate AB and glass plate CD and hence, the superposition of rays takes place. As a result, an interference pattern is formed. The interference pattern which consists of alternate bright and dark fringes is seen through the microscope M.

Consider any one of the bright or dark fringes as the n^{th} fringe. The microscope readings for n, $n + 5$, $n + 10$, $n + 15$ up to $n + 15$ fringes are taken. The fringe width β for the pattern is obtained from the microscope reading. Let l be the length of the glass plate and t the thickness of the paper

Therefore, the thickness of the paper

$$t = \frac{l\lambda}{2\beta} \tag{7.24}$$

where λ is the wavelength of monochromatic light.

Equation (7.24) is used to determine the thickness of the given paper.

The McGraw·Hill Companies

7.4 ANTI-REFLECTION COATING

When a material is coated on the surface of a glass plate, it minimises the reflection of light from its surface. The coating is known as *anti-reflection coating* or *non-reflecting film*. The material whose refractive index lies between the refractive index of air ($\mu = 1$) and glass plate (($\mu = 1.5$) is used for coating. A typical anti-reflection coating on a glass plate and the performance of light reflection is shown in Fig.7.5. Consider n_a, n_2 and n_2 are respectively the refractive index of air, coating and glass material. When a light ray *AB* is incident on the upper coating surface S_1, it gets reflected along *BC*. The transmitted ray *BD*, gets reflected from the surface S_2 as *DE* and then refracted at the surface S_1, as *EF*. There is a path difference between the rays *ABC* and *ABDEF*, and hence, a phase change π is obtained between the two rays.

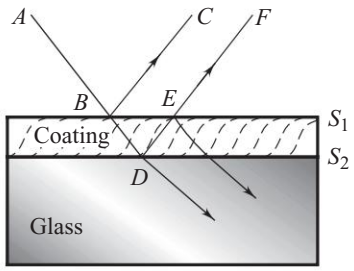


Fig. 7.5 *Anti-reflection coating*

We know that for destructive interference, the path difference should be equal to $\lambda / 2$.

Therefore,
$$2nt \cos r = \frac{\lambda}{2} \tag{7.25}$$

where r is the angle of incidence, n the refractive index of coating and λ the wavelength of light. We know that for normal incidence, $\cos r = 1$.

Therefore, Eq. (7.25), can be written as,

$$2nt = \frac{\lambda}{2}$$

or

$$t = \frac{\lambda}{4n} \tag{7.26}$$

The optical thickness of the film is

$$nt = \pi/2 \tag{7.27}$$

Consider that a single film of thickness t refractive index n_1, is coated on a glass plate whose refractive index is n_2. The reflection coefficient of the combination of three medium (film, glass and air) is

$$r = \frac{n_1(1-n_2)\cos kt - i(n_2 - n_1^2)\sin kt}{n_1(1+n_2)\cos kt - i(n_2 + n_1^2)\sin kt} \tag{7.28}$$

When the value of $kt = \pi/2$ is equal to the optical thickness of film, $nt = \lambda/4$. Therefore, the reflectance

$$r = \frac{-i\left(n_2 - n_1^2\right)}{-i\left(n_2 + n_1^2\right)} \tag{7.29}$$

Hence, the reflectance of quarter-wave film is,

$$R = \mid r \mid = \frac{\left(n_2 - n_1^2\right)^2}{\left(n_2 + n_1^2\right)^2} \tag{7.30}$$

$R = 0$, when the reflectance is minimum. The above equation can be written as

$$n_1 = \sqrt{n_2} \tag{7.31}$$

The thickness of an optical coating is an effective tool to reduce the reflectance. For a particular wavelength of visible spectrum, the thickness of the coating film is equal to its optical thickness $\lambda/4$. For example, magnesium fluoride ($n = 1.38$) coated glass reduces its reflectance to less than 1% from 4% in an ordinary glass.

In optical instruments, an anti-reflection coating helps to control light reflectance. For example, a multilayer anti-reflection coating used in a high-quality camera lens reduces the haziness and hence improve the brightness of the image. The multilayer coating consists of different materials with different refractive indices. For example, two and three layer coatings are made by an approximate selection of materials. A two-layer coating which is made of up one high and one low refractive index of the materials are used to obtain zero reflectance for a particular wavelength. Similarly, a three-layer coating is used to obtain the zero reflectance for two wavelengths. The reflectance as a function of wavelength for different anti-reflecting film coatings is shown in Fig. 7.6.

Table 7.1 *Different materials used for coating along with layer types and refractive indices*

S.No	*Materials*	*Types of layer*	*Refractive index n*
1.	ZrO_2	H	2.1
2.	TiO_2	H	2.4
3.	$ZnSO_4$	H	2.32
4.	$MgFO_2$	L	1.38
5.	CeF	L	1.63

Generally, a two-layer film is denoted as gHLa represents glass high index layer, low index layer and air medium. The materials used as low and high layers and their refractive indexes are given in Table 7.1.

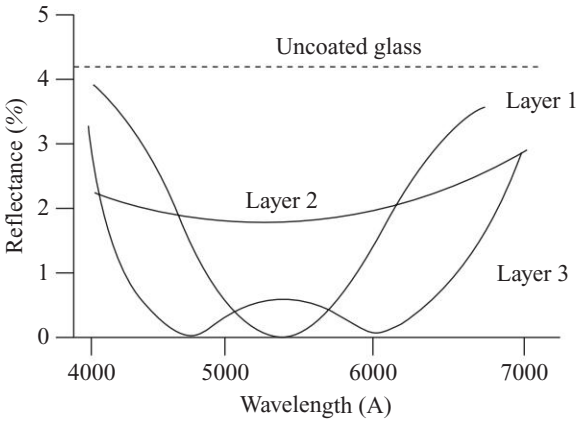
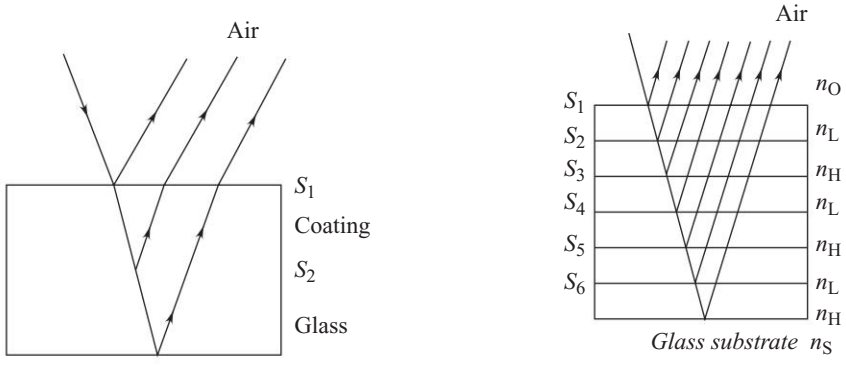


Fig. 7.6 *Reflectance versus wavelength of anti-reflection coating—plot*

7.5 MULTILAYER PERIODIC SYSTEM

Instead of two or three layers of coating, a multilayer coating with number of alternate high and low index materials is used for many applications, such as filters. A schematic diagram of multilayer coating is shown in Fig. 7.7. A simple multilayer periodic system consists of a stack of alternate layer of high-index (n_H) and low-index (n_L) materials, and the layers are represented by S_1, S_2, S_3, S_4, S_5 and S_6. The thickness of each layer is equal to one-fourth of its incident light wavelength. The schematic representation of double high and low refractive index layers, namely double quarter (gHLa) and quarter-wave stack $(g(HL)^3a)$ are shown respectively in Fig. 7.7.



(a) Double quarter (gHLa) (b) Quarter wave-stack $(g(HL)^3a)$

Fig. 7.7 *Multilayer films*

Consider that the refractive indices of air and substrate are equal to unity. Therefore, the reflectance of 2 *n* layers is,

$$R = |r|^2 = \left[\frac{\left(\frac{-n_H}{n_L}\right)^N - \left(\frac{-n_L}{n_H}\right)^N}{\left(\frac{-n_H}{n_L}\right)^N + \left(\frac{-n_L}{n_H}\right)^N}\right]^2 \tag{7.32}$$

Simplifying the above equation, we get,

$$R = \left[\frac{\left(\frac{n_H}{n_L}\right)^{2N} - 1}{\left(\frac{n_H}{n_L}\right)^{2N} + 1}\right]^2 \tag{7.33}$$

Equation (7.33) indicates that the reflectance becomes unity for large value of N. A maximum reflectance is obtained only for a particular wavelength. On the other hand, the reflectance region is broadening either using layers with different thickness or increasing the ratio of index (n_H/n_L). The maximum reflectance is achieved employing a multilayer film with high-index materials, namely, $g(HL)^m Ha$. The multilayer film finds wide applications such as heat reflectivity and transmitting mirrors, filters, etc.

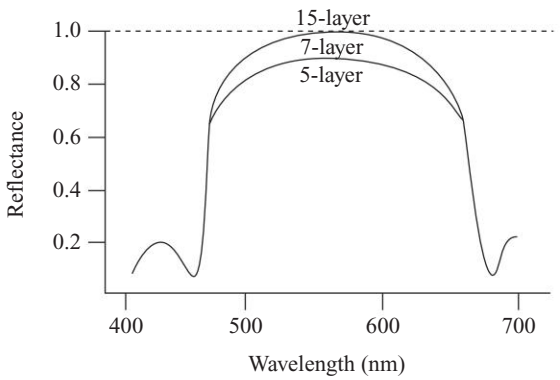


Fig. 7.8 *Reflectance versus wave length of multilayer high reflectance film—plot*

7.6 HIGH-PASS AND LOW-PASS FILTERS

The observed small peak on the short wavelength side of the central zone is decreased exploring a low index with a thickness of λ/8. The corresponding structure is represented as,

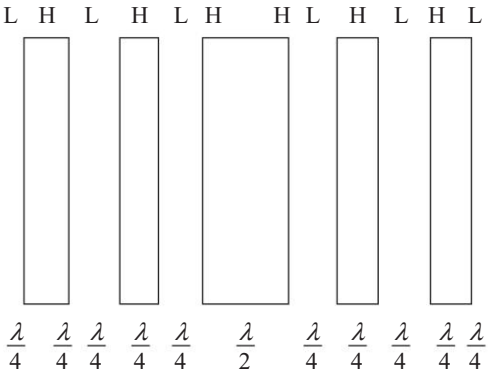
$$g\ (0.5L)\ (HL)^m\ (0.5L)a$$

The transmittance of the above structure is high at high frequency and hence, it is known as high-pass filter. Similarly, a high transmittance at low frequency is obtained using the following structural arrangements. The above structure is used to obtain high transmittance at low frequency and hence it is known as low-pass filter.

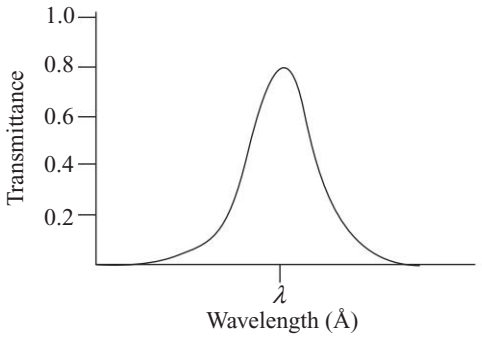
$$g\ (0.5H)\ L(HL)^m\ (0.5H)a$$

7.7 INTERFERENCE FILTER (FABRY–PEROT INTERFERENCE FILTER)

A Fabry–Perot interference filter arrangement is similar to a Fabry–Perot etalon with small spacing, which consists of a layer of dielectric material with a thickness of λ/2. The dielectric material is bound to partially reflecting surfaces on both sides. The transmission curve for this filter arrangement is shown in Fig 7.9. It has peak at the wavelength λ. It is possible to block almost all the higher-order peaks which occur at different wavelengths λ/2, λ/3, λ/4… and so on, by using absorbing filters of colored glass.



(a) Fabry-Perot interference filter



(b) F-P Transmission curve

Fig. 7.9 *Interference filter*

The above structure is made possible in any one of the following non-absorbing Fabry–Perot filter structure employing materials of low index *L* and high index materials with thickness λ/4.

Structure 1 : *g* HLH LLHLHa

Structure 2 : *g* HLHL HH LHLHa

The McGraw·Hill Companies

In both the structures, the central double layer, namely, *LL* and *HH* form the Fabry–Perot cavity with a thickness of λ/2. The influence of this cavity on the reflectance for a particular wavelength is nil. The schematic representation of a Fabry–Perot filter and the response of transmission curve are shown in Fig. 7.9.

7.8 MICHELSON'S INTERFEROMETER

Michelson's interferometer is working based on the principle of interference of light. It consists of mirrors, glass plates and a telescope. The schematic representation of Michelson's interferometer which is used to observe the formation of interference fringes is shown in Fig 7.10. The two plane mirrors M_1 and M_2 are placed at right angles to each other. The mirror M_1 is fixed while M_2 is adjustable. The mirror M_2 consists of three leveling screws at the back portions which are used to adjust the two mirrors M_1 and M_2 to make them perpendicular to each other. G_1 and G_2 are optically parallel glass plates-made up of the same material and having the same thickness. These two glass plates are placed parallel to each other at an angle of 45° to the incident beam. The back side of the glass plate G_1 is semi-coated with silver which helps to split the incident light beam in two beams. The monochromatic light from the source *S* is made to fall on the lens *L* to get a parallel beam of light. The observed interference pattern is viewed through the telescope *T*.

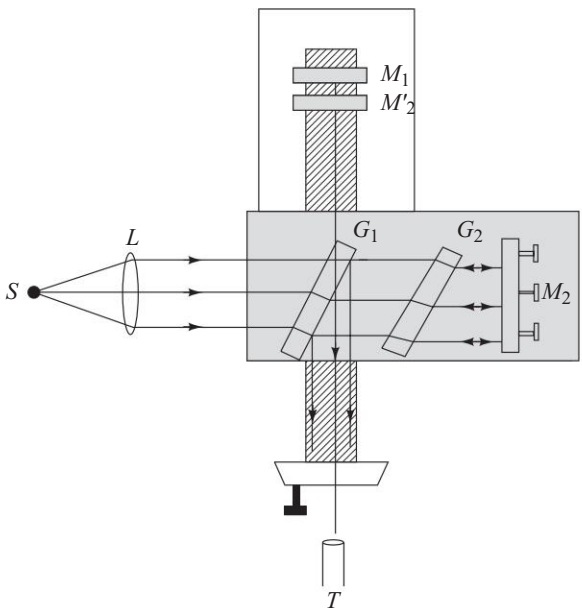


Fig. 7.10 *Michelson's interferometer*

7.8.1 Working

The monochromatic light from the source *S* is made to fall on the lens *L*. The lens *L* converts the monochromatic light into a parallel beam of light. The parallel beam of light is made to fall on a half-silvered glass plate G_1, which is at an angle of 45° to

The McGraw·Hill Companies

the direction of the beam. The incident light is split up into two parts (Fig.7.10) at G_1. The reflected part of the beam travels in a direction at right angles to the incident beam and fall on the plane mirror M_1. The reflected light from M_1 is incident on G_1 and gets refracted through G_1. The refracted light enters the telescope T. On the other hand, the transmitted part of the incident beam travels along the direction of initial beam, and falls normally on mirror M_2.

The light gets reflected from the mirror M_2 and returns to the silicon-coated surface of glass plate G_1 on the same plate. The light gets reflected at G_1 and then is incident on the telescope. The two reflected beams interfere and hence, interference fringes are formed. The observed interference pattern is viewed through the telescope T. The reflected beam at G_1 travels a to-and-fro path from G_1 to M_1 and then, gets refracted through G_1 to reach the telescope. However, the refracted beam from G_1 travels the path G_1 to M_2 and then M_2 to G_1 to reach the telescope. Thus, there is a path difference between the reflected and refracted beam. In order to compensate, a glass plate G_2 of the same material and thickness is introduced between G_1 and M_2 as shown in Fig.7.10. The virtual image of M'_2 of the mirror M_2 is obtained in addition to M_1 while viewing through the telescope. Generally, it appears that the two glass plates are one above the other with a formation of air film between the glass plates. This shows an interference pattern while viewing through telescope. Depending upon the path difference between the mirrors M_1 and M_2, the obtained interference patterns are categorised into different types such as straight, circular and parabolic. Let us discuss the formation of different types of fringes under the following headings.

7.8.2 Types of Fringes

(i) Circular Fringes When the interference pattern is viewed through the telescope, it appears that the virtual image M'_2 is close to M_1. The image M'_2 and M_1 are separated by a distance d. Further, it shows that the light is originated from two virtual sources S_1 and S_2 which are separated by a distance $2d$ as shown in Fig. 7.11. The reflected light from the mirror M_2 has an additional phase difference π or a path difference $\lambda/2$ due to reflection at the semi-silvered surface of the glass plate G_1. This leads to a total path difference of $2d + \lambda/2$.

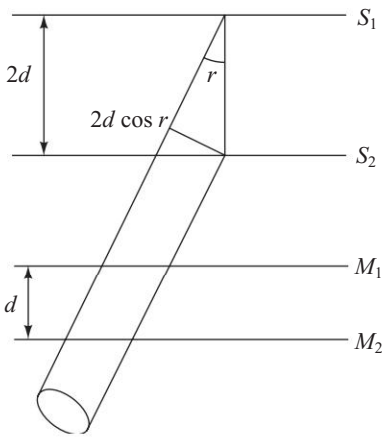


Fig. 7.11 *Michelson's interferometer—Virtual sources*

The McGraw·Hill Companies

Therefore, the total path difference = $2d \cos r + \lambda/2$ (7.34)

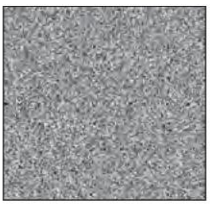
where r is the angle of view and λ, the wavelength of incident light. The mirror M_2 is adjusted to get the path difference equal to $n\lambda$.

Therefore, the condition for bright fringe is

$$2d \cos r + \lambda/2 = n\lambda \quad (7.35)$$

where $n = 1, 2, 3\ldots$ and so on.

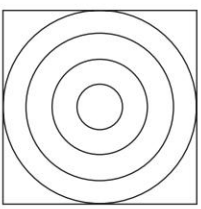
Case I A destructive interference pattern (darkness) is obtained as shown in Fig. 7.12 in the field of view of the telescope only when path difference between the two rays is $\lambda/2$. This is due to the exact coincidence of the virtual image M_2' on the mirror M'_1. In addition, the glass plate G_1 is to be fixed at an angle of 45° to both M'_2 and M_1. When M_2' and M_1 coincides, the value of $2d \cos \theta = 0$ and hence, the path difference is $\lambda/2$.



(a)

Fig. 7.12 *Michelson Interferometer—Destructive interference pattern*

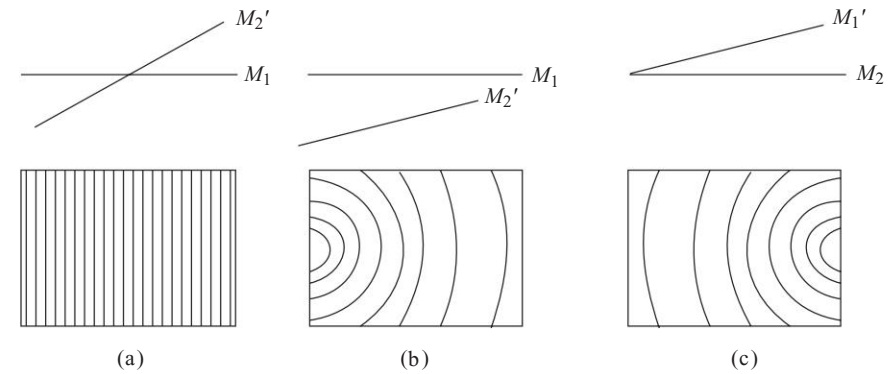
Case II A circular fringe as shown in Fig.7.13 is formed when the mirror M_1 and its virtual image M_2' are exactly parallel to each other. This is achieved by moving the mirror M_1 either in the forward or backward direction to form a small air gap between M_1 and M_2'.



(b)

Fig. 7.13 *Michelson's interferometer—circular pattern*

(ii) Localised Fringes Depending on the position of mirrors M_1 and M_2', the different types of fringes, namely, straight lines, slightly curved and curved fringes are formed as shown in Fig. 7.14. When mirrors M_1 and M_2' intersect as shown in Fig.7.14 (a), straight-line fringes are formed. Similarly, when the mirror M_1 is not intersecting M_2' as shown in Fig 7.14 b, slight curved fringes are formed. On the other hand, when the two mirrors are inclined i.e., the mirror M_1 lies just below the observed image M_2' (not perfectly intersecting) as shown in Fig. 7.14 c, curved fringes are observed. It is interesting to note that the above fringes are made possible only for short wavelengths.



(a) Straight line fringes (b) Slightly curved fringes (c) Curved fringes

Fig. 7.14 *Michelson's interferometer—localised fringes*

(iii) White Light Fringes The monochromatic light source is replaced by white light. The fringes are formed in colour and obtained only when the path difference is very small. The centre of the fringe is dark while the other fringes are coloured due to overlapping with various colours. The observed localised white-light fringes are used to find the zero path difference.

7.8.3 Applications of Michelson's Interferometer

Michelson's interferometer is used for the following applications:

(i) To determine the wavelength of the given source of light
(ii) To determine the refractive index or thickness of a given transparent material
(iii) To determine the resolution of wavelengths
(iv) To standardize one meter

Let us discuss the above applications in brief:

(i) Determination of Wavelength of Monochromatic Light The leveling screw at M_2 is adjusted to make the mirror M_1 and its virtual image M_2' parallel to each other. Thus, it forms an air film of thickness t in between the mirrors and hence, it produces circular fringes. The circular shaped fringes are viewed through the telescope. When the angle of incidence is normal ($\cos r = 1$), the condition for bright fringes is

$$2t + \lambda/2 = n\lambda \tag{7.36}$$

When the two mirrors M_1 and M_2 are at equal distances from the glass plate G_1, the observed pattern will be perfectly dark. The vertical cross wire of the telescope coincides with any one of the fringes. The number of fringes which cross the field of view of the telescope is counted by moving the mirror M_2. Hence, the distance d moved by the mirror is determined from the telescope reading difference. We know that the condition to obtain the interference pattern is $2d = n\lambda$.

The McGraw·Hill Companies

The wavelength of monochromatic light is

$$\lambda = \frac{2d}{n} \tag{7. 37}$$

Equation (7.37) gives the wavelength of the given monochromatic light. The mean value of the wavelength is determined by repeating the experiment a few more times.

(ii) Determination of Refractive Index or Thickness of a Thin Transparent Medium In Michelson's interferometer, white light is used to obtain localised fringes. The field of view is made to coincide with the dark and straight fringes, which produce a zero thickness. A thin film whose refractive index or thickness is to be measured is introduced in between G_2 and M_2. Thus, a path difference of $2(\mu - 1)t$ is introduced due to the thin film. In order to get dark and straight fringes, the mirror M_1 is adjusted either in the forward or backward directions. Let x be the distance moved by the mirror M_1 to obtain n localised fringes.

Therefore, path difference $2t\,(\mu - 1) = n\lambda = 2x$

$$t = \frac{x}{\mu - 1}$$

or

$$\mu = \frac{t + x}{t}$$

Therefore, refractive index $\mu = \frac{n\lambda}{2t} + 1$ (7.38)

Equation (7.38) is used to determine the refractive index of the given film by knowing the thickness of the thin transport film, n and λ . By knowing the values of n, λ and μ, one can also determine the thickness of the film using the above equation.

(iii) Resolution of Spectral Lines Michelson,s interferometer is used to distinguish the two spectral lines of nearly equal wavelength with more resolution. We know that sodium light consists of two spectral lines, namely D_1 and D_2 lines, respectively with wavelengths of $\lambda_1 = 5890$ A° and $\lambda_2 = 5896$ A°. When sodium light is used as a source in the interferometer, it produces alternate bright and dark fringes. The interference pattern produced by λ_1 and λ_2 are different. However, the interference pattern produced by wavelengths λ_1 and λ_2 are not distinct if one views through the telescope, i.e., the bright fringe of the spectral line of λ_1 coincides with the other spectral line λ_2. When the bright fringe of two spectral lines of wavelength λ_1 and λ_2 is viewed through the telescope, it appears as a single line. The fringes produced by the wavelengths λ_1 and λ_2 are separated by adjusting the micrometer screw of the mirror M_1 due to their difference in wavelengths. If one adjusts the mirror M_1, the dark fringe of one spectral line coincides with the bright fringe of another one. When the mirror M_1 is adjusted further, the bright fringe of one spectral line will coincide with the bright fringe of another one. This is possible only when there is a difference in wavelength between the fringes, i.e., the n^{th} fringe of longer wavelength coincides with the $(n + 1)^{th}$ fringe of shorter wavelength. Let

n_1 and n_2 be the order of fringes corresponding to spectral lines λ_1 and λ_2. Therefore, the conditions for the first bright fringes is

$$2d = n_1\lambda_1 = n_2\lambda_2 \tag{7.39}$$

Under this condition, the two spectral lines coincide with each other.

Similarly, the conditions for the next bright fringes of spectrum lines λ_1 and λ_2 will coincide when

$$n_2 = n_1 + 1 \tag{7.40}$$

or

$$n_1\lambda_1 = (n_1 + 1)\lambda_2$$

or

$$n_1 = \frac{\lambda_2}{\lambda_1 - \lambda_2} \tag{7.41}$$

Substituting the value of n_1 from Eq. (7.39), we get

$$2d = \frac{\lambda_1\lambda_2}{\lambda_1 - \lambda_2} \tag{7.42}$$

We can write $\lambda_1\lambda_2 = \lambda^2$ since $\lambda_1 \approx \lambda_2 \approx \lambda$

$$\therefore \quad 2d = \frac{\lambda^2}{\lambda_1 - \lambda_2} = \frac{\lambda^2}{\Delta\lambda} \tag{7.43}$$

Substituting $\Delta\lambda = \lambda_1 - \lambda_2$, in the above equation, we get,

$$\text{i.e,} \quad \Delta\lambda = \frac{\lambda^2}{2d} \tag{7.44}$$

Knowing the mean value of d for more than 15 successive positions of maximum distance, the value of the difference in wavelength is determined using the relation Eq.(7.44).

(iv) Standardisation of a Metre In order to standardise the length of 1 metre, an optical device known as etalon is used. In 1895, Michelson and Benoit used the cadmium red line to standardise the length of 1 metre. An optical device called an etalon has been designed with different sizes, covering ranges from 0.390625 mm to 10 cm.

7.8.4 Etalon

A schematic representation of an etalon is shown in Fig.7.15. It consists of two optically plane glass plates G_1 and G_2 . The front surface of the glass plates is coated

The McGraw·Hill Companies

with optically semi-silvered or semi-alimunised coating. These two glass plates are separated by a distance of *l* m. The screw attachment with the back of the plates helps to adjust for parallelism. The experimental procedure to standardise *l* m consists of two parts, namely,

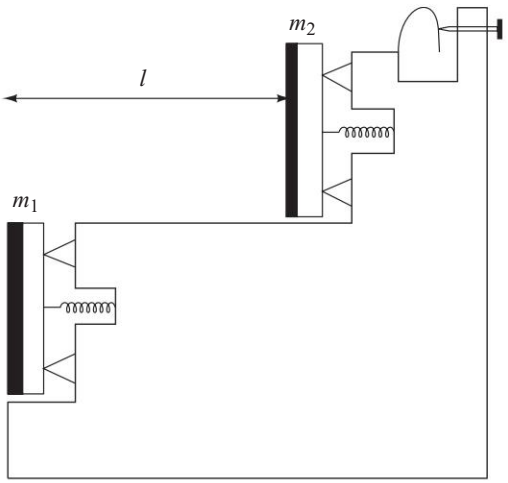


Fig. 7.15 *Etalon*

(i) determination of the number of fringes, and
(ii) calibration of *l* m by comparing the length of the shorter etalon with the longer etalon. After comparison, the length of the etalon is calibrated in terms of cadmium wavelength. Therefore, the length of 1 metre is calibrated in terms of red, green and blue lines of cadmium wavelengths.

Experiment A schematic representation of Michelson's interferometer used to standardise the length of 1 metre is shown in Fig. 7.16. It consists of a source *S*, a lens *L*, a semi-silvered mirror M_1, 100 % reflecting mirrors M_2, M_3 and M_4, and a telescope *T*. A monochromatic beam of light from the source *S* is made to fall on a semi-silvered mirror M_1 after passing through a lens *L*. When a light been incident on M_1, a part of the beam gets reflected towards the mirror M_1 while the other part of the light is transmitted towards the mirror M_2. The reflected light from M_2 is made to fall on the mirror M_4 and the two mirrors m_1 and m_2 of the first and second etalon respectively. The mirror M_4 is fixed, whose plane of axis is given as *x*. The first etalon mirror m_1 is placed in the same axis as that of M_4, while the second etalon mirror m_2 is placed just above the mirrors m_1 and M_4. In order to obtain the interference pattern, initially the plane of the mirrors M_4, M_1 and M_2 are made parallel.

Consider that a small air gap is formed between plane *x* and mirrors M_1 and M_2. As a result, circular fringes are formed. On the other hand, straight-line fringes are formed when the plane *x* of the mirror m_1 is slighting inclined to the plane of the mirror M_2. The observed straight-line fringes are shown in Fig. 7.17 a.

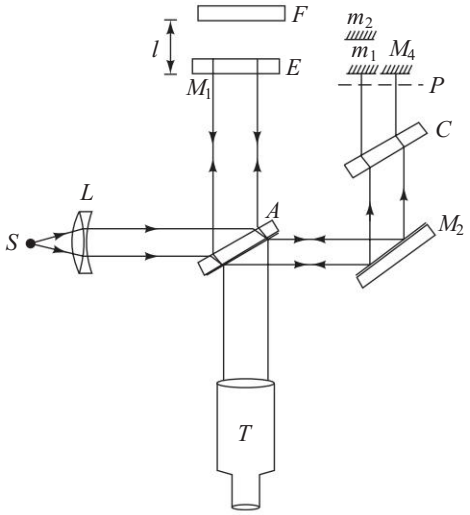


Fig. 7.16 *Michelson's interferometer—standardisation of 1 metre*

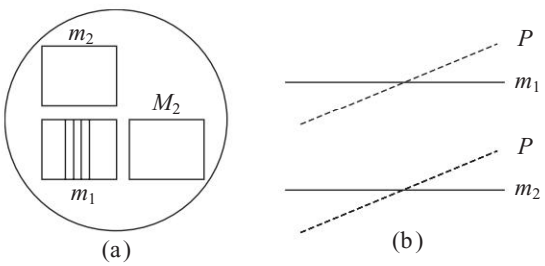
The white-light source is replaced by a cadmium light source. The position of the mirror M_4 is adjusted to make its plane parallel to the plane of x of the mirror M_1 and slightly crossed to mirrors M_1 and M_2. The interference pattern formed under this condition is made of circular fringes, as shown in Fig. 7.17 d.

The crosswire in the telescope is focused to any one of the fringes. The number of fringes crossing the crosswire is counted by moving the mirror from position P_1 and P_2 until the plane of x intersects the plane of the mirror M_2. Straight-line fringes are formed as obtained while using white light. The observed straight-line fringes are as shown in Fig. 7.17 c. The number of circular fringes n which is crossing the field of view and the distanced moved by the mirror M_1 from P_1 to P_2 is noted.

Therefore,

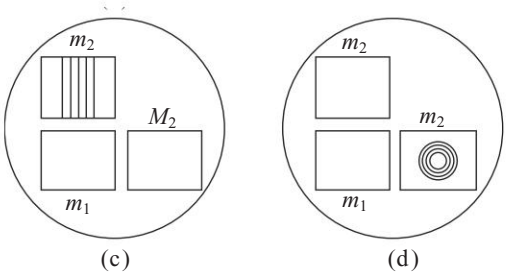
$$d = \frac{n\lambda}{2} \tag{7.45}$$

Equation (7.45) is used to measure separation between the mirrors m_1 and m_2 of the etalon. Michelson and Benoit have determined the number of half-wavelengths of cadmium red line as 1212.



(a) Straight line fringes in M_2 (b) Etalon m_1 and m_2 and the mirror M_2

The McGraw·Hill Companies



(c) Straight line fringes in m_1 (d) Circular fringes at m_2

Fig. 7.17 *Different types of fringe pattern—etalons m_1 and m_2 and the mirror M_2*

In order to calibrate 1-metre wavelength, first the length of the shortest etalon to the longest etalon is compared. Consider two etalons, namely, m_1m_2 is the shortest length while etalons m_3m_4 are the next longest. These two etalons are arranged side by side as shown in Fig. 7.18 (a). White light is used to compare the length of any two successive etalons. Hence, monochromatic light is replaced by white light. The plane of the mirrors M_1 and M_3 are slightly crossed to the plane of mirror M_1 as shown in Fig. 7.18 a. This results in straight-line fringes as observed in Fig. 7.18 (a). The telescope crosswire is adjusted to coincide with the central dark fringe of the mirror M_1. Straight the line fringes are observed as shown in Fig 7.18 (b), when the plane of the mirror M_2 intersects the plane of the mirror M_1 by moving the mirror from P_1 to P_2. The dark fringes are adjusted to coincide with the crosswire by moving the screw on the mirror M_1. The etalon m_1m_2 is moved backward by keeping the mirror M_1 in a fixed position, until it intersects the mirror M_1. To plane of the mirror M_1, straight-line fringes are observed as shown in Fig. 7.18 c. The mirror M_2 is adjusted until the plane of mirrors M_2 and M_4 intersect the plane of the mirror M_1. Hence, the central dark fringe coincides with the crosswire of the telescope. When the length of etalon m_3m_4 is twice as that of etalon m_1m_2, it gives a straight-line fringe in the field of view of mirrors M_2 and m_4 as shown in Fig. 7.18 (d). On the other hand, the etalon m_3m_4 is ending twice as that of m_1m_2 and hence the straight-line fringes are not seen.

(a) Straight-line fringes at M_1 and M_2 (b) Straight-line fringes at M_2

(c) Straight-line fringes at M_1 (d) Straight-line fringes at M_2 and M_4

Fig. 7.18 *different types of interference pattern—comparison of etalons m_1m_2 and m_3m_4*

Similarly, the experiment is repeated by comparing with the next higher wavelength of etalons like m_3m_4 with m_5m_6 and m_5m_6 with m_7m_8 up to the longest etalon of 10 cm. The longest etalon (10 cm) is compared with the standard metre. The results of wavelengths of red, green and blue lines of cadmium are given in Table 7.2. It is interesting to note that isotope of Hg^{198} of wavelength 5460.7532 Å is sharp in nature and is used to standardise the standard metre.

Table 7.2 *Wavelength of the cadmium lines and its equivalent of 1 metre*

S.No	Cadmium line colour	Wavelength Å	Equivalent value of 1 metre
1.	Red	6438.4722	1553, 163.5 λ_R
2.	Green	5085.8240	1966, 249.7 λ_G
3.	Blue	4799.9107	2083, 372.1 λ_B

Keypoints to Remember

- The formation of maximum intensity due to the superposition of two light energies is known as a constructive interference pattern. It is formed due to the overlapping of the crest of one wave on the crest of other waves.
- The formation of minimum intensity due to the overlap of the crest of one wave on the trough of other waves is known as destructive interference pattern.
- The wedge-shaped air film obtained in between two glass plates is known as air wedge.
- The condition for bright fringe is $2\mu t \cos r = \left(\frac{2n+1}{2}\right)\lambda$.

The McGraw·Hill Companies

- The condition for dark fringe is $2\mu t \cos r = n\lambda$.
- An air-wedge experiment is used to test the flat surfaces of glass plates and thickness of a thin sheet of paper.
- In the air-wedge experiment, the thickness of the paper is determined using the equation $t = \frac{\lambda l}{2\beta}$.
- The material which is on the surface of glass plate minimises the reflection of light from its surface and is known as anti-reflection coating or non-reflecting film.
- In optical instruments, the anti-reflection coating is used to control light reflectance.
- The structure of refractive indices of a coating material to obtain high-pass filter is g (0.5L) $(HL)^m$ (0.5L)a.
- The structure of refractive indices of a coating material to obtain low-pass filter is g(0.5H) L $(HL)^m$ (0.5 H) a
- A Michelson's interferometer's working based on the principle of interference of light.
- In Michelson's experiment, the wavelength of monochromatic light is determined using the equation $\lambda = \frac{2d}{n}$.
- In Michelson's experiment, the refractive index is determined using the relation $\mu = \frac{n\lambda}{2t} + 1$.
- In Michelson's interferometer, the two spectral lines of nearly equal wavelength are distinguished with more resolution using the relation $\Delta\lambda = \frac{\lambda^2}{2d}$.
- Etalon is an optical device used to standardise the length of 1 metre.

Solved Problems

Example 7.1 A wedge-shaped air film is formed between two glass plates by placing a paper at one of the sides. On illuminating this film by a light of 6000 Å wavelength, 10 fringes are seen in one cm. If the light is incident normally, find the angle of the wedge.

Given Data

Wavelength of the light used, $\lambda = 6000 \times 10^{-10}$ m

Fringe width, $$\beta = \frac{1}{10} \text{ cm} = 0.1 \text{ cm}$$

$$= 1 \times 10^{-3} \text{ m}$$

Solution We know that, the angle of the air wedge,

$$\theta = \frac{\lambda}{2\beta}$$

Substituting the values of λ and β, we get

$$= \frac{6000 \times 10^{-10}}{2 \times 1 \times 10^{-3}}$$

$$= 0.12$$

The angle of the air wedge is 0.12 radian.

Example 7.2 Interference fringes are produced with monochromatic light falling normally on a wedge-shaped air film. The angle of the wedge is 40 seconds of an arc and the distance between the successive fringes is 0.125 cm. Calculate the wavelength of the light.

Given Data

Angle of the wedge, $\theta = 40$ s
Fringe width, $\beta = 0.125 \times 10^{-2}$ m

Solution We know that the fringe width is given by,

$$\beta = \frac{\lambda}{2\theta}$$

Therefore,

$$\lambda = 2\beta\theta$$

Substituting the values of β and θ, we get,

$$= 2 \times 0.125 \times 10^{-2} \times 40 \times (\pi/180) \times (1/60) \times (1/60)$$

$$= 4.848 \times 10^{-7} \text{ m}$$

The wavelength of the light used is 4848 Å.

Example 7.3 A shift of 125 circular fringes is observed when the movable mirror of a Michelson's interferometer is moved by 0.04 mm. Calculate the wavelength of the light.

Given Data

Number of fringes observed, $n = 125$
Distance moved, $d = 0.04 \times 10^{-3}$ m

Solution We know that the condition for interference is $2d = n\lambda$

$$\lambda = \frac{2d}{n}$$

Substituting the values, we get,

$$= \frac{2\times0.04\times10^{-3}}{125}$$

$$= 6.4 \times 10^{-7} \text{ m}$$

The wavelength of the light used is 6400 Å.

Example 7.4 A shift of 200 circular fringes is observed when the movable mirror of a Michelson's interferometer is moved by 0.589 mm. Calculate the wavelength of the light

Given Data

Number of fringes observed, $n = 200$
Distance moved, $d = 0.589 \times 10^{-3}$ m

Solution We know that, the condition for interference is $2d = n\lambda$

$$\lambda = \frac{2d}{n}$$

Substituting the values, we get,

$$= \frac{2\times0.589\times10^{-3}}{200}$$

$$= 5.89 \times 10^{-7} \text{ m}$$

The wavelength of the light used is 5890 Å.

Example 7.5 A shift of 200 circular fringes is observed when the movable mirror of the Michelson's interferometer is moved by 0.06 mm. Calculate the wavelength of the light.

Given Data

Number of fringes observed, $n = 200$
Distance moved, $d = 0.06 \times 10^{-3}$ m

Solution We know that the wavelength of light is

$$\lambda = \frac{2d}{n}$$

Substituting the values, we get,

$$= \frac{2\times0.06\times10^{-3}}{200}$$

$$= 6.0 \times 10^{-7} \text{ m}$$

The wavelength of the light used is 6000 Å.

Example 7.6 A glass plate with a refractive index of 1.5 is introduced in between the glass plate G_2 and the mirror M_2 in a Michelson's interferometer. If the distance between any two dark fringes of zero path difference is 1.24 mm, find the thickness of the glass plate.

Given Data

The McGraw·Hill Companies

Refractive index of the plate $\mu = 1.5$
Distance between dark fringes of zero path difference $d = 1.24 \times 10^{-3}$ m

Solution We know that the distance between dark fringes of zero path difference is

$$d = (\mu - 1)t$$

Therefore, the thickness of the glass plate is

$$t = \frac{d}{\mu - 1}$$

$$t = \frac{1.24 \times 10^{-3}}{1.5 - 1}$$

$$= 2.48 \text{ mm}$$

The thickness of the glass plate is 2.48 mm.

Example 7.7 1000 circular fringes collapsed at the centre when a movable mirror of Michelson's interferometer is moved by 0.3165 mm. Calculate the wavelength of the monochromatic source of light.

Given Data

Number of fringes observed $n = 1000$
Distance moved, $d = 0.3165 \times 10^{-3}$ m

Solution We know that, the wavelength of light is $2d = n\lambda$

$$\lambda = \frac{2d}{n}$$

Substituting the values, we get,

$$= \frac{2 \times 0.3165 \times 10^{-3}}{1000}$$

$$= 6.33 \times 10^{-7} \text{ m}$$

The wavelength of the monochromatic source light is 6330 Å.

Example 7.8 Calculate the distance through which the mirror of Michelson's interference has to be moved between two consecutive positions of maximum distinctness of D_1 and D_2 of sodium lines. Given λ for D_2 line is 5890 Å and λ for D_1 line is 5896 Å.

Given Data

Wavelength of D_2 line, $\lambda_1 = 5890$ Å
Wavelength of D_1 line, $\lambda_2 = 5896$ Å

Solution We know that the resolution of two spectral lines is

$$\lambda_1 - \lambda_2 = \frac{\lambda_1 \lambda_2}{2d}$$

Therefore, $$2d = \frac{\lambda_1 \lambda_2}{\Delta\lambda}$$

Substituting the values of λ_1 and λ_2, we get,

$$d = \frac{5890 \times 5896 \times 10^{-20}}{2 \times 6 \times 10^{-10}}$$

$$= 2.894 \times 10^{-4} \text{ m}$$

The two spectral lines of the sodium lines are resolved, when d is 0.2894 mm.

Example 7.9 A thin wire used in an air-wedge experiment produces straight-line parallel fringes having a mean fringe width of 1.5 mm with sodium light of 589.3-nm wavelength. If the length of the air wedge is 4.5 cm, find the thickness of the wire used.

Given Data

Mean fringe	$\beta = 1.5 \times 10^{-3}$ m
Wavelength of the light	$\lambda = 589.3 \times 10^{-9}$ m
Length of the air wedge	$l = 4.5 \times 10^{-2}$ m

Solution We know that, the thickness of the wire $t = \frac{\lambda l}{2\beta}$

Substituting the values, we get,

$$= \frac{589.3\times10^{-9} \times 4.5\times10^{-2}}{2\times1.5\times10^{-3}}$$

$$= 8.8395 \times 10^{-6} \text{ m}$$

The thickness of the wire is 8.84 μm.

Example 7.10 The movable mirror of a Michelson's interferometer in moved a distance of 0.02603 mm. Find the number of fringes shifted across the crosswire if the wavelength used is 5206Å wavelength.

Given Data

Distance moved	$d = 0.02603 \times 10^{-3}$ m
Wavelength of the light	$\lambda = 5206 \times 10^{-10}$ m.

Solution We know that the condition for interference is i.e., the number of fringes shifted

$$n = \frac{2d}{\lambda}$$

Substituting the values, we get,

$$= \frac{2\times0.02603\times10^{-3}\, m}{5206\times10^{-10}}$$

$$= 100$$

The number of fringes shifted across the cross wire is 100.

Objectives Questions

7.1 When the crest of one wave overlaps with the crest of another wave, it gives _____ pattern.

7.2 _____ interference patterns is formed due to the overlap of the crest of one wave on the trough of another wave.

7.3 The condition for bright fringe is _____

a) $2\mu t \cos r = n\lambda$ b) $2\mu t \sin r = \frac{(2n+1)}{2}\lambda$

c) $2\mu t \sin r = n\lambda$ d) $2\mu t \cos r = \frac{(2n-1)}{2}\lambda$

7.4 The condition for dark fringe is _____

a) $2\mu t \sin r = n\lambda$ b) $2\mu t \cos r = n\lambda$

c) $2\mu t \sin r = \frac{(2n+1)}{2}\lambda$ d) $2\mu t \cos r = \frac{(2n-1)}{2}\lambda$

7.5 _____ is a wedge-shaped air film formed in between the two glass plates.

7.6 In the air-wedge experiment, the thickness of the paper is determined using the equation

a) $t = \lambda l/2\beta$ b) $t = 2\lambda l/\beta$

c) $t = \lambda l/\beta$ d) $t = 2\beta l/T$

7.7 The plainness of glass-plate is tested using air-wedge experiment. (True/False)

7.8 Anti-reflection coating is used to _____ the reflection of light from its surface.

7.9 For destructive interference, the path difference is equal to _____

a) 2λ b) $2\lambda/4$

c) $\lambda/4$ d) $\lambda/2$

7.10 For normal incidence, the value of $\cos r$ = _____

7.11 High-quality camera lenses are obtained using _____ anti reflection coating

7.12 In anti-reflection coating, the multilayer coating consists of different materials with same refractive indices. (True/False)

7.13 The device used to obtain high transmittance at high frequency is known as _____ filter.

7.14 The device used to obtain high transmittance at low frequency is known as_____ filter.

7.15 The structure of refractive index of coating for a high-pass filter is _____.

a) g $(0.5L)^m$ (HL) (0.5L)a b) g (0.5L) $(HL)^m$ (0.5L)a

c) g (HL) $(0.5L)^m$a d) g $(0.5L)^m$ $(HL)^m$a

7.16 The structure of refractive index of coating for a low-pass filter is _____.

a) g $(0.5H)^m$ $(HL)L^m$ (0.5H)a b) g (0.5H) (HL) L^m (0.5H)a

c) g (0.5L) H $(HL)^m$ (o.5H)a d) g (0.5H) $L(HL)^m$ (0.5H)a

7.17 _____ principle is used in Michelson's interferometer.

The McGraw·Hill Companies

7.18 The wavelength of monochromatic light is fixed. (True/False)

7.19 In Michelson's interferometer, the wavelength of monochromatic light is determined using the relation

a) $\lambda = 2d/n$ b) $\lambda = n/d$

c) $\lambda = 2n/d$ d) none of these

7.20 In Michelson's interferometer, the refractive index of a given thin medium is determined using the relation

a) $\mu = \frac{n\lambda}{a} + 1$ b) $\mu = \frac{n\lambda}{3t} + 1$

c) $\mu = \frac{n\lambda}{2t} + 1$ d) $\mu = \frac{2\lambda}{nt} + 1$

7.21 The two spectral lines of nearly equal wavelength are distinguished with more resolution using the relation in Michelson's interferometer,

a) $\Delta = \lambda^2/d$ b) $\Delta\lambda = \lambda^2/d$

c) $\lambda = 2\lambda^2/d$ d) $\Delta\lambda = \lambda^2/d$

7.22 The colours of cadmium lines are _____, _____ and _____.

7.23 The wavelength of cadmium lines red, green and blue are respectively _____, _____ and _____.

KEY

7.1	constructive interference	7.2	destructive
7.3	d	7.4	b)
7.5	Air wedge	7.6	a)
7.7	True	7.8	Minimise
7.9	d)	7.10	1
7.11	Multi layer	7.12	False
7.13	high pass	7.14	low pass
7.15	b)	7.16	d)
7.17	Interference of light	7.18	True
7.19	a)	7.20	c)
7.21	d)	7.22	Red, Green and Blue
7.23	6438.4722, 5085.8240 and 4799.9107		

Exercises

7.1 An air-wedge experiment is used to find the thickness of a thin sheet of paper using a monochromatic light of 5893 Å wavelength falling normally on it. If the length of the glass plate is 4.6 cm and 10 fringes are observed in 0.8 cm, find the thickness of the paper.

7.2 Two glass plates are separated with a thin wire and illuminated by light of 5893 Å wavelength falling normally on it. If the angle of the wedge is 0.024

radian, at what distance from the edge of the wedge will the 15th fringe be observed by reflected light?

7.3 In an experiment using a Michelson's interferometer, a mica sheet of 1 mm thickness is inserted between the glass plate G_2 and the mirror M_2. If the wavelength of the light used is 6000 Å, find the refractive index of the mica sheet. The distance moved by the micrometer screw between two adjacent dark fringes of zero path difference using white light is 0.8 mm.

7.4 Calculate the distance between the successive positions of a movable mirror of a Michelson's interferometer giving best fringes in the case of the spectral lines of wavelengths of 5500 Å and 5520 Å.

Short–Answer Questions

7.1. What is interference?
7.2. What is constructive interference?
7.3. Define destructive interference.
7.4. What is meant by an air wedge?
7.5. What is an air film?
7.6. Differentiate reflections from refractions.
7.7. Write the condition to obtain bright fringes.
7.8. Write the condition to obtain dark fringes.
7.9. What are the advantages of the air-wedge experiment?
7.10. Explain how the flat surfaces tested using air-wedge experiment.
7.11. If the wavelength of the light used is double, what may be the nature of the thickness of the specimen in the air wedge?
7.12. Mention the conditions to produce destructive and constructive interference in thin films.
7.13. What do you mean by anti-reflection coating?
7.14. Mention the application of anti-reflection coating.
7.15. What is the role of refractive index of materials in anti-reflection layers?
7.16. Differentiate the high and low-pass filter.
7.17. Explain how a Michelson's interferometer is used to find the refractive index of a transparent medium.
7.18. Differentiate monochromatic from polychromatic light.
7.19. What are the applications of Michelson's interferometer?
7.20. Explain how a Michelson's interferometer is used to resolve spectral lines.
7.21. How will you calculate the wavelength of the given source of light using a Michelson's interferometer?
7.22. Explain the localized fringes produced by a Michelson's interferometer
7.23. Under what conditions are circular fringes formed in a Michelson's interferometer?
7.24. What type of fringes will be formed if white light is used in a Michelson's interferometer?
7.25. Explain why a monochromatic light is used in Michelson's interferometer.
7.26. What is etalon?
7.27. Mention few applications of etalons.

The McGraw·Hill Companies

Descriptive Questions

7.1 Derive an expression for the path difference produced in a thin film of thickness t and of refractive index μ and hence obtain the condition for constructive interference.

7.2 Deduce an expression for an air wedge to find the thickness of a thin sheet of paper and hence explain an experimental method to find the thickness of a thin sheet of paper using air-wedge experiment.

7.3 Explain the construction and working of a Michelson's interferometer and hence explain how it is used to find the refractive index of a transparent layer.

7.4 Describe the construction and working of a Michelson's interferometer and also explain with necessary theory, a laboratory experiment of finding the wavelength of a monochromatic source of light using Michelson's interferometer?

7.5 Illustrate the different types of fringes formed in a Michelson's interferometer and the corresponding positions of their mirrors.

7.6 Explain the standardization of 1 metre using Michelson's interferometer.

Chapter

Polarisation

8

OBJECTIVES

- To understand the nature of polarisation.
- To study the different types of polarisation of light
- To study the role of calcite crystal and Nicol prism in polarisation and its limitations
- To derive the mathematical expression for polarisation mechanism
- To understand quarter wave plate and half wave plate for double refraction process.
- To understand the production and detection of plane, circularly and elliptically polarised light

8.1 INTRODUCTION

Light exhibits dual nature. Certain properties of light, like the photoelectric effect, Compton effect, etc., are explained based on the particle nature of light. Some other properties such as interference, polarisation, etc., are explained based on the wave nature of light. Therefore, light is said to possess dual nature, namely, the particle nature and the wave nature.

There are two types of wave motion, namely, longitudinal and transverse wave motions. In longitudinal motion, the particles in the medium vibrate about their mean position along the direction of propagation of the wave. On the other hand, in transverse wave motion, the particles in the medium vibrate perpendicular to the direction of propagation of the wave. Sound waves exhibit longitudinal wave motion, whereas light waves exhibit transverse wave motion. An ordinary light has vibrations in all directions. The vibration of an ordinary light is symmetrically distributed over the plane of vibration and it is represented by a double arrow head and a dot (↕). The polarisation property of light is used to identify the transverse wave motion of light.

Let us discuss the principles of wave motion, polarisation of light, its production and detection in details in the following section.

The McGraw·Hill Companies

8.2 FUNDAMENTALS OF POLARISED LIGHT

8.2.1. Transverse Wave Motion

A simple mechanical experiment which demonstrates the transverse wave motion is shown in Fig. 8.1. Consider a string AB is passing through two thin slits S_1 and S_2. The two slits are parallel to each other. Let one end of the string B be fixed. If one end of the string is vibrated, then the vibration produced at one end passes up to the other end B of the string. Thus, the string vibrates between the ends A and B.

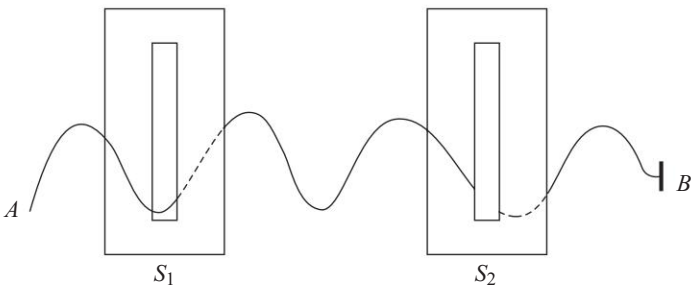


Fig. 8.1 *Vibration of a string between two parallel slits*

Let us consider that the second slit is rotated through an angle of 90º and it is kept perpendicular to the first slit S_1, as shown in Fig. 8.2. If the string is fixed at B and it is vibrated at A, the string vibrates up to the slit S_2. Between the slit S_2 and the fixed end B, there is no vibration of the string. This shows that the string is vibrating transversely. The wave in which the vibration is perpendicular to the direction of propagation of the wave is said to be a transverse wave. Similar observation can be made when a ray of light is passed through a tourmaline crystal.

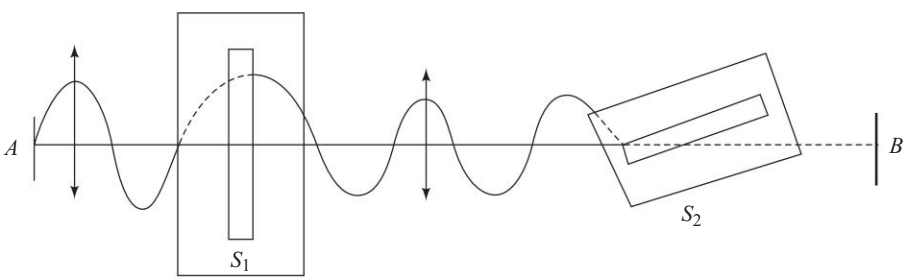
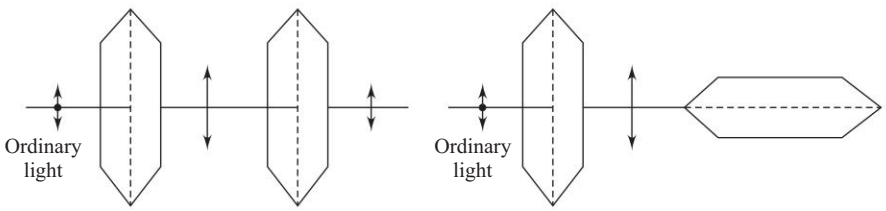


Fig. 8.2 *Vibration of strings between two perpendicular slits*

8.2.2 Polarisation of Light

Consider an ordinary light is passed through a tourmaline crystal as shown in Fig. 8.3. The light emerging from the tourmaline crystal is found to be slightly

coloured. If the tourmaline crystal is rotated by taking the incident beam as an axis, there is no variation in the intensity of the emergent beam. The tourmaline crystal has the property of allowing the components of light that are vibrating parallel to its axis to pass through it, and it absorbs all the other components of the light that are not parallel to the axis of the crystal. The emergent beam from the tourmaline crystal vibrates in only one direction, i.e., it vibrates parallel to the axis of the crystal. The light rays that are vibrating in only one direction are known as *polarised light*. The device that produces the polarised light is called as *polariser*. Consider that the emergent beam from the first tourmaline crystal is passed through another tourmaline crystal kept perpendicular to the first crystal. No light is emerging from the second tourmaline crystal.



(a) Parallel motion (b) Perpendicular motion

Fig. 8.3 *Polarisation—tourmaline crystal*

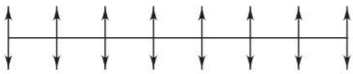
Consider the second tourmaline crystal is rotated by taking the incident beam as an axis. The intensity of the emergent beam from the second tourmaline crystal varies. The intensity of the emergent beam from the second tourmaline crystal has two times the maximum intensity and two times the minimum intensity within one complete rotation. When the second tourmaline crystal is parallel to the first one, the emergent beam from the second crystal is maximum. The intensity of the emergent beam from the second crystal is minimum when these two crystals are perpendicular to each other. The second tourmaline crystal is used to analyse the polarised light. Therefore, the second crystal is said to be an analyser. This property of light ray confirms the transverse nature of light waves.

There are three different types of polarised lights, as given below.

(i) Plane polarised light
(ii) Circularly polarised light, and
(iii) Elliptically polarised light

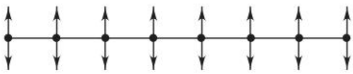
8.2.3 Pictorial Representation of Light

The plane polarised light, vibrating parallel to the plane of this paper, is represented as shown in Fig. 8.4 a. When the plane polarized light is vibrating perpendicular to the plane of this paper, it is represented as shown in Fig. 8.4 b. The unpolarised or ordinary light has vibrations that are both parallel and perpendicular to the plane of the paper and it is represented as shown in Fig. 8.4 c



(a) Plane polarised light– parallel to plane of this paper

(b) Plane polarised light– perpendicular to the plane of pape



(c) Unpolarised light

Fig. 8.4 *Representation of plane polarised light*

8.2.4 Plane of Vibration and Plane of Polarisation

The plane in which the light is vibrating is known as the *plane of vibration*, and the plane perpendicular to the plane of vibration is known as the *plane of polarisation* and is shown in Fig. 8.5

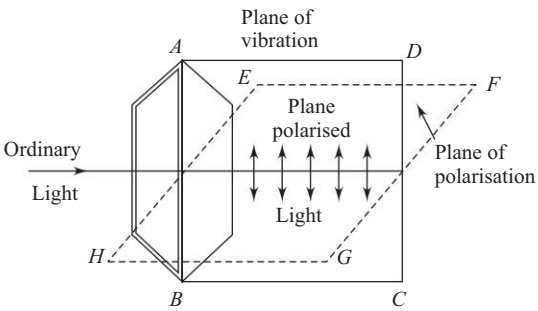


Fig. 8.5 *Plane of vibration and polarisation*

In Fig. 8.5, the emergent beam vibrates parallel to the axis of the polariser. The plane *ABCD* is parallel to the axis of the polariser and hence, it is known as plane of vibration. The plane perpendicular to the plane of vibration is known as plane of polarisation. The plane *EFGH* is perpendicular to the axis of the polariser. It should be noted that no vibration takes place in this plane *EFGH*.

8.2.5 Geometry of Calcite Crystal

Calcite is a transparent colourless crystal. It is chemically known as hydrated calcium carbonate ($CaCO_3$). It can be reduced into a rhombohedron by cleavage or breakage. The rhombohedron-shaped calcite crystal is shown in Fig. 8.6. It consists of six faces. These six faces are parallelograms having angles of 102º and 78º. Strictly speaking, the angles of the parallelograms are 101º55′ and 78º5′. The corners *C* and *F* consist

The McGraw·Hill Companies

of three obtuse angles and hence, these corners are said to be blunt corners. The other corners of the crystal consist of two acute and one obtuse angle.

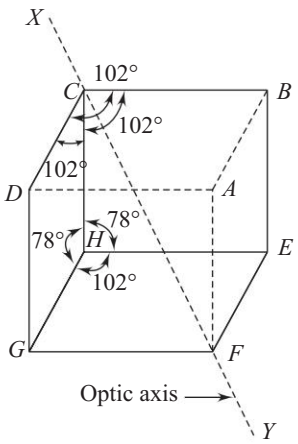


Fig. 8.6 *Calcite crystal—Rhombohedron*

8.2.6 Optic Axis

A line joining any two blunt corners is known as the *optic axis* of the crystal. An optic axis is not an axis, but it is a direction in the crystal. Whenever the light rays are passing through the optic axis, there is no splitting of light into two. Since optic axis is a direction, any parallel direction (line) to the optic axis is also known as optic axis.

8.2.7 Uniaxial and Biaxial Crystals

Certain crystals like calcite, quartz, tourmaline, etc., have only one optic axis. Hence, they are called uniaxial crystals. The crystals like borax, mica and selenite possess two optic axes and hence, they are called biaxial crystals.

8.2.8 Principal Section and Principal Planes

A plane containing the optic axis and perpendicular to two opposite faces of the crystal is known as the *principal section* of the crystal. The principal section of the calcite crystal is a parallelogram having angles of 71º and 109º. The principal plane of the ordinary ray is the plane containing the optic axis and the ordinary ray. The plane consisting of the extraordinary ray and the optic axis of the crystal is known as the *principal plane* of the extraordinary ray.

8.3 NICOL PRISM

It is an optical device made from calcite and is used to produce and analyse plane polarised light. It was invented by William Nicol, a Scotch physicist, in 1828, and is known as *Nicol prism*. It is based on the principle of double refraction exhibited by

calcite. The Nicol prism is designed in such a way that it can transmit the extraordinary rays and cut off the ordinary rays. The emergent beam from a Nicol prism is a plane polarised light.

8.3.1. Construction

The Nicol prism is constructed using a calcite crystal having a length three times as that of its breadth. Consider a calcite crystal *AB′CDEFGH′* in which *B′* and *H′* are blunt corners. It is rhombohedron shaped and is shown in Fig. 8.7. The plane *B′DH′F* is known as the principal section of the crystal since it is perpendicular to the planes *ABCD* and *EFGH′*, and it consists of the optic axis. The plane *B′DH′F* is a parallelogram and its angles are 71º and 108º. The angles *B′DH′* and *B′FH′* are 71º, The upper and lower faces of the crystal, *AB′CD* and *EFGH′*, are grounded in such a way that the angles *B′DH′* an *B′FH′* are reduced from 71º to 68º. Let *ABCD* and *EFGH* be the new upper and lower faces of the crystal. The new principal section is *BDHF*. The crystal is cut into two along the plane *JHKB*.The plane *JH′ KB′* is perpendicular to the principal section *B′ DH′ F* of the crystal. The two cut surfaces are highly polished into optical flatness and then cemented together by means of Canada balsam.

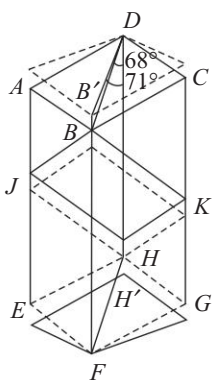


Fig. 8.7 *Calcite rhombohedron*

The refractive indices of Nicol prism for sodium light of wavelength $\lambda = 5893$ Å are 1.55 (μ_{CB}), 1.65837 (μ_o) and 1.48641 (μ_e), respectively, for Canada balsam layer, ordinary rays and extraordinary rays. The two end faces of the crystals are kept open, while the sides are blackened so as to absorb the light by total internal reflections.

8.3.2 Action of a Nicol Prism

Consider a ray of light *SM* is made incident on the front face *DB′* of the Nicol prism. The incident beam is split into two beams, namely, ordinary and extraordinary rays.

The McGraw·Hill Companies

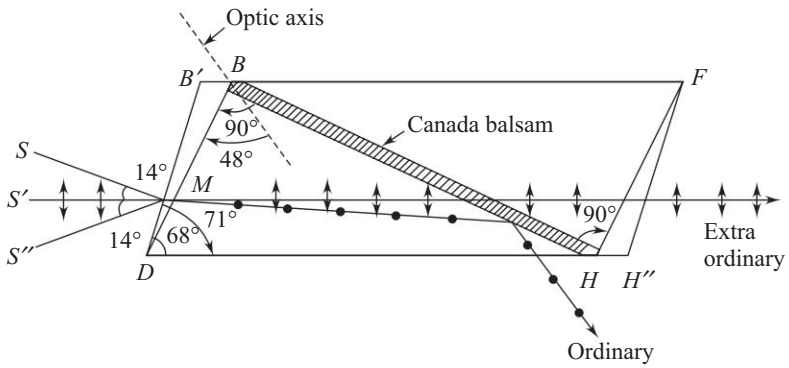


Fig. 8.8 *Nicol prism*

The Canada balsam layer is a rarer medium for the ordinary ray. Therefore, the ordinary ray is travelling from a denser (calcite) to a rarer (Canada balsam) medium. Hence, it undergoes total internal reflection, if the angle of incidence made by the ordinary ray at the Canada balsam layer is greater than the critical angle. The critical angle for the ordinary ray for calcite to Canada balsam layer is

$$\theta_c = \sin\left[\frac{1.55}{1.658}\right] \tag{8.1}$$

$$= 69.2^{\circ}$$

In order to make the angle of incidence made by the ordinary ray at the calcite and Canada balsam interface more than the critical angle, the angle of the upper and bottom faces of the crystal is grounded from 71° to 68° and the length of the crystal is taken as three times as that of its breadth. For the extraordinary rays, the Canada balsam layer is a denser medium. Therefore, it travels from a rarer (calcite) to a denser (Canada balsam) medium, and hence it travels in a straight line and emerges out from the Nicol prism. The emergent beam is a plane polarised light.

8.3.3 Limitations of Nicol Prism

The Nicol prism acts as a polariser only for the light rays having slightly convergent and slightly divergent beams. If the incident beam is more divergent or more convergent, the Nicol prism does not act as a polariser.

Let the incident beam be divergent and making a larger angle with the front face *DB′*. In this case, the extraordinary ray will become parallel to the optic axis of the crystal, and hence it travels alongwith the ordinary ray. It is also total internal reflected back by the Canada balsam layer. Therefore, no light emerges from the Nicol prism.

Let the incident beam be convergent and making a smaller angle with the face *DB′*. The ordinary ray makes an angle less than the critical angle at the Canada balsam layer. Therefore, the ordinary ray is also transmitted with the extraordinary rays. This limits the angle of incidence for the crystal to 14°. The angle between the extreme incident rays is limited to 28°, so as to avoid the transmission of the ordinary ray with the extraordinary rays and to avoid the total internal reflection of the extraordinary rays. Hence, a Nicol prism cannot be used for highly convergent and divergent light beams.

The McGraw·Hill Companies

8.4 QUARTER WAVE PLATE

A Quarter Wave Plate (QWP) is a device from a doubly refracting uniaxial negative or positive crystal whose refracting faces are cut parallel to the direction of the optic axis. The thickness of this device is cut in such a way that it can produce a path difference of $\lambda/4$ or a phase difference of $\pi/2$ between ordinary and extraordinary rays.

Let μ_e and μ_o be the refractive indices of the extraordinary and ordinary rays, respectively. Let t be the thickness of the crystal. Then, for a negative crystal, the path difference between these two rays is $(\mu_o - \mu_e)\, t$. The QWP is cut in such a way that it can produce a path difference of $\lambda/4$.

$$\therefore \qquad (\mu_o - \mu_e)t = \lambda/4$$

i.e.
$$t = \frac{\lambda}{4(\mu_o - \mu_e)} \tag{8.2}$$

For a positive crystal, the path difference between the ordinary and extraordinary rays is $(\mu_e - \mu_o)t$. The thickness of a QWP made up of a positive crystal can be obtained by equating this path difference to $\lambda/4$.

i.e $\qquad (\mu_e - \mu_o)\, t = \lambda/4$

or
$$t = \frac{\lambda}{4(\mu_e - \mu_o)} \tag{8.3}$$

Equations (8.2) and (8.3) give the thickness of a QWP made up of a negative and a positive crystal, respectively.

Applications The following are some of the important applications of QWP:

(i) A QWP is used to produce circularly polarised light, if the optic axis of the crystal makes an angle of 45º with the incident beam of light.

(ii) A QWP is used to produce elliptically polarised light, if the incident beam makes an angle other than 45º with the optic axis of the crystal.

Limitation The QWP is designed for a particular wavelength. Hence, it is not useful for the other wavelengths.

8.5 HALF WAVE PLATE

A Half Wave Plate (HWP) is a doubly refracting uniaxial negative or positive crystal cut in such a way that it can produce a phase difference of π or a path difference of $\lambda/2$ between the ordinary and extraordinary rays.

Let μ_e and μ_o be the refractive indices of the crystal for the extraordinary and ordinary rays. For a negative crystal like calcite, the path difference between the ordinary and extraordinary rays is given by $(\mu_o - \mu_e)\, t$. Therefore, the thickness of the HWP required to produce a path difference of $\lambda/2$ is given by,

$$(\mu_o - \mu_e)t = \lambda/2$$

i.e.,
$$t = \frac{\lambda}{2\,(\mu_o - \mu_e)} \tag{8.4}$$

The McGraw·Hill Companies

For a positive crystal like quartz, the path difference is given by $(\mu_e - \mu_o)\, t$. The thickness of the crystal required to produce a path difference of $\lambda/2$ is

$$(\mu_e - \mu_o)t = \lambda/2$$

i.e
$$t = \frac{\lambda}{2\,(\mu_e - \mu_o)} \tag{8.5}$$

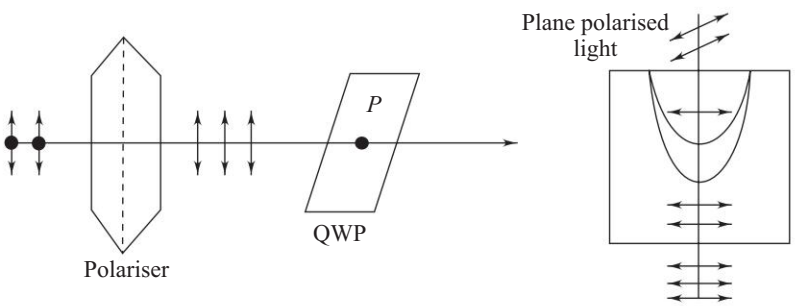
Applications Some of the applications of the HWP are given below:

(i) The HWP produces a phase difference of π between the ordinary and extraordinary rays.
(ii) The emergent beam from a HWP is a plane polarised light.
(iii) It rotates the azimuth of a beam of plane polarised light by 90º, provided the incident light makes an angle of 45º with the optic axis of the half wave plate.

Limitation The thickness of a HWP is cut in such a way that it can produce a path difference of $\lambda/2$ between the ordinary and extraordinary rays. Therefore, it is useful only for a particular wavelength.

8.6 THEORY OF POLARISED LIGHT

Consider that a plane polarised light produced by a tourmaline crystal is passed through a quarter wave plate whose faces have been cut parallel to the optic axis. The production of the plane polarised light and the particle vibrations are shown in Fig. 8.9. The incident plane polarised light is split into ordinary and extraordinary rays. These two rays travel in the same direction, but with different velocities, and hence there is a phase difference introduced between them.



(a) Production (b) Vibrations of light

Fig. 8.9 *Polarised light*

Let the plane polarised light incident on the quarter wave plate be represented as $y = A \sin \omega t$. A phase difference of $\pi/2$ is introduced between the ordinary and extraordinary rays due to the introduction of a quarter wave plate. Therefore, the emergent beams can be represented as,

$$x = A \sin \omega t \cos \theta$$

and

$$y = A \sin \omega t \sin \theta \tag{8.6}$$

where θ is the angle made by the incident beam with the optic axis of the quarter wave plate.

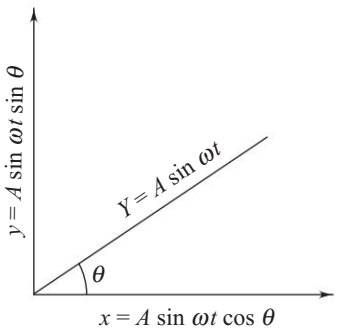


Fig. 8.10 *Incident and emergent beams*

The ordinary and extraordinary rays travel with different velocities in the same direction, and hence a phase difference is introduced between these two rays. Let d be the phase difference between the ordinary and extraordinary rays. The emergent beams from the quarter wave plate can be represented as,

$$x = A \cos\theta \sin(\omega t + \delta)$$

$$y = A \sin\theta \sin\omega t \tag{8.7}$$

Taking $A \cos\theta = a$ and $A \sin\theta = b$, the above equations can be written as,

$$x = a \sin(\omega t + \delta)$$

$$y = b \sin\omega t \tag{8.8}$$

Equation (8.8) can be writte as,

$$\frac{x}{a} = \sin(\omega\ t + \delta) \tag{8.9}$$

$$\frac{y}{b} = \sin\omega t \tag{8.10}$$

Equation (8.9) can be writte as,

$$\frac{x}{a} = \sin\omega t \cos\delta + \cos\omega t \sin\delta \tag{8.11}$$

Substituting the value of $\sin\omega t$ and $\cos\omega t$ from Eq. (8.10) in Eq. (8.11), weget

$$\frac{x}{a} = \frac{y}{b}\cos\delta + \sqrt{1-\frac{y^2}{b^2}}\ \sin\delta \tag{8.12}$$

Squaring Eq. (8.12), we get,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2}\cos^2\delta - 2\frac{xy}{ab}\cos\delta = \left(1-\frac{y^2}{b^2}\right)\sin^2\delta$$

Simplifying the above equation we, get,

The McGraw·Hill Companies

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-2\frac{xy}{ab}\cos\delta=\sin^2\delta \tag{8.13}$$

Equation (8.13) is a general equation for an ellipse.

Case I When $\delta = 0$, $\sin\delta = 0$ and $\cos\delta = 1$
Equation (8.13) beomes

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-2\frac{xy}{ab}=0$$

i.e.
$$\left(\frac{x}{a}-\frac{y}{b}\right)^2=0$$

or
$$\frac{x}{a}=\frac{y}{b}$$

$$y=\frac{b}{a}x \tag{8.14}$$

Equation (8.14) is an equation of a straight line. This shows that when the phase difference between the two rays is zero, the emergent light will be plane polarised. It is shown in Fig. 8.11a.

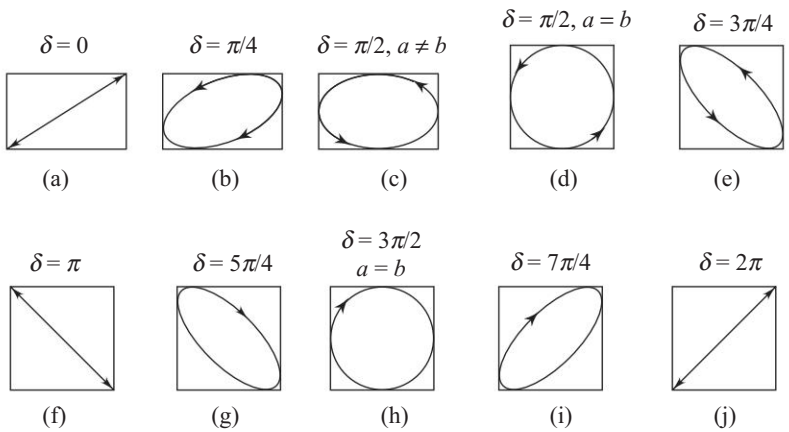


Fig. 8.11 *Production of polarised light–Different types*

Case II When $\delta = \pi/2$, $\cos\pi/2 = 0$, $\sin\pi/2 = 1$ and $a \neq b$.
Equation (8.13) beomes

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \tag{8.15}$$

Equation (8.15) represents the equation of a symmetrical ellipse. Therefore, the emergent beam is an ellipse provided $a \neq b$, and it is shown in Fig. 8.11c.

The McGraw·Hill Companies

Case III When $\delta = \pi$, $\cos \pi = -1$, $\sin \pi = 0$
Equation (8.13) beomes

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + 2\frac{xy}{ab} = 0$$

i.e.
$$\left(\frac{x}{a} + \frac{y}{b}\right)^2 = 0$$

or
$$y = -\frac{b}{a}x \tag{8.16}$$

The emergent beam is a plane polarised light.

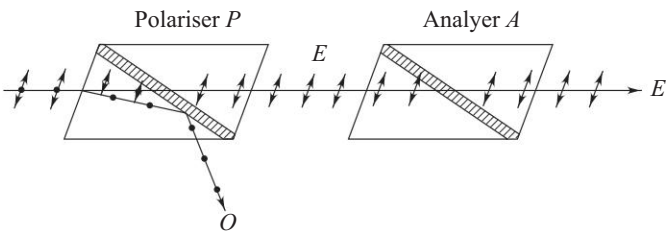
Case IV When $\delta = \pi/2$ and $a = b$
Equation (8.13) becomes

$$(x^2 + y^2) = a^2 \tag{8.17}$$

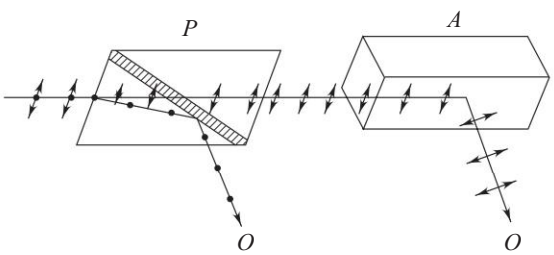
Equation (8.17) is represents the equation of a circle. Therefore, the emergent beam is a circle and is shown in fig 8.11(e).

8.7 PLANE POLARISED LIGHT BY NICOL PRISM

A Nicol prism is used to produce and analyse plane polarised light. Consider an unpolarised light is passed through a Nicol prism, *P*. It is splitted into ordinary and extraordinary rays. The ordinary ray is cut off when it is passing from calcite to Canada balsam layer. The extraordinary ray is passed through it. If it is passed through another Nicol prism, *A*, that is kept parallel to the first one, the extraordinary ray is passed throgh it.



(a) Production of plane polarised light



(b) Analysis of plane polarised light

Fig. 8.12 *Nicol prism*

The McGraw·Hill Companies

If the second Nicol prism is rotated by taking the incident beam as an axis, the emergent beam from the second Nicol prism is extinguished two times and it has two times the maximum intensity within one complete rotation. The first Nicol prism is known as *polariser* and the second Nicol prism is known as *analyser*.

8.8 PRODUCTION OF PLANE, CIRCULARLY AND ELLIPTICALLY POLARISED LIGHT

The different types of polarised lights are produced by employing polarisers and analysers, described below.

8.8.1 Plane Polarised Light

Consider an unpolarised light passed through a tourmaline crystal. The schematic representation of the experimental arrangements is shown in Fig. 8.13. The tourmaline crystal passes the light that is vibrating parallel to the axis, and it absorbs all the remaining lights. The emergent beam from the polariser is found to be slightly coloured and it vibrates only in one direction, and hence, it is said to be a plane polarised light.

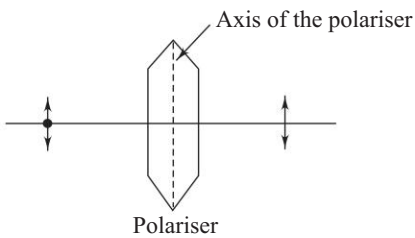


Fig. 8.13 *Plane polarised light–production*

8.8.2 Circularly Polarised Light

The circularly polarised light is produced using a quarter wave plate and a polariser. The plane polarised light produced by a polariser is passed through a quarter wave plate. The quarter wave plate produces two refracted beams. The experimental arrangement used for the production of the circularly polarised light is shown in Fig. 8.14.

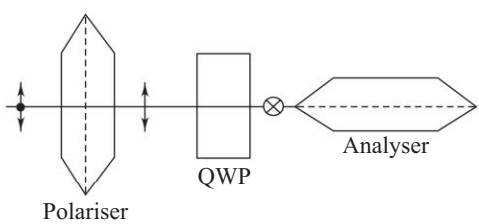


Fig. 8.14 *Circularly polarised light–production*

The QWP introduces a path difference of $\lambda/4$ between these two refracted beams. These refracted beams are passed through an analyser kept perpendicular to the polariser. The emergent beam from the polariser produces a bright field of view.

The McGraw·Hill Companies

By continuously viewing the emergent beam from the analyser, the QWP is rotated by taking the incident beam as an axis. The intensity of the emergent beam from the analyser decreases and it is completely extinguished. When the intensity of the emergent beam from the analyser is zero, the optic axis of the QWP is parallel to the incident beam. From this position, the QWP is rotated through an angle of 45º and it is fixed. When the optic axis of the QWP is making an angle of 45º with the incident beam, the emergent beam from the QWP is found to be circularly polarised.

8.8.3 Elliptically Polarised Light

Elliptically polarised light is produced by passing a plane polarised light through a quarter wave plate. The quarter wave plate splits the incident beam into two and it produces a phase difference of $\pi/2$ between the two refracted beams. The arrangements of the polariser and analyser used for the production of the elliptically polarised light is shown in Fi. 8.15.

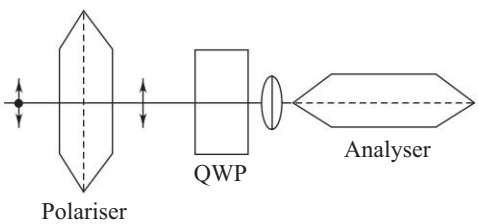


Fig. 8.15 *Elliptically polarised light–production*

If the angle made by the optic axis of the QWP is equal to 45º, the magnitude of these two emergent beams will be same. If the optic axis of the QWP makes an angle other than 45º with the incident beam, the magnitude of these two emergent beams will be unequal. Hence, in order to produce elliptically polarised light, the angle made by the optic axis of the QWP with the incident beam should be made as other than 45º. The emerging elliptically polarised light is analysed using an analyser. If the analyser is rotated by taking the incident beam as an axis, the emergent beam varies from a maximum value to a minimum value. If the major axis of the elliptical light is parallel to the axis of the analyser, the emergent beam has maximum intensity. If the minor axis of the ellipse is parallel to the axis of the analyser, the emergent beam has minimum intensity. The variation of the intensity of the emergent beam ranging from maximum to minimum show that the light emerging from the QWP is elliptically polarised.

8.9 DETECTION OF PLANE, CIRCULARLY AND ELLIPTICALLY POLARISED LIGHT

The existence of the plane, circularly and elliptically polarised light can be analysed employing polariser and analyser. The details of the experimental set up and the method of analysis are given in brief below.

8.9.1 Plane Polarised Light

A plane polarised light emerging from a polariser is analysed by passing it through another tourmaline crystal as shown in Fig. 8.16. By taking the incident beam as

The McGraw·Hill Companies

an axis, if the second tourmaline crystal is rotated, in one complete rotation, the emergent beam from it is completely extinguished two times and consequently the emergent beam has

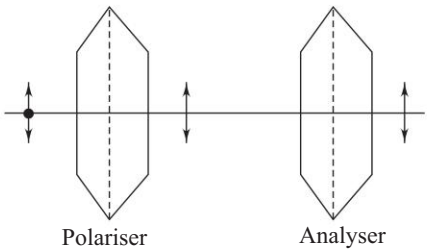


Fig. 8.16 *Plane polarised light–analysis*

a twice the maximum intensity. The second tourmaline crystal is used to analyse the plane polarised light and hence, it is known as an analyser. If the axes of the polariser and the analyser are same, the intensity of the emergent beam is maximum. The intensity of the emergent beam is minimum when the axis of the polariser and the analyser are perpendicular.

8.9.2 Circularly Polarised Light

The given circularly polarised light is passed through the analyser and then, the analyser is rotated by taking the incident beam as an axis, as shown in Fig. 8.17. The emergent beam from the analyser is found to have no variation in intensity because the magnitude of the light vectors in a circularly polarised light is same. In case of an unpolarised light also, if it is passed through an analyser, there is no variation in the intensity of the emergent beam.

To distinguish between the unpolarised light and the circularly polarised light, the beam is first passed through a quarter wave plate. The quarter wave plate splits the light into two and hence, it produces a path difference of $\lambda/4$ between these two rays. If a circularly polarised light is incident on the QWP, the emergent beam will be a plane polarised light. This plane polarised light will be analysed using a polariser. If an unpolarised light is passed through the QWP, the emergent beam will have a phase difference of $\pi/2$. If these emergent beams are passed through an analyser, the beam emerging from the analyser has no variation in the intensity.

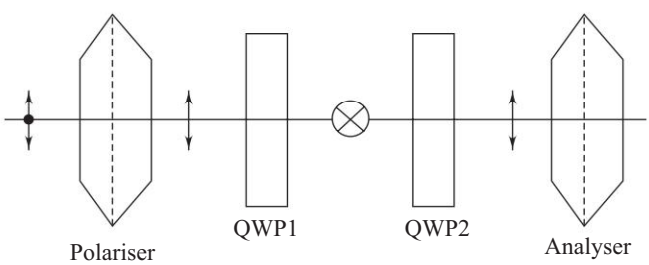


Fig. 8.17 *Circularly polarised light–analysis*

8.9.3 Elliptically Polarised Light

The given elliptically polarised light is first passed through an analyser as shown in Fig. 8.18. The beam emerging from the analyser varies from a maximum to a minimum. In a mixture of polarised and unpolarised light also, the emergent beam varies from a minimum to a maximum.

In order to distinguish between an elliptically polarised light and the mixture of polarized and unpolarised light, the given beam is first passed through a QWP. If the incident beam is an elliptically polarised one, the emergent beam from the QWP is a plane polarised light. Then it is passed through an analyser and it is studied. The emergent beam from the analyser has maximum intensity and minimum intensity two times. If the incident beam is a mixture of unpolarised and polarised light, the emergent beam will have a phase difference of $\pi/2$. If these emergent beams are passed through an analyser, there is no variation in intensity of the beam.

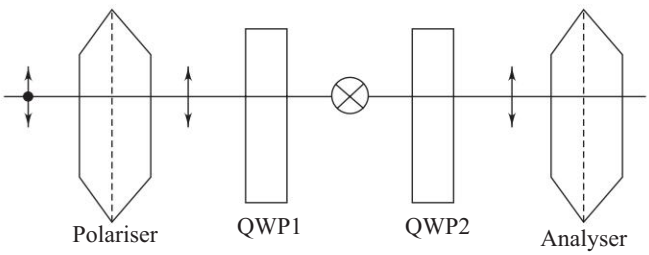


Fig. 8.18 *Elliptically polarised light*

Keypoints to Remember

- Light rays which vibrat in only one direction are known as polarised light.
- There are three different types of polarised lights, namely, plane polarised, circularly polarised and elliptically polarised.
- The plane in which a bean of light is vibrating is known as plane of vibration.
- The plane perpendicular to the plane of vibrations is known as the plane of polarisation.
- A line joining any two blunt corners is known as the optic axis.
- The crystals which have only one optic axis are known as uniaxial crystals.
- The crystals which have two optic axies are known as biaxial crystals.
- Nicol prism is an optical device used to produce and analyse polarised light.
- In a Nicol prism, the sides are blackened to absorb the total internal reflected rays.
- A quarter wave plate is a device from a doubly reflecting uniaxial negative or positive crystal whose faces are cut parallel to the direction of optic axis.
- A quarter wave plate introduces a path difference of l/4 between the two refracted beams.

The McGraw·Hill Companies

- A half wave plate is a device from a doubly refracting uniaxial negative or positive crystal cut in such a way that it can produce a phase difference of p or path difference of l/2 between the ordinary and extraordinary rays.
- Plane polarised light is produced by polarisor, namely, tourmaline crystal.
- Quarter wave plate is used to produce circularly polarised light.
- Elliptically polarised light is produced by passing a plane polarised light through a quarter wave plate.

Solved Problems

Example 8.1 Calculate the thickness of a quarter wave plate. Given that $\mu_e = 1.553$, $\mu_o = 1.544$ and $\lambda = 6 \times 10^{-5}$ cm.

Given Data

The refractive index of the extraordinary ray, $\mu_e = 1.553$
The refractive index of the ordinary ray, $\mu_o = 1.544$
Wavelength, $\lambda = 6 \times 10^{-7}$ m

Solution The thickness of a quarter wave plate,

$$t = \frac{\lambda}{4(\mu_e - \mu_o)}$$

$$= \frac{6 \times 10^{-7}}{2(1.553 - 1.544)}$$

$$= 1.667 \times 10^{-5} \text{ m.}$$

The thickness of the quarter wave plate is 1.667×10^{-5} m.

Example 8.2 Calculate the thickness of a half wave plate of quartz for wavelength 6000 Å. For quartz, $\mu_e = 1.553$ and $\mu_o = 1.544$.

Given Data

Wavelength, $\lambda = 6 \times 10^{-7}$ m

The refractive index of the ordinary ray, $\mu_o = 1.544$

The refractive index of the extraordinary ray, $\mu_e = 1.553$

Solution The thickness of a quarter wave plate,

$$t = \frac{\lambda}{4(\mu_e - \mu_o)}$$

$$\frac{6 \times 10^{-7}}{2(1.553 - 1.544)}$$

$$= 3.33 \times 10^{-5} \text{ m.}$$

The thickness of the half wave plate is 3.33×10^{-5} m.

Example 8.3 Calculate the thickness of a quarter wave plate manufactured from calcite using the following data: $\mu_e = 1.486$, $\mu_o = 1.658$ and $\lambda = 6000$ Å. [AU June 2006]

Given Data

Wavelength, $\lambda = 6 \times 10^{-7}$ m

The refractive index of the ordinary ray, $\mu_o = 1.658$

The refractive index of the extraordinary ray, $\mu_e = 1.486$

Solution The thickness of a quarter wave plate, $t = \dfrac{\lambda}{4(\mu_e - \mu_o)}$

$$= \frac{6 \times 10^{-7}}{4(1.658 - 1.486)}$$

$$= 8.7 \times 10^{-7} \text{ m.}$$

The thickness of the half wave plate is 8.7×10^{-7} m.

Example 8.4 Calculate the wavelength of light used in a half wave plate of thickness 16.4×10^{-7} m. The HWP is made up of calcite (Given: $\mu_e = 1.486$ and $\mu_o = 1.658$).

Given Data

Thickness, $t = 1.64 \times 10^{-6}$ m

The refractive index of the ordinary ray, $\mu_o = 1.658$

The refractive index of the extraordinary ray, $\mu_e = 1.486$

Solution The wavelength of light, $\lambda = 2t(\mu_o - \mu_e)$

$$= 2 \times 1.64 \times 10^{-6} \times (1.658 - 1.486)$$

$$= 0.564\ \mu\text{m.}$$

The wavelength of light is 0.564×10^{-6} m.

Example 8.5 A quarter wave plate is 12.5×10^{-6} m thick. Calculate the wavelength for which it acts as a quarter wave plate. The difference in the principle refractive indices is 0.01

Given data

Thickness $t = 12.5 \times 10^{-6}$m

The difference in the principal Refractive indius $\mu_0 - \mu_e = 0.01$

Solution:

The wavelength of the light $\lambda = 2t\,(\mu_o - \mu_e)$

Substituting the value, we get,

$$= 2 \times 12.5 \times 10^{-6} \times 0.01$$
$$= 0.25 \times 10^{-6}$$

The wavelength of light is 0.25×10^{-6} m.

Objectives Questions

8.1 _____ is used to produce polarised light.

8.2 When the crystals are perpendicular to each other, the intensity of the emergent beam from the second crystal is _____.

8.3 The ordinary light has vibration of both _____ and ____ to the plane of the paper.

8.4 The chemical name of calcite crystal is _____.

8.5 Examples for uniaxial crystals are _____, _____, and _____.

8.6 Crystals like borax, mica and selenite possess two optic axes and hence, are known as _____ crystals.

8.7 The emergent beam from a Nicol prism is _____.

8.8 The structure of calcite is _____.

8.9 In a Nicol prism, the total internal reflection is obtained due to the presence of

a) optic axis b) Canada balsam layer

c) ordinary rays d) extra ordinary rays

8.10 The thickness of the HWP (calcite) required to produce a path difference of $\lambda/2$ is

a) $t = \dfrac{\lambda}{2(\mu_0 - \mu_e)}$ b) $t = \dfrac{\lambda}{2(\mu_e - \mu_o)}$

c) $t = \dfrac{\lambda^2}{2(\mu_0 - \mu_e)}$ d) $t = \dfrac{\lambda}{2(\mu_0 + \mu_e)}$

8.11 The thickness of negative HWP, required to produce a path difference of $\lambda/2$ is

a) $t = \dfrac{\lambda}{2(\mu_e + \mu_o)}$ b) $t = \dfrac{\lambda^2}{2(\mu_0 - \mu_e)}$

c) $t = \dfrac{\lambda}{2(\mu_0 - \mu_e)}$ d) $t = \dfrac{\lambda}{2(\mu_e - \mu_o)}$

8.12 Equation for a symmetrical ellipse is

a) $\frac{x^2}{a^2}+\frac{y^2}{b^2}=0$ b) $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$

c) $\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$ d) $\frac{x^2}{a^2}-\frac{y^2}{b^2}=0$

8.13 The equation for the emergent beam of a circle is

a) $(x^2-y^2)=a^2$ b) $(x^2+y^2)=2a$

c) $(x^2+y^2)=a^2$ d) $(x^2+y^2)=a$

8.14 _____ is used for the production and analysis of plane polarised light.
8.15 _____ crystal is used to convert the unpolarised light into plane polarised light.
8.16 The circularly polarised light is produced by _____ plate.
8.17 _____ wavelength is designed for a particular wavelength and hence, it is not useful for other wavelengths.

KEY:

8.1 Polariser
8.2 minimum
8.3 parallel, perpendicular
8.4 hydrated calcium carbonate
8.5 calcite, tourmaline, quarty
8.6 bioaxial
8.7 plane polarised
8.8 rhombohedrun
8.9 b)
8.10 a)
8.11 d)
8.12 b)
8.13 c)
8.14 nicol pism
8.15 Tourmaline
8.16 QWP
8.17 QWP

Exercises

8.1 Calculate the thickness of (i) a quarter wave plate, and (ii) a half wave plate, given that μ_e = 1.553, μ_o = 1.544 and λ = 5000 Å.
8.2 Find the thickness of a quarter wave plate for a wavelength of 5893 Å. Given that μ_e =1.553 and μ_o = 1.544.
8.3 A half wave plate of 5 μm thickness is used to produce circularly polarised light. If μ_e = 1.553 and μ_o = 1.544, What is the wavelength of the light used?
8.4 Find the thickness of a half wave plate made up of quartz for λ = 500 nm. For quartz, μ_e = 1.553 and μ_o= 1.544.

The McGraw·Hill Companies

Short Questions

8.1 What is meant by polarisation of light?
8.2 What is plane polarised light?
8.3 What is circularly polarised light?
8.4 What is elliptically polarised light?
8.5 Define the term plane of polarisation.
8.6 Define the term plane of vibration.
8.7 What is an optic axis?
8.8 What are uniaxial and biaxial crystals?
8.9 What is meant by principal plane?
8.10 What is a principal section?
8.11 What are the limitations of a Nicol prism?
8.12 Explain the production of plane polarised light by Nicol prism.
8.13 What is a quarter wave plate?
8.14 Mention the uses and drawbacks of a quarter wave plate.
8.15 What is a half wave plate?
8.16 Mention the uses and drawbacks of a half wave plate.
8.17 Explain how circularly polarised light is produced.
8.18 Explain how eliptically polarised light is produced.
8.19 What do you understand by quarter and half wave plates

Descriptive Questions

8.1 (a) Explain the construction and working of a Nicol prism.
(b) Enumerate the actions of a Nicol prism.
(c) What are the limitations of a Nicol prism?
8.2 Explain with suitable mathematical derivation, the formation of plane, circularly and elliptically polarised lights.
8.3 Explain in detail the production and analysis of plane, circularly and elliptically polarised lights.
8.4 Distinguish between the (1) Circularly Polarised light and (2) elliptically polarised and partially polarised light

Chapter

9

Photoelasticity

OBJECTIVES

- To acquire knowledge on the distribution of stress / strain in a material under loading conditions
- To understand the theory of photoelasticity
- To study the use of polariser and analyser in photoelastic analysis
- To characterise the distribution of stress / strain on a material under applied load
- To study the possible applications of photoelasticity in industry

9.1 INTRODUCTION

In order to determine the suitability of the materials used in bridges, roller bearings, wheels of a railway wagon, etc., it is essential to study their mechanical properties under different conditions. The experimental stress analysis aims to find the state of stress at various points of the material for the known loading conditions. The experimental methods used for the stress analysis may be classified as *whole field methods* and *point-by-point methods*.

Whole field methods give information about the distribution of stress or strain in the whole area of the material, e.g., photoelasticity, grid method, holography, etc. However, point-by-point methods give information only at some selected points, e.g., electrical strain gauges, optical extensometers, etc. Further, these methods may be classified as static or dynamic, and destructive or non-destructive methods. A non-destructive method used for the quality testing has been discussed earlier. In this chapter, we will discuss the photoelastic method.

Photoelastic Method The photoelastic method was first introduced by David Brewster in 1814, and later it was developed by Maxwell, Neuman and many others. Photoelasticity is one such technique specially useful at the designing stage.

In this method, a model of the material is prepared using a suitable transparent component (plastic or glass) and is placed in the path of a beam of polarised light. On application of stress on the model, an interference pattern is formed on the screen. This pattern yields immediate qualitative results, i.e., it indicates the area of high stress, low stress and the resultant peak stress at the boundary and the interior parts

The McGraw·Hill Companies

of the model. The polarised light used for this purpose can be either plane polarised or circularly polarised.

Let us first discuss the optical terms used in this section.

9.2 DEFINITIONS

9.2.1 Polarised Light

According to the wave theory, light waves have transverse wave nature. If the light waves vibrate in only one direction, it is known as polarised light.
There are three different forms of polarised light.

(i) Plane polarised light
(ii) Circularly polarised light, and
(iii) Elliptically polarised light

(i) Plane Polarised Light In a plane polarised light, the light vector *S* vibrates perpendicular to the direction of propagation in a single plane as shown in Fig. 9.1.

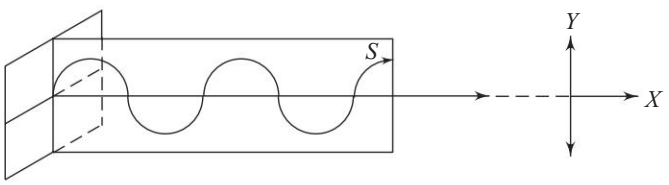


Fig. 9.1 *Plane polarised light*

(ii) Circularly Polarised Light The light vector *S* rotates round the line of propagation and describe a circular helical path, which is either right-or-left handed as shown in Fig. 9.2.

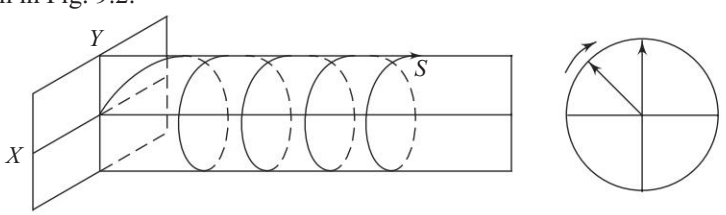


Fig. 9.2 *Circularly polarised light*

(iii) Elliptically Polarised Light The tip of the light vector *S* travels forward in an elliptical helical path and at the same time rotates around the axis of propagation as shown in Fig. 9.3.

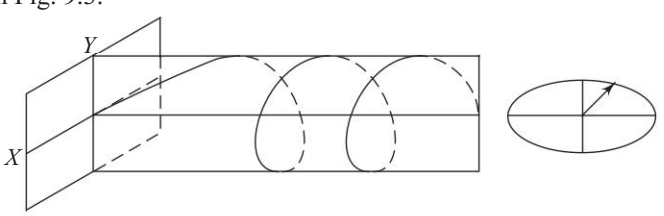


Fig. 9.3 *Elliptically polarised light*

The McGraw·Hill Companies

9.2.2 Double Refraction

When a beam of unpolarised light is incident on a crystalline material such as mica, calcite, etc., the beam upon entering the crystalline material gets split into two components of plane polarised light. This phenomenon is known as double refraction and the material exhibiting this behaviour is known as doubly refracting material. The two resultant light vectors travel with different velocities and emerge at the other end with a phase difference as shown in Fig. 9.4.

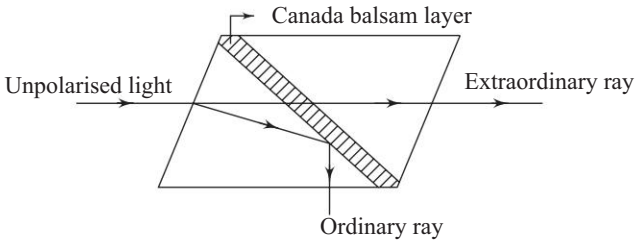


Fig. 9.4 *Double refraction in Nicol prism*

Certain crystalline materials undergo double refraction only when they are subjected to strain. The double refraction vanishes when the load is removed. This phenomenon is known as *temporary or artificial double refraction*.

9.2.3 Wave Plate (or Retardation Plate)

Certain crystalline materials have the ability to resolve the light vectors into two orthogonal components and transmit each one of them at different speed and phase. This phase difference is proportional to the thickness of the plate. Such plates are known as wave plates or retardation plates. A typical wave plate is shown in Fig. 9.5.

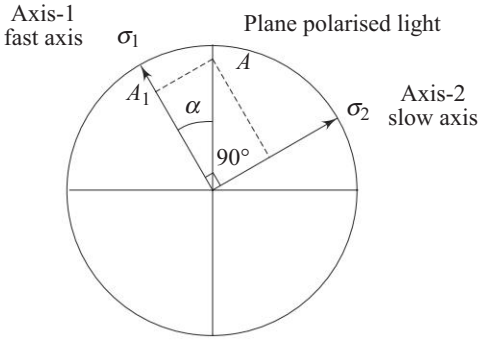


Fig. 9.5 *Wave plate*

The wave plate is of two types, namely, Quarter Wave Plate (QWP) and Half Wave Plate (HWP), depending on the path difference ($\lambda/4$ and $\lambda/2$ respectively) produced between the two orthogonal components.

The quarter wave plate will convert a plane polarised light into a circularly polarised light, when the plane of vibration of the incident beam makes an angle of 45° with the

direction of the optic axis of the QWP. When one of the components is suppressed, the resultant light is plane polarised.

9.3 THEORY OF PHOTOELASTICITY

According to Brewster (1814), certain transparent materials like bakelite, synthetic glass, ebonite, etc., exhibit temporary double refraction when subjected to stress on its own plane. The amount of double refraction at a point is proportional to the maximum shear at that point. The photoelasticity is an optical method used to determine the degree of double refraction existing in a strained body. It provides a clear picture of the stress distribution throughout the model irrespective of its shape and elastic constant of the material.

Stress–optic Law The stress–optic law gives the relationship between the changes in the refractive indices and the principal stresses of a strained photoelastic materials.

Let σ_1 and σ_2 be the principal stresses at a point in a two-dimensional photoelastic material. Let n_1 and n_2 be the refractive indices along the two principal axes associated with σ_1 and σ_2. Then according to Maxwell, the relationship between the principal stresses and refractive indices can be written as,

$$n_1 - n_0 = c_1\sigma_1 + c_2\sigma_2 \tag{9.1}$$

$$n_2 - n_0 = c_1\sigma_2 + c_2\sigma_1 \tag{9.2}$$

where n_0 is the refractive index of the material in the unstressed state, c_1 and c_2 are the direct and transverse stress optic coefficients. Subtracting Eq. (9.1) from Eq. (9.2) to eliminate n_0, we get,

$$n_2 - n_1 = (c_2 - c_1)(\sigma_1 - \sigma_2)$$

or
$$n_2 - n_1 = c(\sigma_1 - \sigma_2) \tag{9.3}$$

where $c = (c_2 - c_1)$ is a constant known as stress optic coefficient.

This equation is known as *stress–optic law*. Therefore, the values of $(\sigma_1 - \sigma_2)$ give the relative phase difference (retardation) of the two components of light through the model.

Effect of Stressed Model in a Plane Polariscope Consider a doubly refracting plate (QWP) of thickness h placed in a field of plane polarised light $S(a \sin \omega t)$. The light vectors make an angle α with the fast axis and get resolved along the direction of the principal planes, namely, σ_1 and σ_2 as shown in Fig. 9.6.

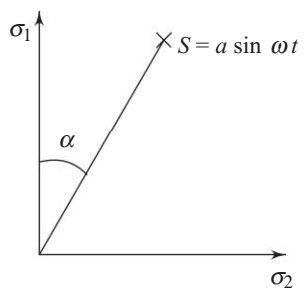


Fig. 9.6 *Relation between the principal stresses σ_1 and σ_2*

The McGraw·Hill Companies

If we resolve the light vector into two components along fast and slow axes, it can be written as (see Fig. 9.7)

$$S_1 = a \cos \alpha \sin \omega t \text{ parallel to } \sigma_1$$

$$S_2 = a \sin \alpha \sin \omega t \text{ parallel to } \sigma_2$$

Let t_1 and t_2 be the time required for the transmission on two principal planes σ_1 and σ_2. Then, the two components leaving the model can be represented as,

$$S'_1 = a \cos \alpha \sin \omega(t - t_1) \text{ and } S'_2 = a \sin \alpha \sin \omega(t - t_2)$$

and their angular phase difference is $\omega(t_2 - t_1)$.

When the two vibrations emerging out from the model are allowed to pass through another polariser, whose plane of polarisation is at right angles to the plane of vibration of the incident light on the model (the second polariser is known as analyser, see Fig. 9.7), then the resultant light is represented by

$$S = S_1'' - S_2'' \tag{9.4}$$

where $S_1'' = a \cos \alpha \sin \alpha \sin \omega(t - t_1)$ and

$S_2'' = a \sin \alpha \cos \alpha \sin \omega(t - t_2)$

Substituting the S_1'' and S_2'' values in Eq. (9.4), we get,

$$S = a \cos \alpha \sin \alpha \sin \omega(t - t_1) - a \sin \alpha \cos \alpha \sin \omega(t - t_2) \tag{9.5}$$

Using the standard formulas $\sin C - \sin D = 2 \sin\left(\frac{C-D}{2}\right)\cos\left(\frac{C+D}{2}\right)$, and $\sin 2\alpha = 2 \sin \alpha \cos \alpha$ the above equation can be written as,

$$S = a \sin 2\alpha \sin\left[\omega \frac{t_2 - t_1}{2}\right] \cos \omega\left[t - \frac{t_1 + t_2}{2}\right] \tag{9.6}$$

The intensity of the emergent beam from the analyser is proportional to the square of the amplitude of light

i.e., $$I \propto a^2 \sin^2 2\alpha \sin^2\left[\omega \frac{t_2 - t_1}{2}\right] \cos^2 \omega\left[t - \frac{t_1 + t_2}{2}\right] \tag{9.7}$$

The amplitude of the resultant vibration is a function of α and phase difference $\omega(t_2 - t_1)$. The intensity of the transmitted light at any point depends on the direction of the principal stress relative to the polarising plane and also on the difference of principal stress at that point.

The intensity of the transmitted light is zero at every point.

i.e. $\sin^2 2\alpha = 0$ or $\sin^2 \omega(t_2 - t_1)/2 = 0$ (9.8)

Case I For $\alpha = 0$ or $\pi/2$, we get the loci or focus of the point for which extinction of the light takes place. Under this condition, the fringes formed are known as *isoclinic*.

Case II For $\omega(t_2 - t_1)/2 = n\pi$

where $n = 0, 1, 2, 3$, etc., we get a series of coloured bands is known as *isochromatic fringes*.

The McGraw·Hill Companies

The determination of the isoclinic fringes gives the direction of the principal stress and isochromatic fringes give maximum shear stress values.

The stress optic law can be written from Eq. (9.3.)

$$R_t = n\lambda = \frac{c(\sigma_1 - \sigma_2)}{d}$$

or
$$\sigma_1 - \sigma_2 = d\,\frac{\lambda n}{c} \qquad (9.9)$$
$$= F_m n$$

where F_m is a constant for a model called 'fringe constant' and n the 'fringe order'.

Let us discuss the experimental methods used to determine the strength of materials.

9.4 ARRANGEMENT OF OPTICAL ELEMENTS IN A POLARISCOPE

The commonly used polariscopes for photoelastic analysis are

(i) Plane polariscope
(ii) Circular polariscope

9.4.1 Plane Polariscope

This is the simplest optical system consisting of a light source, polariser and an analyser arranged as shown in Fig. 9.7.

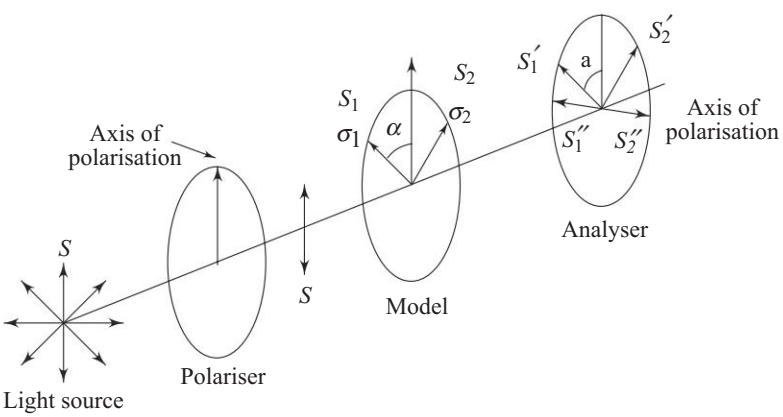


Fig. 9.7 *Plane polariscope*

The polariser N_1 and the analyser N_2 are both made of Nicol prism. In the plane polariscope set-up, the axes of the polariser and analyser are at right angles to each other. Therefore, no light is transmitted through the analyser. This results in a dark field on the screen. The photoelastic model is inserted between the two crossed elements and the resultant pattern is viewed through the analyser for different loading conditions.

9.4.2 Circular Polariscope

In this optical system, two additional QWPs are inserted, one between the polariser and the model and the other between the model and the analyser. The arrangement of optical elements in a circular polariscope is shown in Fig. 9.8.

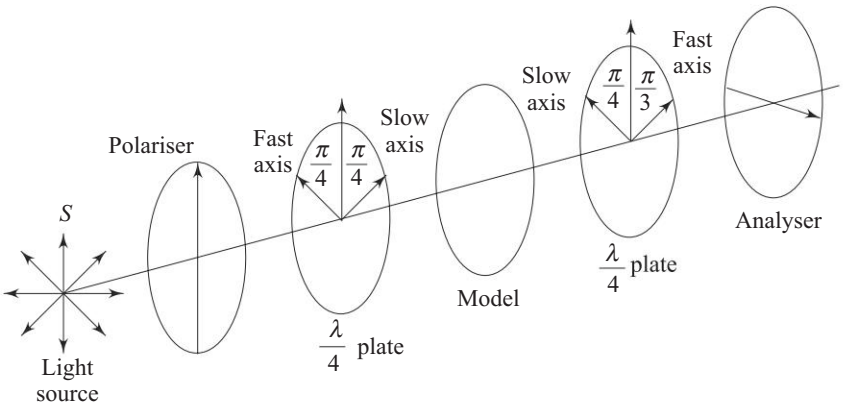


Fig. 9.8 *Circular polariscope*

In this arrangement, the axis of the quarter wave plates (QWP1 and QWP2) are arranged at an angle of $\pi/4$ to the plane of polarisation. The first quarter wave plate (QWP1) converts the plane polarised light into circularly polarised light. Similarly, the second quarter wave plate (QWP2) converts circularly polarised light into plane polarised light, whose fast axis is parallel to the slow axis of the first quarter wave plate. Thus, the resultant light again vibrates in the vertical plane. This arrangement of optical elements constitute a circular polariscope and it produces a dark field. The resultant field in the polariscope depending on the position of QWPs are described in the following Table 9.1.

Table 9.1 *Arrangements of optical elements in a circular polariscope*

Arrangements	*Quarter wave plate*	*Polariser and analyser*	*Field*
1	Crossed	Crossed	dark
2	Crossed	Parallel	light
3	Parallel	Crossed	light
4	Parallel	Parallel	dark

9.5 INTERPRETATION OF ISOCLINIC AND ISOCHROMATIC FRINGES

9.5.1 Isoclinic Fringes

The model is inserted between the polariser and analyser in a plane polariscope. When the model is viewed through the analyser for a known loading condition, the resultant pattern appears as dark areas superimposed on a bright coloured pattern, as shown in Fig. 9.9a.

In these fringes, the intensity of the transmitted light will be zero.

i.e. $\sin^2 2\alpha = 0$ (9.10)

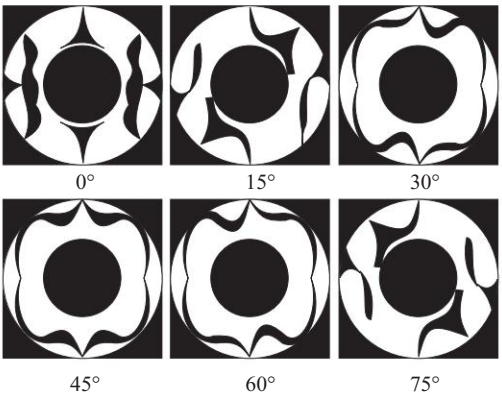


Fig. 9.9(a) *Isoclinic fringes*

The dark areas are called isoclinic fringes. This represents the locus of all points where the principal stress is parallel to the plane of polarisation. If one measures the angle from the vertical reference axis, it is known as 0°-isoclinic. If the polariser and analyser are turned through 30° by keeping the optical axis always crossed, the dark area is shifted to another position, i.e., 0°- isoclinic vanishes and a new dark band is observed on the screen. This is known as 30°-isoclinic. Thus, the position of the new locus represents the orientation of principal stress. By rotating the polariser and analyser for every 10° or 15°, the isoclinic for the entire model can be represented by tracing the loci.

9.5.2 Isochromatic Fringes

In a plane polariscope, the loaded model forms a coloured pattern as shown in Fig. 9.9 b, which gives information about the magnitude of the difference between the principal stresses, i.e., the maximum shearing stress.

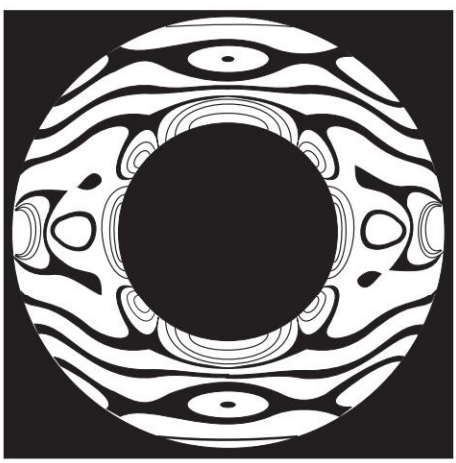


Fig. 9.9(b) *Isochromatic fringes*

The McGraw·Hill Companies

We know that σ_1 and σ_2 are the principal stresses in an element at any particular point, then the maximum stress at any point can be written as,

$$\tau_{max} = \frac{\pm(\sigma_1 - \sigma_2)}{2} \tag{9.11}$$

Each coloured point, is the locus of the points having a constant difference between the principal stresses.
From Eq. (9.8), at the points of extinction

$$\sin^2\left[\frac{\omega(t_1 + t_2)}{2}\right] = 0 \tag{9.12}$$

Equation (9.12) is satisfied only when the phase difference is an integral multiple of π.

i.e., $$\frac{\omega(t_2 - t_1)}{2\pi} = n = \frac{c(\sigma_1 - \sigma_2)}{dy}$$

$\therefore$ $$\sigma_1 - \sigma_2 = \frac{\lambda n \times d}{c} \tag{9.13}$$

where $n = 0, 1, 2, 3$, etc. This type of fringes are known as *isochromatic fringes*.

9.6 PHOTOELASTIC BENCH

Basically, photoelastic bench is a circular polariscope which consists of a light source, polariser, two QWPs, model, analyser, compensator and a camera. The schematic diagram of the arrangement of the optical elements is shown in Fig. 9.10.

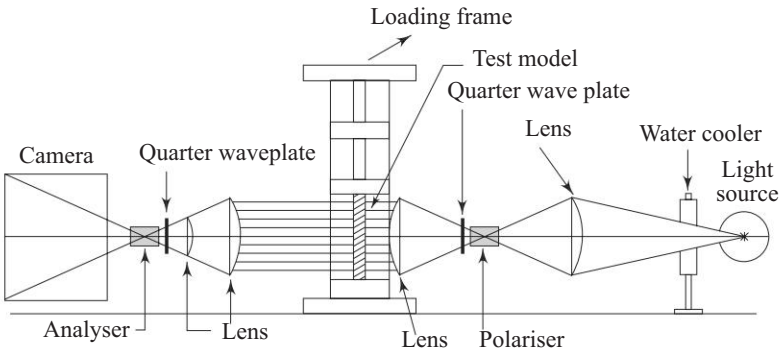


Fig. 9.10 *Photoelastic bench*

Working A white beam of light from the lamp is focused on to the polariser. The polariser converts unpolarised light into a plane polarised light. The plane polarised light emerges as a circularly polarised light from the QWP1. The lenses are used to obtain an image of desired size of the model. The loading frame can impart tension, compression and bending to the model. The QWP2 converts the circularly polarised light into plane polarised light. The interference pattern is viewed through the analyser and using ordinary camera, the pattern is photographed simultaneously. In order to

get a good focusing on every part of the image, there are provisions for adjusting the focal planes of various lenses.

Keypoints to Remember

- Whole field and point-by-point are the experimental methods used to determine the stress analysis in a material.
- The whole field method gives information about the distribution of stress or strain in the whole area of the material.
- The point-by-point method gives information only at the selected points.
- In polarised light, the waves vibrates only in one direction.
- When a beam of unpolarised light is incident on a crystalline material such as mica and calcite, the resultant beam is split into two components of plane polarised light. This phenomena is known as double refraction.
- When the load is removed from materials, double refraction vanishes and this phenomena is known as temporary or artificial double refraction.
- Certain crystalline materials have the ability to resolve the light vectors into two components and transit each one of them at different speeds and phases. These materials which are in the form of a plate are known as retardation or wave plates.
- Stress optic law gives the relationship between the changes in the refractive indices and the principle stress of strained photoelastic materials.
- The stress optic law is $n_2 - n_1 = c\,(\sigma_1 - \sigma_2)$ where c is the stress optic coefficient, n_1 and n_2 are refractive indices along the two principle axis associated with the stress σ_1 and σ_2.
- When $\alpha = 0$ or $\pi/2$, it froms the focus or loci of the point for which extinction of light takes place. The fringes formed under this condition is known as isoclinic fringes.
- Coloured bands at $n = 0,1,2$ etc., are known as isochromatic fringes.
- Isoclinic fringes are represented by dark areas. It is the locus of all points where the principle stress is parallel to the plane of polarisation.

Solved Problems

Example 9.1 In an experiment using a photoelastic bench, the difference between the principal stresses is equal to 10×10^8 N m^{-2}. The photoelastic material has its relative stress optic coefficient as 1 brewsters (1 brewsters = 10^{-12} m^2 N^{-1}). Calculate the difference between the refractive indices along that principal stresses.

The McGraw·Hill Companies

Given Data

The difference between the principal stress, $\sigma_1 - \sigma_2 = 10 \times 10^8$ N m^{-2}
The stress–optic coefficient, $c = 1 \times 10^{-12}$ m^2 N^{-1}

Solution The difference between the refractive indices

$$N = n_2 - n_1$$

Substituting the c, σ_1 and σ_2 values in the above equation, we get,

$$= c \times (\sigma_1 - \sigma_2) = 1 \times 10^{-12} \times 10 \times 10^8$$
$$= 1 \times 10^{-3}$$
$$= 0.001$$

The difference between the refractive indices along the principal stresses = 0.001.

Example 9.2 The magnitude of stress acting at a point in a stressed material is given by $\sigma_1 = 405 \times 10^6$ N m^{-2} and $\sigma_2 = -105 \times 10^6$ N m^{-2}. Determine the values of the maximum shearing stress.

Given Data

The principal stress, $\sigma_1 = 405 \times 10^6$ N m^{-2}
The principal stress, $\sigma_2 = -105 \times 10^6$ N m^{-2}

Solution The maximum value of the shearing stress

$$\tau_{max} = \frac{\sigma_1 - \sigma_2}{2} = \frac{405 \times 10^6 - \left(-105 \times 10^6\right)}{2}$$
$$= 255 \times 10^6 \text{ N m}^{-2}$$

The maximum shearing stress $= 255 \times 10^6$ N m^{-2}.

Objectives Questions

9.1 _____ method gives information about the distribution of stress or strain in the whole area of the material.
9.2 The point-by-point method gives information only at selected _____.
9.3 _____ is used at the initial designing stage.
9.4 In photoelasticity, the model of the material is prepared using _____ transparent component.
9.5 The following is not a form of polarised light.
a) Plane polarised light b) Circularly polarised light
c) Incident light d) Elliptically polarised light
9.6 The phenomena of removal of double refraction upon removal of applied load is known as _____.

The McGraw·Hill Companies

9.7 Acronym for QWP and HWP is _____.
9.8 The path difference produced by QWP is _____.
9.9 _____ is the path difference produced by HWP.
9.10 The stress optic law is equal to
a) $n_2 - n_1 = c\,(\sigma_1 - \sigma_2)$ b) $n_2 - n_1 = 2c\,(\sigma_1 - \sigma_2)$
c) $n_2 - n_1 = c_2\,(\sigma_1 - \sigma_2)$ d) $n_1 - n_2 = c\,(\sigma_1 - \sigma_2)$
9.11 The maximum stress at any point in an element during the loading is

a) $\frac{\pm(\sigma_1 - \sigma_2)}{4}$ b) $\frac{\pm(\sigma_1 - \sigma_2)}{2}$

c) $\frac{\pm(\sigma_1 + \sigma_2)}{2}$ d) $\frac{\pm(\sigma_1 + \sigma_2)}{4}$

9.12 Photoelastic bench is a _____ polariscope.

KEY

9.1 Whole field 9.2 points
9.3 Photo elasticity 9.4 glass or plastic
9.5 c
9.6 temporary double refraction or artificial double refraction
9.7 Quarter Wave Plate, Half Wave Plate
9.8 $\lambda/4$ 9.9 $\lambda/2$
9.10 a) 9.11 b)
9.12 circular

Exercises

9.1 The principal stress acting at a point is given by $\sigma_1 = 380 \times 10^6$ N m^{-2} and $\sigma_2 = -240 \times 10^6$ N m^{-2}. Determine the maximum shearing stress acting at that point.
9.2 The difference between the two principal stresses acting at a point is 5×10^8 N m^{-2}. If the stress–optic coefficient is 10 brewster, calculate the difference between the refractive indices.

Short Questions

9.1 What is photoelasticity?
9.3 What is meant by photoelastic effect?
9.4 Define plane polarised light.

The McGraw·Hill Companies

9.5 Define circularly polarised light.
9.6 What is meant by elliptically polarised light?
9.7 Define the term, "temporary double refraction".
9.8 What are the common materials employed in the preparation of photoelastic test models ?
9.9 What are isoclinic and ischromatic fringes?
9.10 What are isochromatic fringes?
9.11 Give the applications of photoelasticity.

Descriptive Questions

9.1 What are isochromatic and isoclinic fringes ? How are they analysed by the use of photoelastic bench?
9.2 Explain how stresses in a material are distributed by using stress–optic law.
9.3 Write a note on the arrangements of optical elements in plane and circular polariscope.
9.4 Discuss the stressed model in a plane polariscope and obtain the necessary conditions for isochinic fringe pattern.
9.5 Explain photo elasticity. Explain its use and discuss stress optic law

Chapter

10

Optical and Other Instruments

OBJECTIVES

- To study the principle, mechanism and applications of optical instruments
- To explore the method of finding the altitude of the sun and the height of a tower/ building
- To understand the operating principle and applications of optical microscope, TEM, SEM and STEM
- To understand the role of optical instruments in material characterisation

10.1 INTRODUCTION

Some instruments like photographic camera, telescope, microscope, eyepiece, lenses, and sextants, are based on the principle of the properties of light like reflection, refraction, etc. They are known as optical instruments. Some other instruments, like an electron microscope, use an electron beam, which obeys the optical principles, like refraction, etc. The branch of science, which deals about the optical properties of the electrons, is known as electron optics. In this chapter, the instruments based on the principle of optics and electron optics are discussed.

10.2 SEXTANT

The diagrammatic representation of a sextant is shown in Fig. 10.1. It consists of a circular arc SS' of about 60°. The circular arc (circular scale) SS' is attached with two fixed radial arms A and B. At the centre, there is a movable arc C,

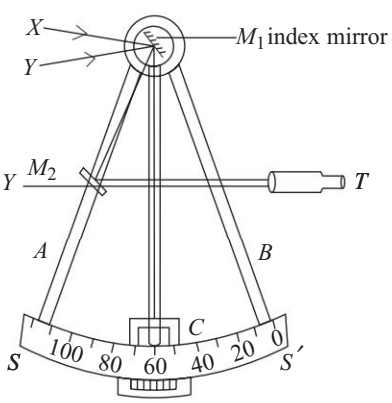


Fig. 10.1 *Sextant*

The McGraw·Hill Companies

known as index arm. A mirror M_1 is placed at one end of the index arm. This mirror is called as index mirror. In another end, a vernier scale V is attached with the index arm. A semi-transparent silvered glass M_2, known as horizon glass, is fixed in the arm A. A telescope is attached with the arm B.

The sextant is based on the principle that if the index mirror M_1 is rotated through an angle of θ, then the reflected ray rotates through an angle of 2θ. Since the reflected ray rotates through an angle of 2θ, the scale SS' is graduated into 120° by taking 0.5° = 1°.

To measure the angle between the points Y and X, the sextant is initially fixed so the telescope receives the image of the point Y directly through the mirror M_2. Then the index mirror is rotated in such a way so as to get the image of the point Y in the field of view of the telescope. The light ray coming from Y reaches the telescope after getting reflected by the index glass M_1 followed by another reflection by the mirror M_2. The two images obtained are made to lie side by side and then the vernier scale is coincided to zero of the main scale (SS'). If it is not possible to coincide the vernier scale to zero of the main scale, note the vernier scale reading and it should be added or subtracted depending on the errors. Then the index mirror M_1 is adjusted so as to get the reflected image of the point X in the field of view of the telescope. The image of X reflected by the index glass M_1 and the image of Y through the horizon glass should made to lie side by side and the position of scale is noted. From the difference between the two readings, the angle between the points X and Y is determined. A sextant is also used to measure the height of an inaccessible tower and the altitude of the sun.

10.2.1 Measurement of the Height of an Inaccessible Tower

Let A and B be the bottom and top portions of a tower. Consider our aim is to find the distance between the points A and B using a sextant. To measure the distance between the points A and B of the tower, a base line is drawn from the tower. The direct image of A is seen through the telescope. The mirror M_1 is adjusted so as to get the image of A in the telescope side by side [Fig. 10.2a] and the reading in the scale SS' is noted. Then the mirror M_1 is rotated so as to receive the image

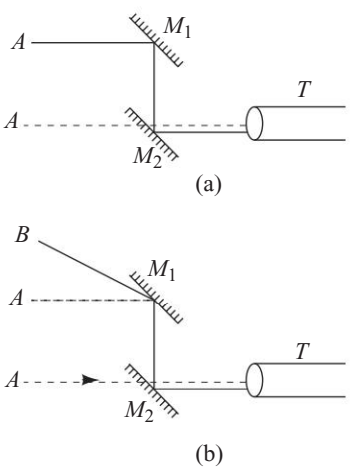


Fig. 10.2 *Sextant—height of a tower*

of the top of the building or tower. The image of the top of the tower or building *B* and the image of the bottom level *A* are seen side by side [Fig.10.2b] in the telescope. The reading in the scale is noted and the angle θ_1 through which the index arm rotated is found out.

Now move the sextant through a distance of *d* meter (say 10 m) and repeat the entire experiment again. The angle θ_2 through which the index arm is rotated is determined. Let the sextant be first positioned at *C* and then at *D*.

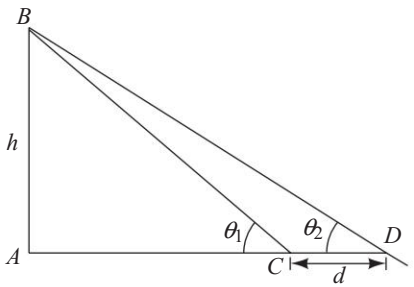


Fig. 10.3 *Measurement of the height of an inaccessible tower*

From Fig. 10.3, $\cot \theta_1 = \frac{AC}{h}$ and $\cot \theta_2 = \frac{AD}{h}$

By measuring the distance between *C* and *D*, we get,

$$\cot \theta_2 - \cot \theta_1 = \frac{AD - AC}{h} = \frac{d}{h}$$

$$\therefore \quad h = \frac{d}{\cot \theta_2 - \cot \theta_1}$$

The height of the building or the tower is calculated using the above equation.

10.2.2 Measurement of the Altitude of the Sun

To measure the altitude of the sun, since it is not possible to find the horizon of the sun, an artificial horizon is used. A dish containing mercury is used as an artificial horizon. The telescope of the sextant is turned towards the mercury surface. One image *A′* of the sun from the mercury is seen through a transparent mirror M_2 and another image due to the reflection from the index glass M_1 and the horizon glass M_2 are obtained in the field of view of the telescope (Fig. 10.4). These two images should be made to lie side by side and the reading in the scale is noted. From these readings θ_1 is determined. The index glass is now rotated to receive the image of the sun directly and mirror M_2 reflects it. This image should lie side by side with the image of the sun in mercury as seen through the transparent mirror M_2. The position of the scale is also noted. The value of θ_2 is determined from these readings. The angular elevation of the sun, a is determined using the relation, $\alpha = (\theta_2 - \theta_1)/2$.

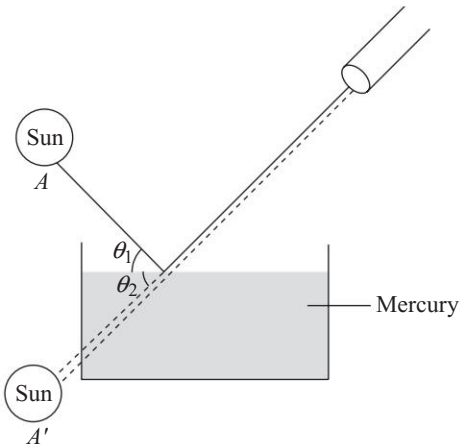


Fig. 10.4 *Measurement of the altitude of the sun*

10.3 METALLURGICAL MICROSCOPE

In a biological microscope, the transmitted light is used to view the specimen. Since the metals are opaque to light, the metallurgical microscope uses reflected light to view the specimen.

The diagrammatic representation of a metallurgical microscope is shown in Fig. 10.5. In a metallurgical microscope, a beam of light produced by an incandescent lamp (halogen or xenon lamp) is made to fall on a glass plate inclined at an angle of 45°. The light reflected by the inclined glass plate is used to illuminate the object. An objective lens receives the light reflected from the object and it focuses the image at its focal point. An eye-piece receives the image and enlarges it. The enlarged image is seen through the eye-piece lens.

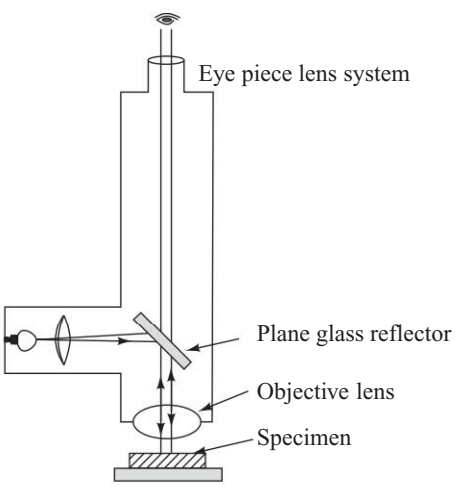


Fig. 10.5 *Metallurgical microscope*

The McGraw·Hill Companies

Usually, the objective lens with different magnifications like 5x, 10x, 20x, 50x and 100x are provided in a turret mount. The eye-piece lenses with different magnifications like 5x, 10x, 20x, etc., are also provided. The total magnification of the microscope depends upon the product of the magnifications of the eyepiece lens and that of the objective lens. Thus, a magnification from 20x to 2000x is easily achieved using a metallurgical microscope.

The overall magnification is approximately given by DN/F, where D is the length measured from back of objective to lower end of eye-piece, N the power of eye-piece and F the focal length of the objective lens.

The resolution of image is given by $\chi = 0.5\lambda/(\mu \sin a)$ where λ is the wavelength of light used, μ the refractive index between lens and specimen and a is half the angle subtended by maximum cone of rays entering the objective.

For λ = 5000 Å, the theoretical value of the resolution cannot exceed 2000 Å. Resolution may be increased either by decreasing refractive index of the medium in between the sample and the objective lens or by increasing λ. A drop of oil of high refractive index (e.g. cedar oil of $\mu = 1.5$) is placed between the objective and specimen to increase the resolution and then viewed.

10.4 FOCUSING OF ELECTRON BEAMS

An ordinary beam of light is focused with the help of lenses. Since electrons are charged particles, they cannot be focused using ordinary lenses. Focusing of parallel or diverged electron beams are achieved with the help of an electrostatic lens or an electromagnetic lens.

10.4.1 Electrostatic Lens

The first electrostatic lens was built by Bruch and Johnson in 1932. An electrostatic lens consists of two coaxial cylinders A and B maintained at different potentials as shown in Fig. 10.6. The potential difference of B is greater than that of A. If a beam of electrons is passing through the axis of these coaxial cylinders, it moves normal to the equipotential surface and finally comes to focus at F. By varying the potential difference of the coaxial cylinders A and B, the electron beams are focused either near to the lens or far away from the lens. This set-up acts as lens for an electron beam similar to an optical lens for visible light. In some cases, more than two coaxial cylinders are used to get good focusing of the electron beams.

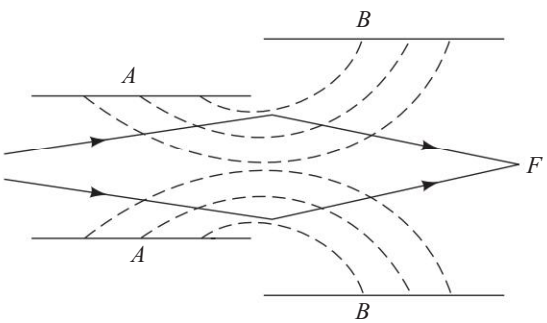


Fig. 10.6 *Electrostatic lens*

10.4.2 Electromagnetic Lens

Electromagnetic lens is based on the principle that a magnetic field produces deflection when an electron beam is obliquely passing through it. The diagrammatic representation of an electromagnetic lens is shown in Fig. 10.7. There is a small hole (separation) between the pole pieces of the magnetic lens of an electromagnet for the electrons to pass through. The pole pieces of the electromagnet consists of iron shells wounded with the coil of wire and air gaps for air circulation. If the electromagnet is switched on, the magnetic field is established in the fine perforated section. When an electron beam is passing through the magnetic field, it takes a spiral path and brings to focus just like glass lenses focuses a beam of ordinary light.

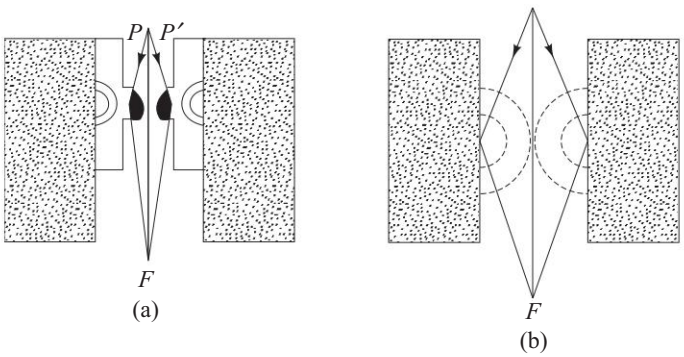


Fig. 10.7 *Electromagnetic lens*

10.5 ELECTRON MICROSCOPE

The principle of an electron microscope (EM) is similar to that of an optical microscope. An optical microscope uses optical lenses at various stages of the microscope for magnification. In an electron microscope, either an electrostatic lens or an electromagnetic lens is used at different stages of the microscope. Using an optical microscope, a magnification of about 1500 times is achieved, whereas an electron microscope magnifies the image up to 100000 times. Since the wavelength of the electron beam is 10^4 to 10^5 times smaller than that of an ordinary light ray, greater resolution can be achieved using electron microscope.

The first electron microscope was designed in the year 1932. But today, with considerable improvement in technique, the electron microscopes with a number of facilities are available in the market. Most of the instrument uses the electromagnetic lens for magnification, even though both electrostatic and electromagnetic lenses can be used for magnification.

A schematic diagram of an electron microscope is shown in Fig. 10.8 with an optical microscope. In an electron microscope, E is an electron gun and it produces electron beams, whereas in an optical microscope, an incandescent lamp S is used as a source. An electron microscope is operated at very high vacuum. Hence, the apparatus is evacuated to very high vacuum using a vacuum pump P. The electromagnetic lens in the electron microscope acts as a condensing lens, similar to the lens L_1 in the optical microscope. The condensed electron beam in the electron microscope is made to fall

The McGraw·Hill Companies

on the sample kept at O. The sample is mounted on a thin film of cellulose supported on a wire-gauze. High-speed electrons pass through the sample and the remaining are stopped by it. The electron beam passing through the specimen is magnified 100 times by the electromagnetic objective lens L_2. Then, the image is made to fall on an intermediate screen S_1. The electromagnetic lens L_3 again magnifies the image 100 times and then the image is made to fall on a fluorescent screen S_2. The lens L_3 is known as the projector lens, since it project a small portion of the first image on the screen S_2. The window W can view the image on the screen S_2. The image seen through the window has a magnification of about 10000 time. This image is further magnified 10 times with the help of an optical lens, so as to get a total magnification of 10^5 times. When photomicrographs are required, the viewing screen is replaced by a photographic plate.

An electron microscope is very useful for research on smoke, dust particles, on pigments for paints and on disease-bearing bacteria and viruses. The surface of the crystalline materials is studied by it.

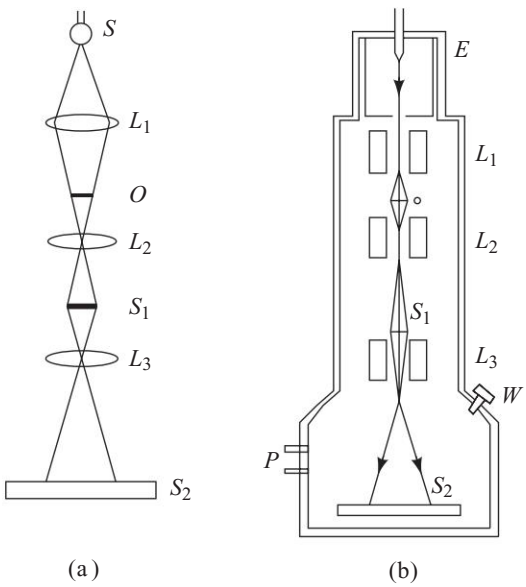


Fig. 10.8 *(a) Optical microscope (b) Electron microscope*

Though the principle of the electron microscope is same as that of an optical microscope, it differs from the optical microscope with regard to its size, weight, appearance and cost.

10.6 SCANNING ELECTRON MICROSCOPE

Knoll invented scanning electron microscope (SEM) in 1935. The schematic representation of a scanning electron microscope is shown in Fig. 10.9. It is used to produce a three dimensional image of a specimen of any size and thickness. The electrons produced by the hot filament are accelerated by electric and magnetic fields. The emerging electrons with a spot size of nearly 10 nm is made to be incident on a

The McGraw·Hill Companies

conducting sample under study. If the sample is non-conducting, a thin layer of 5–50 nm thick gold or any other metal coating is made and then the sample is used for study. A scanning coil deflects the electron beam, so that the sample is scanned point by point. The electrons striking on the sample produce secondary electrons. The number of secondary electrons produced depends upon the geometry and other properties of the samples. The secondary electrons are collected by a positively charged electron detector, which accelerates the electrons to very high energy in the order of 10 keV. Then, the electrons are made to strike a scintillator. The scintillator produces a large number of photons. These photons are made to fall on a photomultiplier tube. The photomultiplier tube converts the electrons into a highly amplified electric signal. This amplified electric signal is passed into a cathode ray tube, through which the image of the scanned surface is seen.

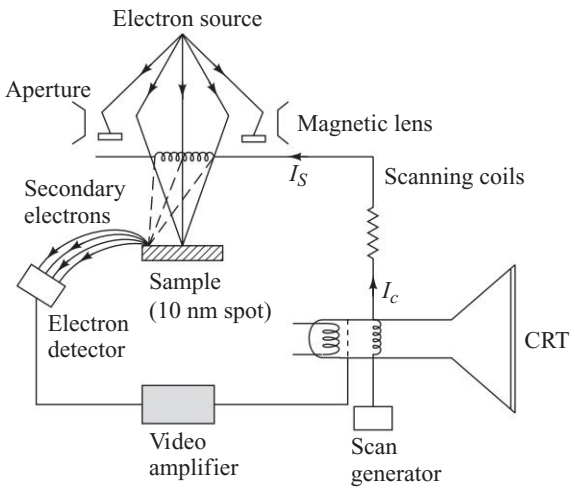


Fig. 10.9 *Scanning electron microscope*

The resolution of the image is about 10–20 nm. The magnification of the microscope can be continuously varied from 15 to 10^5 times. The magnification depends upon the ratio of the variable current I_s in the scanning coil to the current I_c in the deflection coil of the cathode ray tube.

Nowadays, due to considerable development in electron optics, the electron probes with a diameter of nearly 2.5 nm are used in scanning electron microscope. This instrument is known as electron probe microanalysis (EPMA). It is used to find the chemical compositions of alloys, metals, semiconductors, etc.

10.7 TRANSMISSION ELECTRON MICROSCOPE

The Transmission Electron Microscopy (TEM) is one of the modern characterisation tools used to obtain structural images of the materials. One can explore the surface nature, structural and the properties of the materials. It finds wide applications in almost all areas like chemistry, materials science, geology and biology. The basic

The McGraw·Hill Companies

composition of the materials can be explored with the help of detectors through the energy loss spectrum of the transmitted electrons. On the other hand, the modern TEMs are used to produce the images in the scanning range of 0.1nm with a magnification of 50 million times.

The schematic representation of the TEM is shown in Fig. 10.10. A stream of monochromatic electron beams is produced by the electron gun, namely, virtual source located at the top. Two condenser lenses are used to obtain coherent electron beam as shown in Fig. 10.10. The first lens is used to determine the spot size while the second lens is used to change the size of the spot on the sample. The electron beam is made to incident on the selected points on the sample. A part of the beam is transmitted into the sample while the remaining part is being scattered. The transmitted portion is focused by the objective lens. The objective aperture enhances the contrast of the image by preventing larger diffracted electrons. Thus, the image is formed after passing through the projector lenses. The transmitted beam of electron strikes the phosphor image screen and thereby it generates light. The obtained image consist of dark and bright patterns. The dark pattern refers to the less transmission of electron from the sample which may be due to denser or thicker nature of the sample. Similarly, the bright pattern reveals the high transmission of electron beam through the sample.

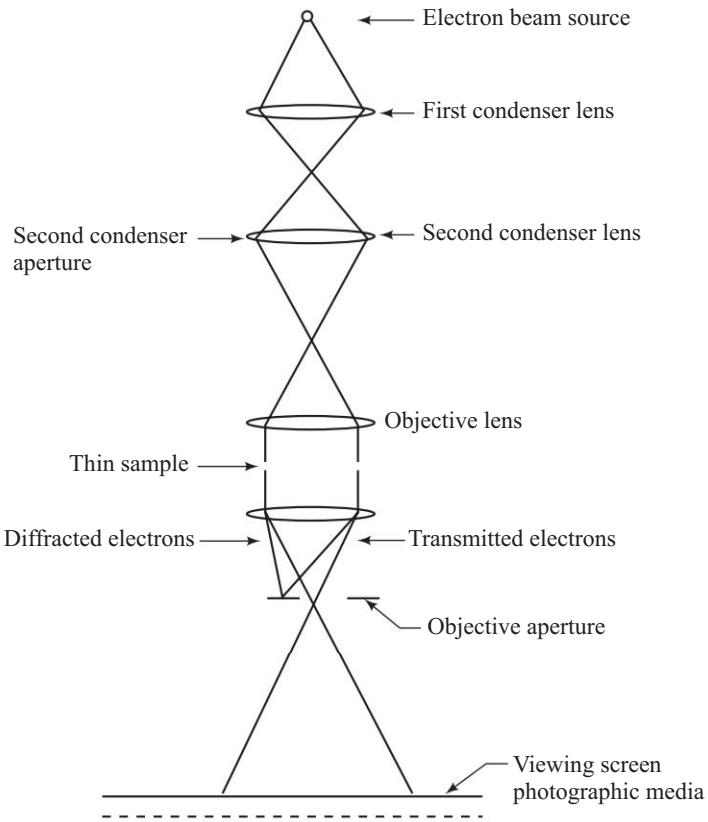


Fig. 10.10 *Transmission electron microscope*

A high intensity electron beam is made incident on the sample. The interaction of electron beam in crystalline material is based on the diffraction principle. The intensity of the diffraction pattern depends on the orientation of atomic planes in the crystal. The intensity of the transmitted beam may be altered by the volume and density of the materials through which it passes.

A high contrast image is obtained by TEM only by allowing the transmitted electrons through the sample along the optic axis. The scattered electrons are deflected from the optical axis and hence, prevented from reaching the phosphor screen using the objective apertures. The scattered electrons are not taken into account during the formation of an image.

In High Resolution Transmission Electron Microscopy (HRTEM), the image can be formed due to the differences in the phases of the scattered electron waves through the sample during transmission. The images obtained using HRTEM will facilitate to explore the crystal structure of the material.

Applications The applications of the TEM are given below:

(i) One can obtain the imaging of individual molecules (or) macromolecular assemblies.
(ii) It is used heavily in materials, metallurgy and biological sciences.
(iii) Computer modeling of the images is an added advantage in TEM characterisation.

Limitations The limitations of the TEM are given below:

(i) The sample must be thin enough to characterise using TEM and hence, it requires extensive sample preparation (electron transparent)
(ii) The structure may be changed during the process of sample preparation.
(iii) In biological materials, the electron beam may damage the samples.

10.8 SCANNING TRANSMISSION ELECTRON MICROSCOPE

Scanning Transmission Electron Microscope (STEM) is an advanced version of Transmission Electron Microscope (TEM) used for research and development in all fields. Like other electron microscopes, the high energy electron beam is made to be incident on the sample not only at a particular point, but line by line with the help of scanning coils. The incident high energy electron beam is transmitted into the sample. The transmitted electrons are collected by number of detectors and hence, the images are formed. The scanning of electron beam across the sample provides details of the material such as mass determination and mapping. These data are used extensively for the direct correlation of image and the structures quantitatively. In STEM, the image and the structural informations of the sample can be obtained simultaneously. It is one of the most versatile tools used for the determination and mass mapping of bimolecular structures.

The schematic diagram for the STEM is shown in Fig. 10.11. The field emission gun is placed at the top of the instrument, which emits the high energetic electron beam. A fine narrow beam of electrons of size 2.5 Å diameter is obtained using an objective lens.

The McGraw·Hill Companies

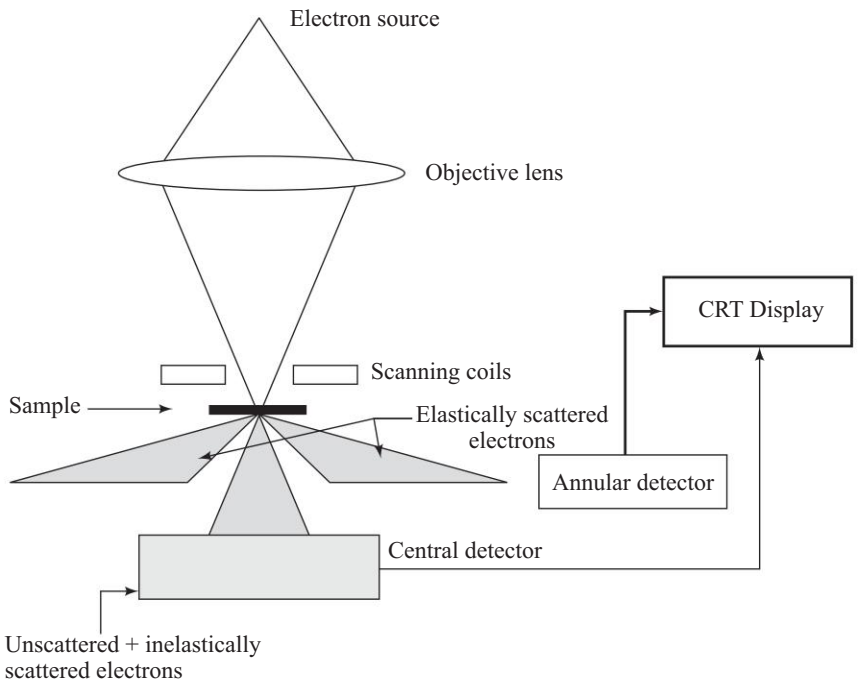


Fig. 10.11 *Schematic diagram of scanning transmission electron microscope*

The direction of the beam of electrons can be controlled employing the scanning coils. The transmitted electrons are of three categories, namely, elastically scattered, in-elastically scattered and the unscattered electrons. The various detectors which are placed in different angles below the sample, collect the above three categories of electrons.

A high energetic electron beam is made to be incident on the sample and then scan in a 2-dimensional raster across the sample with the help of the scanning coils. During the transmission, the electron scattering has occurred both in elastically and in-elastically. The scattering angle of the elastically scattered electrons is relatively low with the same scattering region of unscattered electrons. The above electrons are collected by the same central detector. However the elastically scattered electrons are scattered over a large solid angle and these can be collected by annular detector below the sample. All the scattered electrons are collected by a variety of detectors which are placed behind the sample. The image is generated by step-by-step movement of the focused beam over the sample. Variety of signals collected simultaneously can be discriminated depending on the different scattering angle and/or energy loss. The different structural and chemical informations of the sample are obtained from the images due to the various scattered angles. The simultaneous and controlled acquisition of information can give good quantitative analysis which cannot be obtained by other microscopes. The inelastic and unscattered electrons can be differentiated using their difference in energies with the help of magnetic prism near the central detector. During the scanning process, the analog signal from the detector is fed into the Cathode Ray

The McGraw·Hill Companies

Tube (CRT) for display. The magnification factor is the ratio of the raster of CRT and electron probe as in written as,

$$\text{Magnification} = \frac{\text{CRT raster}}{\text{Electron probe raster}} \tag{10.1}$$

The following are the **advantages** of STEM:

(i) High elemental mapping resolution of any analytical technique
(ii) Provide crystallographic information of even small area of the sample
(iii) It is used to characterize the nanoparticle, agglomeration due to annealing, etc

The following are the **limitations** of the STEM:

(i) The sample preparation is tedious as that of STEM
(ii) Some materials not stable to high energetic electron beam

Keypoints to Remember

- A sextant is used as an optical instrument working on the principle of light reflection and is used to measure inaccessible towers or buildings and the altitude of sun.
- The metallurgical microscope is based on the principle of reflection of light, while a biological microscope is based on the principle of transmission of light.
- Electrostatic and electromagnetic lenses are used to focus electron beams in a microscope.
- An electron microscope is different from optical microscope in terms of size, weight, appearance and cost.
- Electron Probe Microscope Analysis (EPMA) is used to find the chemical composition of alloys, metals and semiconductors.
- Transmission Electron Microscope (TEM) is one of the modern characterisation tools used to obtain the information such as surface nature, structure and properties of materials.
- In TEM, the contrast image is obtained by allowing the transmitted electrons through the sample along the optic axis.
- In High Resolution Transmission Electrons Microscope (HRTEM), the image is obtained due to the difference in the phases of the scattered electrons waves through the sample.
- STEM is used for the determination and mass mapping of biomolecular structures.
- In STEM, there are three types of transmitted electron namely electrically scattered, in-elastically scattered and unscattered electrons.

The McGraw·Hill Companies

Objectives Questions

10.1 The altitude of the sun is measured using _____.

10.2 A sextant is works on the principle of_____.

10.3 The height of an inaccessible tower is measured using_____.

10.4 Using a sextant, the height of a building is measured using the reflection

a) $h = \dfrac{d}{\cos\theta_1 + \cos\theta_2}$ b) $h = \dfrac{d}{\cos\theta_1 - \cos\theta_2}$

c) $h = \dfrac{d}{\cot\theta_1 + \cot\theta_2}$ d) $h = \dfrac{d}{\cot\theta_2 - \cot\theta_1}$

10.5 The transmitted light is used to view the specimen in _____ microscope.

10.6 The reflected light is used to view the specimen in _____ microscope.

10.7 The magnification of metallurgical microscope is equal to

a) $\dfrac{DN}{F}$ b) $\dfrac{DN}{F^2}$

c) DN d) DNF

10.8 The resolution of the image in a metallurgical microscope is

a) $\dfrac{0.5\ \lambda^2}{\mu \sin a}$ b) $\dfrac{0.5\ \lambda}{\mu \sin a}$

c) $\dfrac{0.5}{\mu \sin a}$ d) $\dfrac{0.5\ \lambda}{\sin a}$

10.9 Electron beams are focused using _____ or _____ lenses.

10.10 The wavelength of an electron beam is smaller than ordinary light. True / False?

10.11 The electron microscope is not different from an optical microscope in terms of

a) size b) weight

c) principle d) cost

10.12 Acronym for SEM is _____.

10.13 In SEM, the energy of the accelerated electrons is in the order of _____.

10.14 The chemical composition of alloys and metals are analysed using _____.

10.15 Acronym for EPMA is_____.
10.16 Acronym for TEM is_____.
10.17 HRTEM is used to explore_____ of material.
10.18 Expand DRTEM_____.
10.19 Acronym for STEM is_____.

KEY:

10.1 sextant
10.2 reflection
10.3 sextant
10.4 d)
10.5 biological
10.6 metallurgical
10.7 a)
10.8 b)
10.9 electrostatic or electromagnetic
10.10 True
10.11 c)
10.12 Scanning Electron Microscope
10.13 10 ke V
10.14 Electron Probe Microscope
10.15 Electron Probe Microanalysis
10.16 Transmission Electron Microscope
10.17 crystal structure
10.18 High Resolution Transmission Electron Microscope
10.19 Scanning Transmission Electron Microscope

Short Questions

10.1 What is a sextant?
10.2 What are the uses of a sextant?
10.3 Explain how the altitude of the sun is measured using a sextant.
10.4 Explain how the height of a tower is measured using a sextant.
10.5 What is an optical microscope?
10.6 What is a metallurgical microscope?
10.7 What is the difference between an optical microscope and a metallurgical microscope?
10.8 What is an electron microscope?
10.9 What are the uses of an electron microscope?
10.10 What is a scanning electron microscope?
10.11 What is the principle behind transmission electron microscope.
10.12 Explain the principle of scanning transmission electron microscope.
10.13 Compare the merits and demerits of SEM and STEM.

Descriptive Questions

10.1 Explain with neat sketch, the construction and working of a sextant. Explain how is a sextant used to find the height of a tower and the altitude of the sun.

The McGraw·Hill Companies

10.2 Explain the construction and working of a metallurgical microscope.
10.3 Explain with neat sketch, the construction and working of an electron microscope.
10.4 Explain with neat sketch the construction and working of a scanning electron microscope.
10.5 Explain with a neat sketch, the working of TEM with STEM. Mention their applications and limitations.

Chapter

11

Laser

OBJECTIVES

- To understand the principle, theory, operations and applications of laser
- To discuss the fundamental terms such as absorption, spontaneous and stimulated emission
- To derive Einstein expressions for stimulated emission and to study its importance
- To explore the different methods of population inversion
- To derive Schawlow and Townes condition for achieving laser action
- To study the different types of lasers and their applications
- To discuss the laser applications such as holography, and the use of lasers in the fields of computers, industry, medicine, etc

11.1 INTRODUCTION

Laser is an acronym for Light Amplification by Stimulated Emission of Radiation. The theoretical explanation for laser oscillation was given by A L Schawlow and C H Townes in the year 1958. The first laser, namely, ruby laser was demonstrated by T H Maiman, in the year 1960. The discovery of laser made an enormous impact in the scientific world and showed that the function of optics was very much alive.

The laser beam has the properties given below which distinguish it from an ordinary beam of light.

(i) Directionality The light beam can travel as a parallel beam up to a distance of d^2/λ, where d is the diameter of the aperture through which the light is passing and λ the wavelength of the light used. After travelling the distance d^2/λ, the light beam spreads radially. In an ordinary light beam, the angular spread is given by $\Delta\theta = \frac{\lambda}{d}$. For a typical laser beam, the angular spread is 1 mm per 1 m, but for an ordinary source of light, the angular spread is 1 m per 1 m. This shows the directionality of the laser beam. For example, the laser beam can be focused to moon from the earth with an angular spread of a few kilometres.

The McGraw·Hill Companies

(ii) Intensity The intensity of the laser beam is very high. If a person is allowed to observe an ordinary light, emitted by a 100 W bulb at a distance. One foot from the source, he can perceive only one thousand watt of light. While, if the person is allowed to observe the laser beam from the same distance, the entire laser beam penetrates through his eye. It will damage the eye of the observer. This shows the high intensity of the laser beam.

(iii) Monochromaticity The laser beam is strictly monochromatic (single frequency) than any other conventional monochromatic source. The bandwidth ($\Delta\nu = 0$) of the laser beam is narrow, while ordinary light spreads over wide range of frequencies. Obtaining absolute monochromaticity is difficult; however, one can get high degree of monochromaticity in laser beam with very small bandwidth, of about one kilocycle per second. For example, the line width $\Delta\lambda$ emitted by ruby laser is 5×10^{-4} Å.

(iv) Coherence The degree of coherence of laser beam is very high than the other sources. The coherence of laser emission results in extremely high power (5×10^{12} W cm^{-2}). The light from a laser source consists of wave trains that are identical in phase and directions of propagation. Further, it can be focused to a very small area of 0.7 mm thickness.

It is, indeed, these properties, that differentiate the laser beam from other conventional light beams. Let us discuss the principle, generation and application of the laser in this chapter in detail.

11.2 PRINCIPLE OF LASER

From the theory of interaction of radiation with matter, we can get an idea regarding the working of laser. Consider an atom that has only two energy levels, E_1 and E_2. When it is exposed to radiation having a stream of photons, each with energy $h\nu$, three distinct processes can take place.

(i) Absorption
(ii) Spontaneous emission, and
(iii) Stimulated emission

(i) Absorption An atom or molecule in the ground state E_1 can absorbs a photon of energy $h\nu$ and go to the higher energy state E_2. This process is known as *absorption* and is illustrated in Fig. 11.1

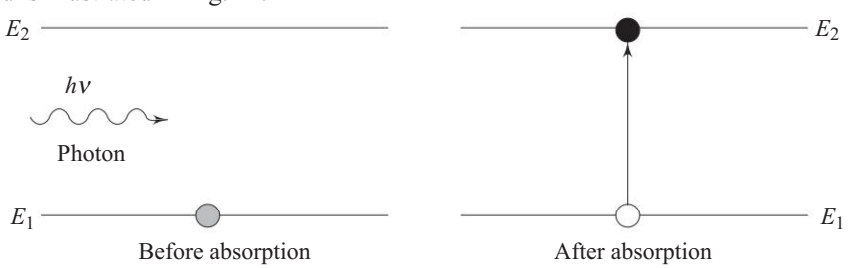


Fig. 11.1 *Absorption*

The rate of upward transition R_{12} from ground state E_1 to excited state E_2 is proportional to the population of the lower energy level N_1 (number of atoms per unit volume) and to the energy density of radiation ρ_ν,

i.e. $R_{12} \alpha \rho_\nu$

αN_1

Thus $R_{12} = B_{12} \rho_\nu N_1$ (11.1)

where B_{12} is the probability of absorption per unit time.

Normally, the higher energy state is an unstable state and hence, the atoms will make a transition back to the lower energy state with the emission of a photon. Such an emission can take place by one of the two methods given below.

(ii) Spontaneous Emission In spontaneous emission, the atoms or molecules in the higher energy state E_2 eventually return to the ground state by emitting their excess energy spontaneously. This process is independent of external radiation. The rate of the spontaneous emission is directly proportional to the population of the energy level E_2,

i.e. $R_{21}\,(\text{SP}) \alpha N_2$

$R_{21}\,(\text{SP}) = A_{21} N_2$ (11.2)

where A_{21} is the probability per unit time that the atoms will spontaneously fall to the ground state and N_2 the number of atoms per unit volume in the state E_2. This process is illustrated in Fig. 11.2

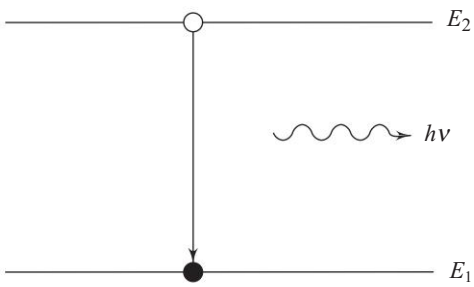


Fig. 11.2 *Spontaneous emission*

(iii) Stimulated Emission In stimulated emission, a photon having energy E, equal to the difference in energy between the two levels E_2 and E_1, stimulate an atom in the higher state to make a transition to the lower state with the creation of a second photon as shown in Fig. 11.3.

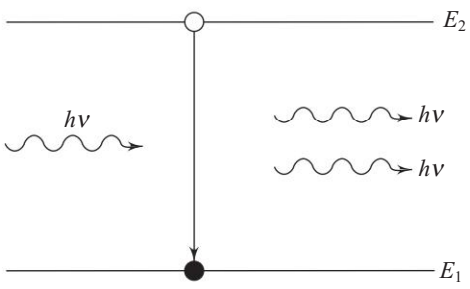


Fig. 11.3 *Stimulated emission*

The McGraw·Hill Companies

The rate of stimulated emission R_{21} (ST) is given as,

$$R_{21}(\text{ST}) = B_{21}\rho_\nu N_2 \tag{11.3}$$

where B_{21} is the probability per unit time that the atoms undergo transition from higher energy state to lower state by stimulated emission.

Under conditions of thermal equilibrium, the population of energy levels obey Boltzmann's distribution function.

11.3 EINSTEIN'S THEORY OF STIMULATED EMISSION

In 1917, Einstein proposed a mathematical expression for the existence of stimulated emission of light. This expression is known as *Einstein's expression*.

Consider a two-level energy system (E_1 and E_2). Let N_1 and N_2 be the number of atoms in the ground state and excited state, respectively. Let us assume that only the spontaneous emission is present and there is no stimulated emission of light.

At thermal equilibrium condition, the rate of absorption = the rate of emission of light

From Eqs (11.1) and (11.2), we can write,

$$\rho_\nu \times B_{12}N_1 = A_{21}N_2$$

$$\rho_\nu = \frac{A_{21}N_2}{B_{12}N_1} \tag{11.4}$$

According to Boltzmann's distribution function, the population of atoms in the upper and lower energy levels are related by,

$$\frac{N_2}{N_1} = \frac{e^{-E_2/kT}}{e^{-E_1/kT}} \tag{11.5}$$

Substituting N_2/N_1 in Eq. (11.4), we get,

$$\rho_\nu = \frac{A_{21}}{B_{21}} e^{-(E_2 - E_1)/kT} \tag{11.6}$$

$$\rho_\nu = \frac{A_{21}}{B_{21}} \frac{1}{e^{h\nu/kT}}$$

According to black-body radiation, the energy density

$$\rho_\nu = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/kT} - 1} \tag{11.7}$$

where h is the Planck's constant and c the velocity of light. Comparing the above two equations [Eqs (11.6) and (11.7)], one can observe that they are not in agreement.

To rectify this discrepancy, Einstein proposed another kind of emission known as *stimulated emission of radiation*. Therefore, the total emission is the sum of the spontaneous and stimulated emissions of radiation.

At thermal equilibrium condition, the rate of absorption = the rate of emission

From Eqs (11.1), (11.2) and (11.3),

$$B_{12}N_1\rho_\nu = A_{21}N_2 + B_{21}N_2\rho_\nu$$

The McGraw·Hill Companies

$$\rho_v = \frac{A_{21}N_2}{B_{12}N_1 - B_{21}N_2} \tag{11.8}$$

Dividing each and every term on R.H.S. of Eq. (11.8.) by N_2 , we get,

$$\rho_v = \frac{A_{21}}{B_{12}(N_1 / N_2) - B_{21}}$$

Substituting for N_1/N_2 from Eq. (11.5), we get,

$$\rho_v = \frac{A_{21}}{B_{12}[e^{(E_2 - E_1)/kT}] - B_{21}}$$

We know that $E_2 - E_1 = h\nu$

$$\rho_v = \frac{A_{21}}{B_{12}e^{h\nu/kT} - B_{21}} \tag{11.9}$$

The coefficients A_{21}, B_{12} and B_{21} are known as *Einstein's coefficients*.

Comparing the above equation with Eq. (11.7), we get,

$$B_{12} = B_{21}$$

and

$$\frac{A_{21}}{B_{21}} = \frac{8\pi h\nu^3}{c^3} \tag{11.10}$$

From Eqs (11.7) and (11.9), the ratio of the stimulated emission to spontaneous emission is given by,

$$\frac{R_{21}(\text{ST})}{R_{21}(\text{SP})} = \frac{B_{21}N_2\rho_v}{A_{21}N_2} = \frac{1}{e^{h\nu/kT} - 1} \tag{11.11}$$

From Eq. (11.11), Einstein proved the existence of the stimulated emission of radiation.

The spontaneous emission produces incoherent light, while the stimulated emission produces coherent light. In an ordinary conventional light source, the spontaneous emission is dominant. For laser action, stimulated emission should be predominant over spontaneous emission and absorption. To achieve this, an artificial condition, known as *population inversion,* is required.

11.4 POPULATION INVERSION

Consider a two-level energy system (E_1 and E_2). Suppose a photon of energy equal to the energy difference between the two levels is incident on the system, then which event is more likely to take place: absorption or stimulated emission? Einstein showed that under normal circumstances, both the processes are equally probable.

In a system containing a very large number of atoms, the dominant process will depend on the virtual number of atoms in the upper and lower states. A large population in the upper level ($N_1 < N_2$) will result in stimulated emission dominating over the absorption. If there are more number of atoms in the lower level, i.e., ground state ($N_1 > N_2$), there will be more absorption than stimulated emission.

Under conditions of thermal equilibrium, Boltzmann's distribution function relating N_1 and N_2 given by Eq. (11.5) is obeyed.

The McGraw·Hill Companies

$$\frac{N_2}{N_1} = e^{-(E_2 - E_1)/kT} \tag{11.12}$$

where k and T are the Boltzmann's constant and the absolute temperature, respectively. For stimulated emission to dominate, it is necessary to increase the population of the upper energy level, so that it is greater than that of the lower energy level. This is known as *population inversion*.

In the above equation, if T is negative, then stimulated emission would dominate over absorption. Such a condition is practically not possible. The population inversion can be obtained even at room temperature by several methods discussed in the following section.

11.5 METHODS OF ACHIEVING POPULATION INVERSION

There are several methods for achieving the condition of population inversion necessary for the laser action to takes place. Some of the most commonly used methods are as follows:

(i) Optical pumping In case of optical pumping, an external optical source like xenon flash lamp is employed to produce a high population in the higher energy level of the laser medium.

This method of excitation, shown in Fig. 11.4a, is used in solid state lasers of which the ruby laser is a prototype.

(ii) Direct electron excitation The direct electron excitation in a gaseous discharge may be used to produce the desired inversion. This method is used in some of the gaseous ion lasers, such as an argon laser. In this type of excitation, as shown in Fig. 11.4b, the laser medium itself carries the discharge current under suitable conditions of pressure and temperture.

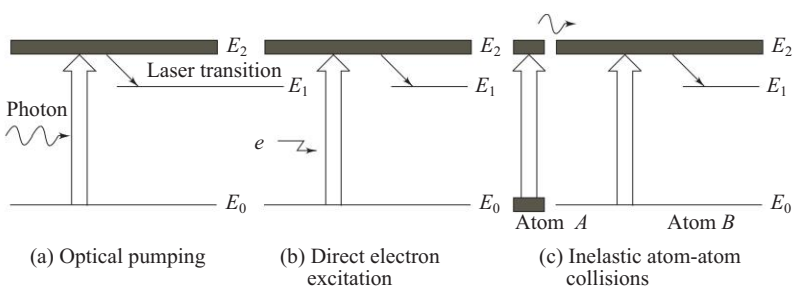


Fig. 11.4 *Population inversion—different-methods of production*

In this method, the electrons directly excite the active atoms to achieve higher population in certain higher energy levels compared to a lower energy level.

(iii) Inelastic Atom–Atom Collisions Here, electric discharge method is employed to cause collision and excitation of the atom. In this method, a combination of two

The McGraw·Hill Companies

types of gases is used, say A and B, both having the same excited state A^* and B^* that coincide or nearly coincide.

In the first step, during electric discharge, A gets excited to A^* (metastable) due to collision with electrons.

$$A + e \longrightarrow A^* + e_1$$
$$A^* + B \longrightarrow A + B^* \qquad (11.13)$$

The excited A^* atoms now collide with B atoms so that the latter atom gets excited to higher energy B^*. This type of transition is used in the He–Ne laser.

(iv) *Chemical reactions* In this method, the molecules undergo chemical changes in which one of the products of the reaction is a molecule or an atom that is left in an excited state under appropriate conditions. Under such conditions, population inversion can occur. An example for this type of lasers is hydrogen fluoride chemical laser, in which hydrogen fluoride molecules in the excited state result from the following chemical reaction:

$$H_2 + F_2 \longrightarrow 2HF \qquad (11.14)$$

11.6 THRESHOLD CONDITION (SCHAWLOW AND TOWNES CONDITION)

The amplitude of the light beam must be amplified in order to get the laser action. Schawlow and Townes discussed the necessary conditions for achieving the laser action. The material in which population inversion can takes place is called *active medium*. This medium may be a gas, liquid or solid, in which light amplification takes place and is placed in between the two mirrors of which one is partial reflecting, while the other one is 100% reflecting, as shown in Fig. 11.15. The light beam bounds back and forth between the two mirrors and hence, the intensity of the beam gets increased. However, there will be losses in each reflection but one can obtain a threshold condition for the laser action to takesplace.

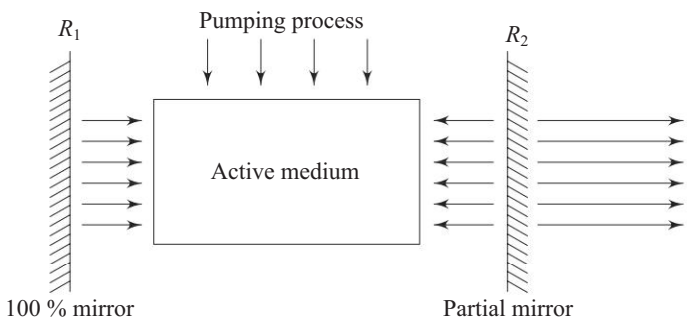


Fig. 11.5 *Optical resonator*

The threshold condition is that the losses in the medium should be overcome by induced transition in the amplifying medium. The total losses are due to the following reasons.

The McGraw·Hill Companies

(i) Transition, absorption and scattering by the mirrors
(ii) Absorption within the medium
(iii) Scattering due to optical inhomogeneity, and
(iv) Diffraction losses by the mirrors

All the losses depend on the lifetime of the photons existing in the cavity and it is represented as τ_{Photon} .

The total rate of losses of photons from the laser medium per second is

$$\frac{1}{\tau_{\text{photon}}} \tag{11.15}$$

By taking two level energy system, the stimulated transition rate ω' is given by

$$\omega' = \rho(\nu) B N_2 \tag{11.16}$$

where $B_{12} = B_{21} = B$. Substituting the B value in Einstein's coefficient Eq. (11.10).

$$\frac{8\pi h\nu^3}{c^3} = \frac{A_{21}}{B_{21}} = \frac{A}{B}$$

$$\omega' = \rho(\nu) \frac{Ac^3}{8\pi h\nu^3} N_2$$

$$\frac{\omega'}{A} = \rho(\nu) \frac{c^3}{8\pi h\nu^3} N_2 \tag{11.17}$$

In practice, the laser line width is very narrow than the line width of atomic transition. Let $g(\nu)$ represent the probability of frequency of the curve during absorption or emission. It is also called *shape of absorption* or *emission curve*, as shown in Fig.11.6.

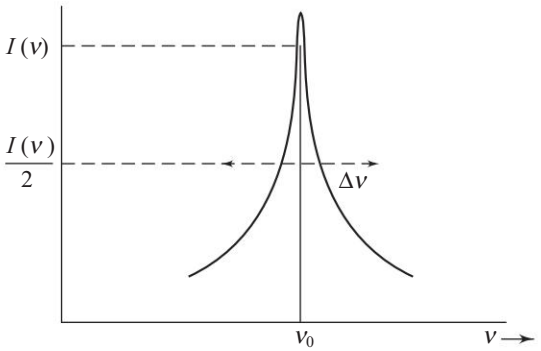


Fig. 11.6 *Absorption curve*

The factor $g(\nu)\, d\nu$ represents the probability of absorption or emission by photons whose energy is $h\nu$ and $h\nu + h\, d\nu$. The curve is usually normalised to unity,

i.e. $$\int_{-\infty}^{+\infty} g(\nu)\, d\nu = 1 \tag{11.18}$$

On account of uncertainty, it is necessary to replace ω' as,

$$\omega' = \omega' g(\nu)\, d\nu \tag{11.19}$$

Similarly, for the spontaneous emission, whose frequency ranges from ν to $\nu + d\nu$ as,

$$A = A g_{\mathrm{L}}(\nu) d\nu \tag{11.20}$$

where $g_{\mathrm{L}}(\nu)\, d\nu$ is the Longevin function for line widths and is equl to

$$\int_{-\infty}^{+\infty} g_L(\nu)\, d\nu = \frac{2}{\pi} \tag{11.21}$$

Equation (11.17) can be writte as,

$$\frac{\omega' g(\nu) d\nu}{A g_L(\nu) d\nu} = \frac{\rho(\nu) c^3}{8\pi h\nu^3} N_2 \tag{11.22}$$

The radiation density $\rho(\nu)$ is simply related to radiation flux or intensity $I(\nu)$ as,

$$I(\nu) = \rho(\nu)\, c \tag{11.23}$$

where c is the velocity of light in the laser medium.
We know, for spontaneous emission,

$$A = \frac{1}{\tau_{\mathrm{SP}}} \tag{11.24}$$

From Eqs (11.23) and (11.24), Eq. (11.22) can be written as,

$$\omega' g(\nu) d\nu = \frac{I(\nu) c^2}{8\pi h\nu^3} \frac{1}{\tau_{\mathrm{SP}}} N_2 g_{\mathrm{L}}(\nu) d\nu \tag{11.25}$$

The total transition rate is obtained by integrating on both sides of Eq. (11.25) as,

$$\omega'(\nu) = \frac{c^2}{8\pi h\nu^3} \frac{2}{\pi} \frac{1}{\tau_{\mathrm{SP}}} I(\nu) N_2 \tag{11.26}$$

Consider a two-level energy system. Let N_1 and N_2 be the populations of ground state and the first excited state. Consider E_1 and E_2 as the energies of the ground state and the first excited state, respectively. Assume a light beam of energy density $\rho(\nu)$ is made to incident on this system. The rate of absorption = $B_{12}\, \rho(\nu)\, N_1$. Absorption of a photon removes an energy $h\nu$ from the incident beam. Therefore, the total energy removed from the incident beam = $B_{12} \rho(\nu) N_1 h\nu$.

Due to the radiation incident on the two-level energy system, stimulated emission takes place. The rate of stimulated emission = $B_{21} r(\rho) \nu_2$. As in the absorption process, the emission of photon produces an energy $h\nu$ and therefore, the total energy added to the incident beam is = $B_{21} \rho(\nu) N_2 h\nu$. The rate of change of energy density in a frequency interval $d\nu$ or ν is

$$\frac{d(\rho_\nu\, d\nu)}{dt} = B_{21} \rho(\nu) N_2 h\nu - B_{12} \rho(\nu) N_1 h$$

i.e.

$$\frac{d\rho_\nu}{dt} = \frac{B\rho(\nu)}{d\nu} (N_2 - N_1) h\nu \tag{11.27}$$

where $B_{21} = B_{12} = B.$

Substituting the vale $\rho_v = \frac{I_v}{c}$ in Eq. (11.27),

we get,

$$\frac{dI_v}{c\,dt} = \frac{B\rho(v)}{dv}(N_2 - N_1)hv$$

i.e., $$\frac{dI_v}{dt} = \frac{B\rho(v)}{dv}(N_2 - N_1)hvc \qquad (11.28)$$

The net rate of change in the energy level will be numerically equal to the product of net photons emitted and its energy.

Usually, there will be always a degeneracy in the energy levels. Let g_1 and g_2 represent the degeneracy in the energy levels E_1 an E_2.

$$\left(\frac{dI}{dt}\right)_{\text{gain}} = \frac{hv}{dv}\left[N_2 - N_1\frac{g_2}{g_1}\right]\left(\frac{\omega'(v)}{N_2}\right)c \qquad (11.29)$$

This results in increased intensity on account of population inversion only if

$$\left(N_2 \geq N_1\frac{g_2}{g_1}\right) \qquad (11.30)$$

If this condition is not satisfied, the laser medium will simply lose energy or that there will be transition losses.

The rate of loss of intensty is

$$\left(\frac{dI}{dt}\right)_{\text{loss}} = \frac{I(v)}{\tau_{\text{photon}}} \qquad (11.31)$$

For laser action,

$$\left(\frac{dI}{dt}\right)_{\text{gain}} - \left(\frac{dI}{dt}\right)_{\text{loss}} \geq 0$$

or $$\frac{hv}{dv}\left(N_2 - N_1\frac{g_2}{g_1}\right)\left(\frac{\omega'(v)}{N_2}c\right) \geq \frac{I(v)}{\tau_{\text{photon}}} \qquad (11.32)$$

Substituting $\omega'(v)$ from Eq. (11.26)

$$\frac{hv}{\Delta v}\frac{hv}{\Delta v}\left(N_2 - N_1\frac{g_2}{g_1}\right)\left(\frac{c^3}{8\pi hv^3}\right)\left(\frac{2\,I(v)}{\pi\,\tau_{\text{SP}}}\right) \geq \frac{I(v)}{\tau_{\text{photon}}} \qquad (11.33)$$

Simplifying and rearranging the above equation, we get

$$\left(N_2 - N_1\frac{g_2}{g_1}\right) \geq \left(\frac{4\pi^2 v^2}{c^3}\right)\left(\frac{\tau_{\text{SP}}}{\tau_{\text{photon}}}\right)\Delta v \qquad (11.34)$$

The McGraw·Hill Companies

The threshold condition represents that the loss of photon is just balanced by the gain of photon, i.e., at threshold condition,

$$\left(N_2 - N_1 \frac{g_2}{g_1}\right) = \left(\frac{4\pi^2 \nu^2}{c^3}\right)\left(\frac{\tau_{SP}}{\tau_{photon}}\right) \Delta\nu \quad (11.35)$$

For non-degenerate energy levels, Eq. (11.35) can be written as,

$$N_2 - N_1 = \left(\frac{4\pi^2 \nu^2}{c^3}\right)\left(\frac{\tau_{SP}}{\tau_{photon}}\right) \Delta\nu \quad (11.36)$$

In terms of the angular frequency, Eq. (11.36) can be written as,

$$N_2 - N_1 = \left(\frac{\omega^2}{2\pi c^3}\right)\left(\frac{\tau_{SP}}{\tau_{photon}}\right) \text{"} \omega \quad (11.37)$$

The laser oscillation takes place at a particular single frequency. To represent this, let us replace ω by ω_0

$$N_2 - N_1 = \left(\frac{\omega_0^2}{2\pi c^3}\right)\left(\frac{\tau_{SP}}{\tau_{photon}}\right) \text{"} \omega_0 \quad (11.38)$$

If n_0 is the refractive index of the amplifying medium (where $n_0 = \frac{c}{v}$), then by taking into account the refractive index of the medium, the above equation can be written as,

$$N_2 - N_1 = \left(\frac{n_0^3 \omega_0^2}{2\pi c^3}\right)\left(\frac{\tau_{SP}}{\tau_{photon}}\right) \text{"} \omega_0 \quad (11.39)$$

This is known as *Schawlow and Townes condition (threshold condition)* for population inversion near the line $g(\nu)$ at $\nu = \nu_0$ for laser oscillation.

11.7 TYPES OF LASERS

Lasers are classified into five major categories based on the type of active medium. They are

(a) Solid state lasers
(b) Gas lasers
(c) Liquid lasers
(d) Dye and chemical lasers, and
(e) Semiconductor lasers

All the above lasers are discussed in the following sections in detail.

11.7.1 Solid State Lasers

(i) Ruby Laser In a solid state laser, the active medium is a crystalline substance. It is the first successful laser achieved by Maiman in 1960. The experimental set-up shown in Fig. 11.7 has three important parts.

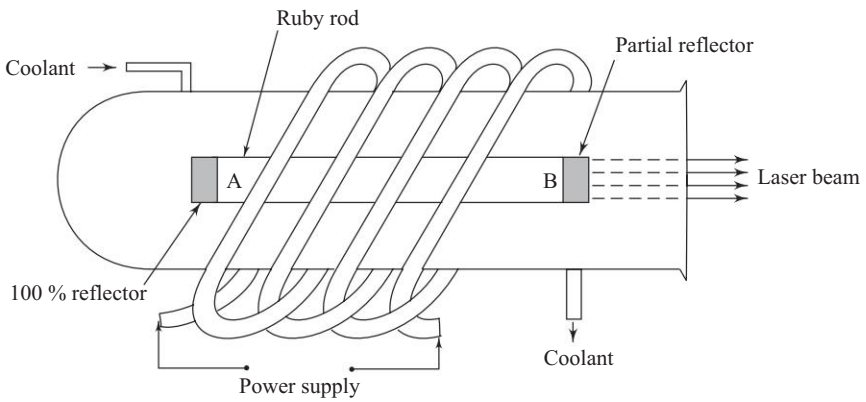


Fig. 11.7 *Ruby laser*

Ruby Rod Ruby is a crystalline substance of aluminium oxide doped with approximately 0.05% by weight of chromium oxide. The resultant pink colour is due to the presence of Cr^{3+} ions in the appropriate concentration which replace aluminium atoms in the crystal lattice.

Resonating Cavity In a ruby laser, a pink rod of 4-cm length and 0.5cm diameter is used. The end faces of the rod are strictly grounded so that they are parallel and polished to a high degree. Further, the end faces are silvered in such a way that one end face becomes fully reflecting while the other end is partially reflecting. In some cases, separate glass pieces are attached at the end faces.

Xenon Flash Tube The laser action is achieved by using optical pumping. A helical xenon flash tube that surrounds the ruby rod, provides the pumping light to raise the chromium ions to the upper energy levels. The flash of the xenon tube lasts for several milliseconds and the tube consumes several thousand joules of energy. Only a part of the energy is used to excite the Cr^3+ ions, while the rest heats up the apparatus. A separate cooling arrangement is used to reduce the temperature.

Working Principle The energy-level diagram illustrating the operation of a ruby laser is shown in Fig 11.8.

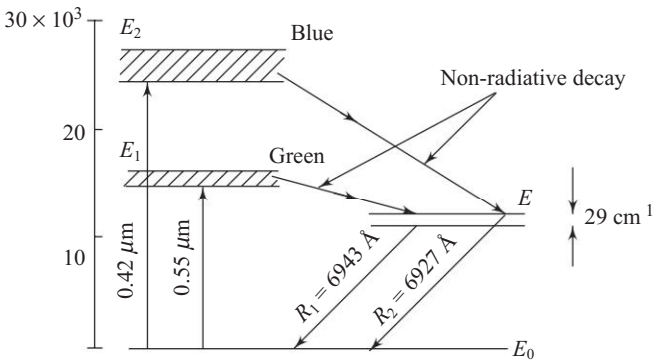


Fig. 11.8 *Energy level diagram—Ruby laser*

The McGraw·Hill Companies

The pumping light from the flash tube is absorbed by Cr^{3+} ions, raising them from the ground state E_0 to the excited states E_1 or E_2. From these levels, a rapid radiationless transition to the level E, which is a metastable state, takes place. Decay from E is relatively slow so that with sufficient excitation, population inversion between E and the ground state E_0 can occur. The photons are allowed to pass back and forth millions of times in the active medium with the help of mirrors at the ends. When the condition for laser action is satisfied, an intense pulse of light of the wavelength 6934 Å, corresponding to the transition E to E_0 is obtained.

(ii) Nd–YAG Laser The Nd-YAG laser is an optically pumped solid state laser and is used to produce very high power emission. *Nd–YAG Rod* YAG is an acronym for yttrium aluminum garnet ($Y_3A1_5O_{12}$) and Nd means neodymium (rare earth element) yttrium ions (Y^{3+}). The aluminum garnet ($Y_3A1_5O_{12}$) is doped with a rare earth neodymium (Nd^{3+}) resulting in a crystalline structure. It is a four-level laser with high power, more gain and low threshold pump power.

Resonating Cavity In the Nd-YAG laser, a rod of 5 to 10 cm length and 6 to 9 mm diameter is used. A cylindrical rod is cut from the crystal and is used as an active medium. The ends of the rod are polished and made optically flat and parallel. One end of the rod is coated with silver to get 100% reflecting mirror while the other end as a partial reflecting mirror. The set-up acts as a resonator and is used to amplify the radiation energy after the stimulated emission of radiation. The length of the rod is used to vary the power out.

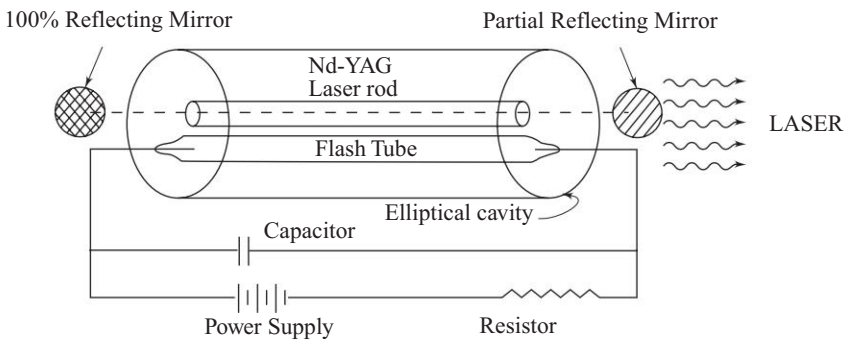


Fig.11.9 *Nd–YAG laser*

Flash Tube *The* principle behind Nd–YAG laser is optical pumping. In this laser, the active medium is Nd-YAG. The population inversion is achieved by a flash light either employing a xenon or krypton flash tube. As a result, Nd^{3+} ions are transported into the excited energy levels. The laser transition from metastable state to ground state takes place with the emission of a laser beam in the infrared region. The schematic representation of Nd–YAG laser is shown in Fig.11.9. The rod and flash tube are kept inside a highly reflecting elliptical cavity. A close coupling between the flash tube and rod is made by placing the flash tube nearer to the centre of the rod. The necessary potential is applied to flash tube and the same is controlled through the power supply. When the flash tube is energised, the rod absorbs energy due to the elliptical cavity arrangements. The energy of the flash tube is produced through the capacitor which is included in the circuit. When the flash tube is on, a large amount of heat is produced

in the elliptical cavity. A very less quantity of energy is required to achieve optical pumping for laser action. Thus, excess heat increases the temperature of the cavity. The temperature inside the cavity is controlled using water or air-cooling arrangements.

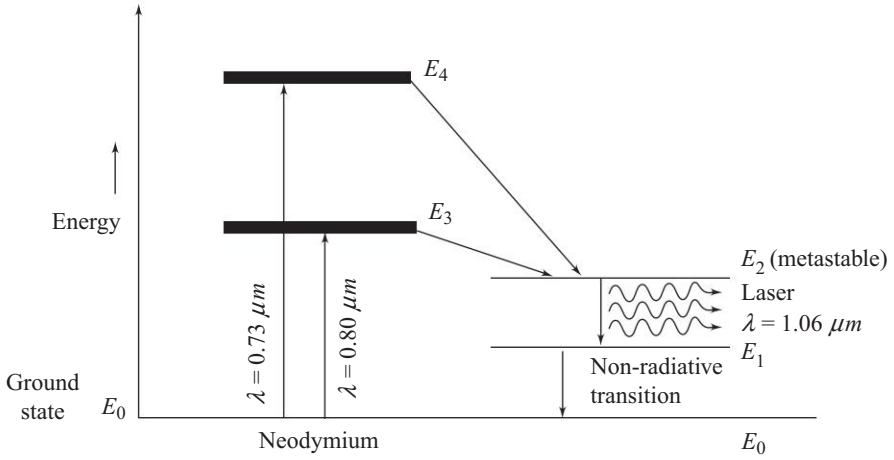


Fig.11.10 *Energy-level diagram*

Working Principle

The energy-level diagram for a Nd-YAG laser is shown in Fig.11.10. When-flash lamp is switched on, neodymium ions acquire energy from the flash light. The neodymium ions are excited to energy levels E_3 and E_4. The transition from the ground state to E_3 and E_4 are due to absorption of energy with wavelengths of respectively 0.73 mm and 0.80 mm. The Nd^{3+} is not stable in the excited state, it makes a non-radiactive transition from E_3 and E_4 states to a metastable state E_2 as shown in Fig.11.10. The metastable state is not a stable state. Therefore, Nd^{3+} ions are in this state until the population inversion is achieved. When the population inversion is achieved between E_2 and E_1, a stimulated emission takes place from the energy levels E_2 to E_1. The emitted energy is amplified between the resonators and then radiates a pulsed laser beam at a wavelength of 1.064 mm in the infrared region. Further, Nd^{3+} ions take a very rapid non-radiactive transition from E_1 to E_0. One can produce a maximum output power of 70 W either in continuous or pulsed mode of operation using the Nd-YAG laser.

Applications

The following are applications of Nd-YAG laser:

- Communication applications such as transmitting signals
- Remote sensing applications
- Medicines for endoscopic applications
- Applications like drilling, welding, micromachining, resistor trimming, scribing, etc.
- Medical surgery, dental surgery, etc.

The McGraw·Hill Companies

11.7.2 Gas Laser

The main advantage of gas lasers is that they can be operated continuously. The gas lasers show exceptionally high monochromaticity and high stability of frequency. The output of the laser can be tuned to a certain available wavelength. Hence, the gas lasers are widely used in industries.

Helium–Neon Laser The first gas laser constructed by Javan in the laboratory is the helium–neon laser as shown in Fig. 11.11. The two important parts of the laser are as follows:

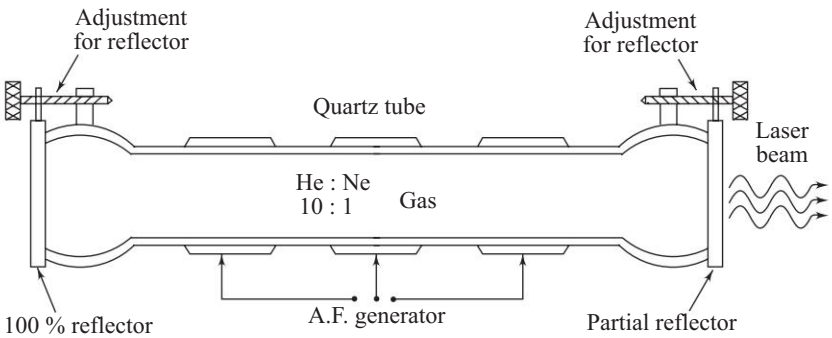


Fig. 11.11 *Helium-Neon laser*

Active Medium The active medium used in this type of laser is a mixture of helium and neon gases. These two gases are mixed under a pressure of 1 mm Hg of helium and 0.1 mm Hg of neon in the ratio of 10:1. The mixture of these gases is filled within the discharge tube for laser action.

Gas Discharge Tube It is made up of a fused quartz tube with a diameter of 11.5 cm and a length of 80.*6* cm. The end faces of the discharge tube are tilted at the *Brewster angle*, known as *Brewster windows*, since the windows are highly transparent for the preferred direction of polarisation. A fully reflecting concave mirror is placed at one end of the discharge tube and a partially reflecting concave mirror is placed at the other end.

Working Principle The population inversion for laser action is achieved by inelastic atom–atom collisions. The collisions of the atoms are made using any one of the following methods given below.

(i) Direct current discharge
(ii) Alternate current discharge
(iii) Electrodeless high frequency discharge, and
(iv) High voltage pulses

Due to the electric discharge in the gas, an energetic electron interact with the ground state helium atom. The impact of the electron results in exchange of some of its energy to the helium atom. As a result, helium atoms are excited to higher levels $1S_0$ and $3S_1$, known as *metastable state*. The lifetime of these levels are relatively low. The collision of the first kind is represented as,

$$He + e_1 \longrightarrow He^* + e_2 \tag{11.40}$$

The McGraw·Hill Companies

These two energy levels are very close to 2s and 3s levels of the neon atom and the collision of the second kind takes place between the He and Ne atoms, and hence, the neon atom goes to the excited state.

$$He^* + Ne \longrightarrow He + Ne^* \tag{11.41}$$

The energy level diagram of He–Ne laser is as shown in Fig. 11.12.

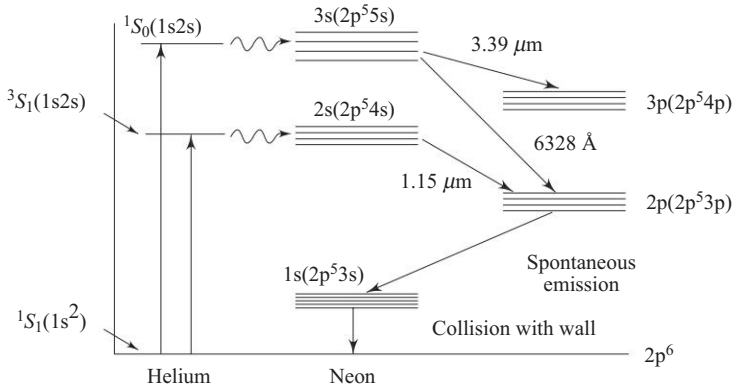


Fig. 11.12 *Helium-Neon laser—energy–diagram*

There are three types of transitions, one from 2s to 2p, the other from 3s to 3p and another from 3s to 2p levels for laser action. These transitions constitute laser beam in the infrared region (11,523 Å , 33,912 Å) and 6328 Å in the visible region. The transition from 3s to 3p at 3.39 μm will have advance effect on the laser emission. This line is suppressed in order to get a maximum power output at 6328 Å.

In Fig. 11.12, the energy levels are represented in terms of Paschen's notation and their corresponding electronic configuration are given in parentheses.

The Neon atom in the terminal level 2p decays very rapidly to the 1s metastable state in 10^{-8} s, much faster than spontaneous rate of decay from 2s to 2p level. Thus, the lower lasing level is relatively kept empty and the population inversion is achieved between 3s and 2p.

11.7.3 Molecular Gas Lasers

The construction of molecular gas lasers is simple and the output of these lasers are continuous. In molecular gas lasers, the laser oscillations are achieved by the transition between the vibrational and rotational levels of the molecules.

Carbon dioxide lasers (CO_2)

It is the first molecular laser and it was developed by C K N Patel.

Principle

In order to understand the working of this laser, one has to recall the rotational and vibrational spectrum of CO_2 molecules. The three atoms can be considered as a ring over a straight line, the outer atom being O with a carbon atom C at the centre. There are three modes of vibrations and in each mode the centre of gravity remains fixed.

(1) As shown in Fig. 11.13a, the carbon atom is fixed in its position and each oxygen atom can vibrate in the opposite direction symmetrical to

The McGraw·Hill Companies

the carbon atom with each other along a straight line and is known as the *symmetric mode of vibration*. The corresponding frequency is called *symmetric stretching frequency*.

(2) As shown in Fig. 11.13b, the oxygen atom and the carbon atom may vibrate at right angles to the line passing through the centre of gravity. This is known as the *bending mode* and the corresponding frequency is *bending frequency*.

(3) As shown in Fig. 11.13c, in the asymmetric mode of operation, the two oxygen atoms may vibrate about the central C atom asymmetrically, and at the same time, the carbon atom also vibrates from its mean position.The corresponding frequency is called as *asymmetric stretching fequency.*

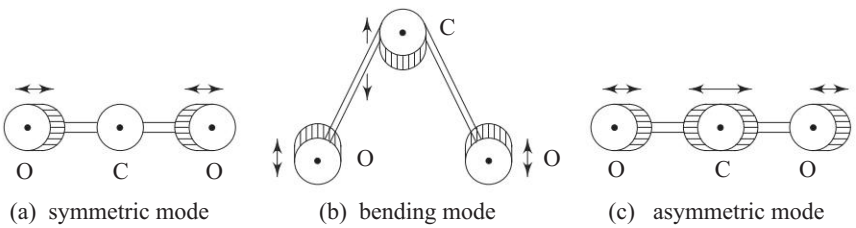


Fig. 11.13 *CO_2 molecules—different modes of vibrations*

In addition to these three vibrational modes, the molecule can also rotate and therefore, quantised rotational energy levels are also possible. A series of rotational levels is associated with each vibrational level and are denoted by *J* values.

Construction The special feature of the CO_2 laser is the dependence of radiation power, i.e., output power on the diameter of the tube. The output power can be raised by increasing the tube diameter. In a powerful CO_2 laser, the length of the discharge tube is in several metres and its diameter will be several centimetres as shown in Fig. 11.14.

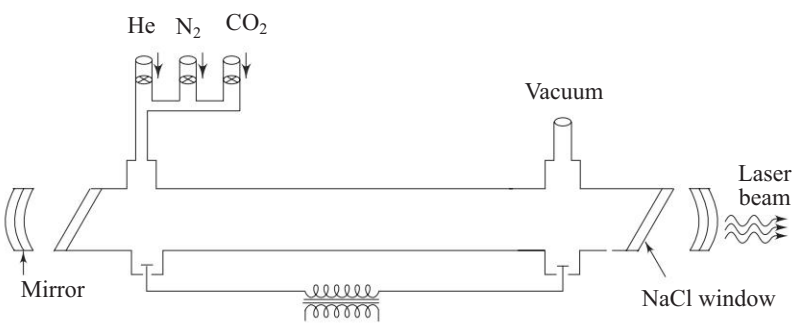


Fig. 11.14 *CO_2 laser*

The laser is powered with an ac supply of frequency 50 cycles or dc supply. In order to get a high power output, a metallic mirror of gold is employed for proper reflection. The lasers are either water-cooled or air-cooled. The efficiency of the CO_2 laser is about 30%. The gas mixture can be pumped either longitudinally or transversely into the gas discharge tube.

The McGraw·Hill Companies

Working Principle The CO_2 laser generally uses two additional gases, N_2 and He. The nitrogen plays a similar role as that of the He in case of He–Ne laser. The N_2 molecules go into an excited state by collision of the first kind with the electrons.

$$N_2 + e_1 \longrightarrow N_2^* + e_2 \tag{11.42}$$

The excited N_2 atom undergoes a collision of the second type and makes the CO_2 molecules to be excited.

$$N_2^* + CO_2 \longrightarrow CO_2^* + N_2 \tag{11.43}$$

The frequency of the CO_2 laser can be pictured in an energy level diagram as shown in Fig. 11.15.

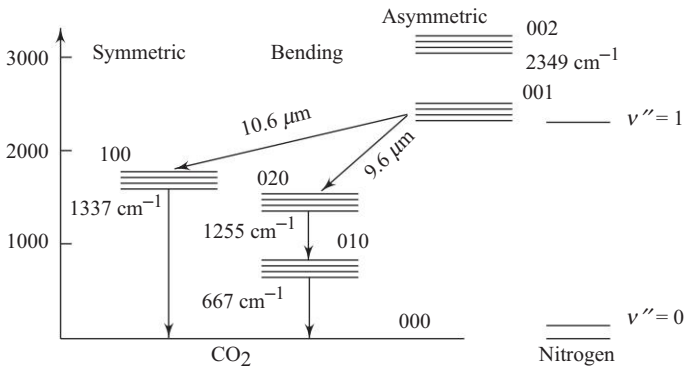


Fig. 11.15

Since the 0 0 1 energy level of CO_2 is very close to the excited level of N_2 atom, the population of 0 0 1 level of CO_2 increases rapidly than the other lower energy levels 1 0 0 or 0 1 0. Thus, the population inversion between 0 01 level and the lower levels 0 2 0 and 1 0 0 is achieved resulting in output radiations of 9.6 and 10.6 μm.

Due to various factors, the most powerful transition in CO_2 laser at normal operating temperature occurs at 10.6 μm. The operating temperature plays an important role in determining the output power of the laser. The contamination of carbon monoxide and oxygen will also have some effect on the laser action. The unused gases can be pumped out and fresh CO_2 must be pumped in. The temperature can be reduced by restricting the tube diameter and also the addition of helium to the mixture of N_2 and He. The helium serves not only to improve the conductivity of heat to the walls of the tube, but also in decreasing the population in the lower levels. The power output coming from this laser is 10 kW.

11.7.4 Semiconductor Laser (Semiconductor Diode Laser)

The most compact of all laser is the semiconductor diode laser, also called the *Injection laser*. In its simplest form, the diode laser consists of a *p–n* junction doped in a single crystal of a suitable semiconductor such as gallium- arsenide.

Principle When a forward bias is applied to the diode, the holes are injected into the *p*-side of the junction and electrons are injected into the *n*-side. The recombination of

The McGraw·Hill Companies

holes and electrons within the junction region results in recombination radiation. If the junction current density is large enough, a population inversion can be obtained between the electron levels and hole levels. Stimulated emission can be obtained for laser action when the optical gain exceeds the loss in the junction layer. In a diode laser, this layer is very thin, typically of the order of few microns and the end faces of the crystal are made partially reflecting to form an optical resonator.

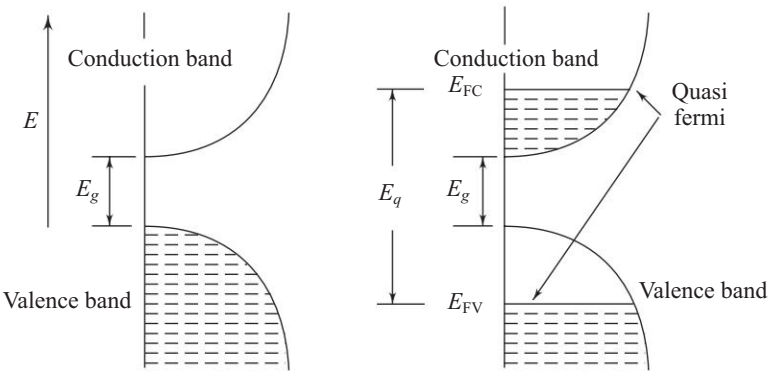
To produce laser action, the following conditions should be satisfied.

(i) Population inversion
(ii) Stimulated emission, and
(iii) Cavity resonator

The laser transition is possible only in direct band gap semiconductors. Therefore, the direct band gap semiconducting materials like Ga–As are used for laser action. Since Si and Ge are said to be indirect band gap semiconductors, they are not used for laser action.

At absolute zero, the conduction band of an intrinsic semiconductor has no electron. However, heavily doped semiconducting materials consist of filled electron states in the conduction band (Fig. 11.16).

If the heavily doped material is irradiated with a radiation having energy greater than E_g but less than the separation of the quasi fermi level ($E_{FC} - E_{FV}$), the incident photon can induce a downward transition of an electron. This electron combines with a hole in the valence band and hence the recombination energy is produced in the form of light. This photon, in turn, may induce another electron in the conduction band and thereby stimulate the emission of the another photon.



(a) In equilibrium (b) With high carrier injection

Fig. 11.16 *Filled electron states of an intrinsic semiconductor at 0 K: (a) in equilibrium and (b) with high carrier injection*

The population inversion in a *p–n* junction is achieved by heavily doping the *p*-type and *n*-type materials, so that the fermi level lies within the conduction band in the *n*-type material and it lies within the valence band in the *p*-type material. Figure 11.17 represents the energy band diagram of a heavily doped *p–n* junction device at thermal quilibrium.

If the junction is forward biased with an applied voltage nearly equal to the band gap voltage, direct conduction takes place. This results in high injection current density leading to the generation of an active region near the depletion layer where the condition

The McGraw·Hill Companies

of population inversion is achieved. At this juncture, if a radiation having frequency ν, where $E_g/h < (\nu) < (E_{FC} - E_{FV})/h$ is made to incident on the *p–n* junction device, the stimulated emission is produced and hence the radiation of this frequency, which is confined to the active region will be amplified.

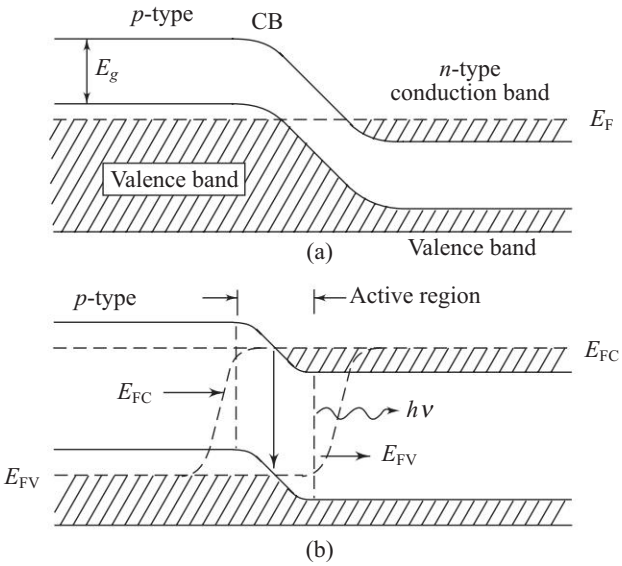


Fig. 11.17 *Energy band diagram of (a) a heavily doped p–n junction, (b) heavily doped p–n junction under heavy forward biasing*

Another requirement of the laser action is the cavity resonator. In a *p–n* junction device, the end faces of the junction diode are polished so that they act as an optical cavity. Other sides of the *p–n* junction device is made as rough surfaces so as to prevent the leakage of light from them.

Ga–As Laser

It is a *p–n* junction diode with the *p*-type and *n*-type regions heavily doped. Under large applied forward bias, electrons and holes are injected into and across the transition region in considerable concentration. As a result, the region around the junction contains a large number of electrons within the conduction band and a large number of holes within the valence band. When the population density is high enough, a condition of population inversion is achieved and recombination may be stimulated resulting in laser action. If the emission is not stimulated, the device is called a light emitting diode. To convert an LED into a laser diode, a high current is required to achieve the population inversion and mirrors are used to cause a feedback. In case of semiconductor laser, there is no need of external mirrors. In a germanium or silicon semiconductor, due to such a recombination, only heat is generated.

The Ga–As laser was constructed by Hall. A typical Ga–As semiconductor laser is as shown in Fig. 11.18.

In case of gallium–arsenide, we get a light radiation in the infrared region. Thus, a Ga–As laser convert electricity into light. The efficiency can go up to 100% when the temperature is reduced to 100 K. The operating current is supplied from a pulse

The McGraw·Hill Companies

generator up to 5–20 μs. The energy separation between the conduction band and the valence band is 1.4 eV and hence, the wavelength of light emitted is 8874 Å at room temperature.

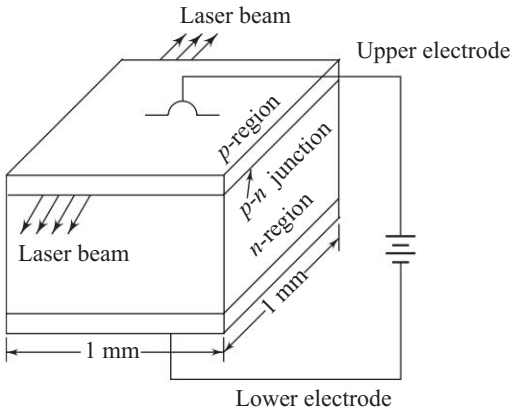


Fig. 11.18 *Ga–As laser*

Advantages

1. The efficiency is more than 10% and it can be increased by decreasing the temperature alone.
2. It can have a continuous wave output or pulsed output.
3. The modulation of the output is possible.
4. Tuning the output is easily possible by applying a magnetic field or mechanical pressure.
5. It is highly economical, and further, the arrangement is compact.

Disadvantage The spatial and temporal coherence are very poor.

11.7.5 Homojunction and Heterojunction Injection Laser

The semiconductor diode or injection laser is of two types, namely, homojunction and heterojunction. Let us discuss the above two types in brief.

(i) Homojunction A *p–n* junction made up of the same material with two regions of different conductivities, *n*-type and *p*-type, is known as a *homojunction*. Consider a *p–n* junction formed by *p*-type and *n*-type Si crystals, then it is said to be a homojunction. The junction is formed at the region where the conductivity changes from one type to another type. A homojunction *p–n* junction may have abrupt or graded junction depending upon the method of preparation. If a *p–n* junction is formed by diffusing a suitable pentavalent impurity from the upper portion of a Si crystal and diffusing another suitable trivalent impurity from the bottom portion of the Si crystal, a *p–n* junction of the Si crystal is formed. This will produce a graded junction. An abrupt Si is coated on a substrate followed by the coating of *n*-type material using a suitable thin film deposition technique.

The homojunction device lacks in carrier containmnet and hence, it is an inefficient light source. The devices like LED, GaAs homojunction laser, etc., are fabricated

with homojunction. The basic structure of a homojunction injection laser is shown in Fig. 11.19.

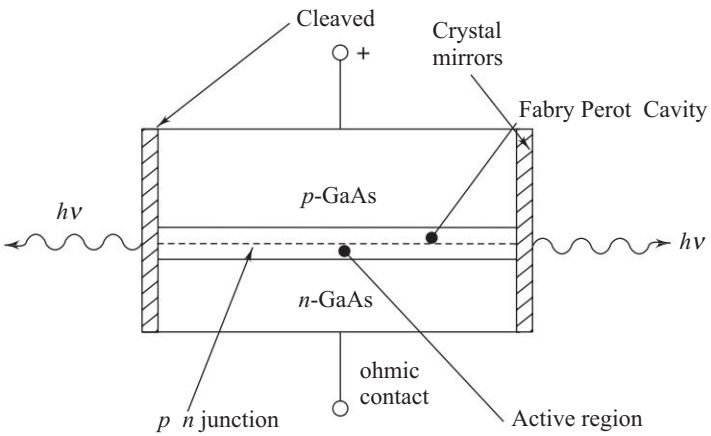


Fig. 11.19 *GaAs laser—Schematic representation*

The cleaved ends of the crystal acts as reflecting mirrors so as to amplify the light sources. This homojunction GaAs laser device had a high threshold current density (greater than 10^{-4} A cm^{-2}) due to lack of carrier containment and hence, it is an inefficient light source. Therefore, the homojunction lasers are operated in a pulsed mode in order to minimise junction temperature and hence, to avoid damage.

(ii) Heterojunction A heterojunction is an interface between two adjoining single crystal semiconductors with different band gap energies. There are two types of heterojunctions, namely, *isotype* and *anisotype heterojunctions*. If the junction is made up of two different materials having same type of conductivity (*n–n* or *p–p*), then it is said to be *isotype heterojunction*. If two different materials having different types of conductivity are joined together, then it is known as *anisotype heterojunction*. An isotype heterojunction is used to confine the minority carriers to a small active region. It reduces the carrier diffusion length, and hence the volume in which the radiative recombination takes place. The isotype heterojunction is mostly used for the preparation of injection lasers and highly radiative LEDs.

The anisotype heterojunctions with sufficiently large band gap differences improve the injection efficiency of the charge carriers (holes and electrons). Both types of heterojunctions provide dielectric steps due to different refractive indices at either side of the junction. This property is used to confine the radiation in the active region. A typical double heterojunction laser is shown in Figs 11.20. a and b represents, respectively, the layer structure shown with applied forward bias and the energy band diagram of two heterostructures: (i) *p–p* at the left, and (ii) *p–n* at the right. The variation of refractive index and electric field distribution with distance is shown in Fig. 11.20.

The requirement of a good quality heterojunction are (i) the lattice constant of the *p*-type and *n*-type material should be nearly equal (i.e. lattice matching), (ii) the electron affinities of these two materials should be compatible, and (iii) the thermal expansion coefficient should be close. If there is lattice mismatching and the thermal expansion coefficient is not very close, it leads to interfacial dislocation at the heterojunction

The McGraw·Hill Companies

interface. The dislocation produces energy discontinuity in the form of a spike in one or both sides of the energy bands.

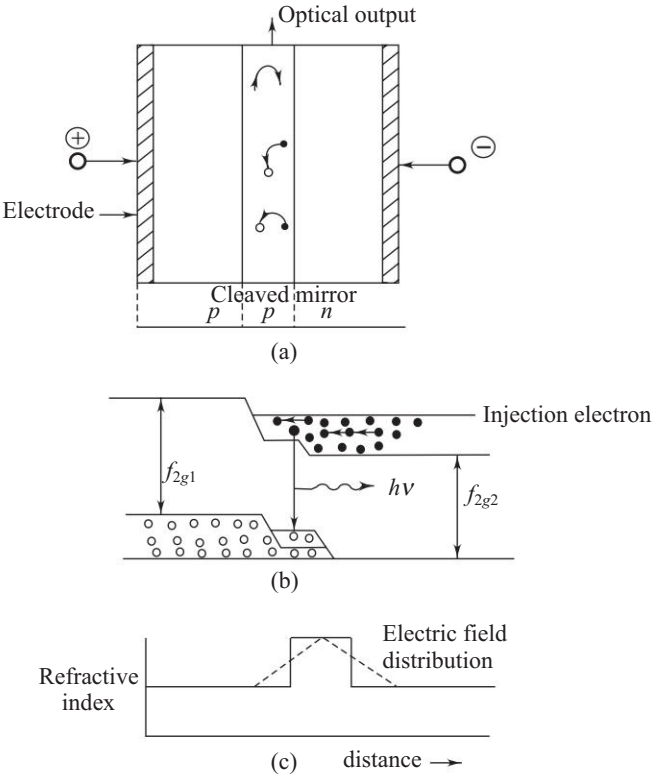


Fig. 11.20 *A double heterostructure laser*

11.8 DETERMINATION OF WAVELENGTH OF LASER USING GRATING

Aim To determine the wavelength of the laser beam using grating.

Apparatus Required A He–Ne laser source; grating; spectrometer; sodium vapour lamp; spirit level; etc.

Formulae

(i) The number of lines per metre drawn on the grating is given by

$$N = \frac{\sin\theta}{m\lambda} \text{ lines m}^{-1} \tag{11.44}$$

where m is the order of spectrum observed and λ the wavelength of sodium vapour lamp (λ = 589.3 nm).

(ii) The wavelength of the given laser source,

$$\lambda = \frac{\sin\theta}{Nm} \tag{11.45}$$

The McGraw·Hill Companies

where θ is the angle of diffraction, N the number of lines drawn on the grating per metre, m the order of spectrum and λ the wavelength.

Experimental Procedure

(i) To determine the number of lines drawn per metre of the given grating The experimental arrangement is shown in Fig. 11.21. One can use the standard procedure to determine the number of lines drawn on the grating per metre. The observed experimental results are tabulated in Table 11.1.

Table 11.1 *Determination number of lines per metre of the grating*

Order	*Diffractig Ray Readins*				*Difference*			$N = \frac{\sin\theta}{m\lambda}$
	Right		*Let*		2θ		*Mean*	
	V_A	V_B	V_A	V_B	V_A	V_B	θ	

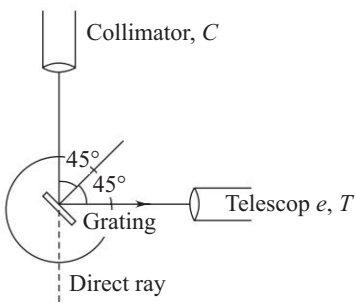


Fig. 11.21 *Adjustment for normal incidence*

(ii) To determine the wavelength of the laser source The experimental arrangement used for the determination of the wavelength of laser is shown in Fig. 11.22. The given He–Ne laser source is switched on and the laser beam is passed through the grating. The diffracted rays are collected on a screen. In the diffraction pattern, the circular spot at the centre is produced due to the direct beam. The other diffracted beams are produced on either side of the direct beams. The diffracted pattern produced by the laser beam is shown in Fig. 11.23.

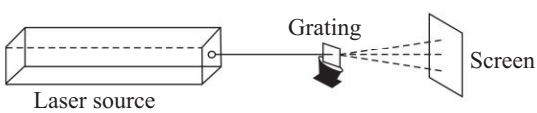


Fig. 11.22 *Determination of wavelength of the laser*

The distance between the first-order diffracted rays (left and right side diffracted beams) is measured and the readings are tabulated. Similarly, the distance between the second order diffracted rays is also measured and the readings are tabulated in

The McGraw·Hill Companies

Table 11.2. The distance between the screen and the grating, l, is also measured. The angle of diffraction is determined using the relation,

$$\theta = \frac{d}{2l} \tag{11.46}$$

where l is the distance between the screen and the grating and d the distance between the dffracted rays.

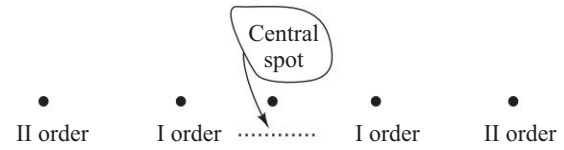


Fig. 11.23 *Diffraction patterns of laser beam produced by grating*

The wavelength of laser beam is determined using therelation,

$$\lambda = \frac{\sin\theta}{Nm} \tag{11.47}$$

where θ is the angle of diffraction, m the order of diffraction and N the number of lines drawn on the grating.

Table 11.2 *The distance between screen and grating, l = … m*

Trial	*Order of diffracted*	*Distance between diffraction beams, d*	*Angle of diffraction,* $\theta = \frac{d}{2l}$
1	I II		
2	I II		

Results

(i) The number of lines drawn on the grating = lines m^{-1}
(ii) Wavelength of the given laser beam = m

11.9 PARTICLE SIZE DETERMINATION BY LASER

Aim To determine the particle size of lycopodium powder using He–Ne laser.

Apparatus Required A He–Ne laser, lycopodium powder, a glass plate; a long focal length lens, a screen, and a scale.

Formulae The particle size is given by $2d = n\lambda\frac{D}{x_m}$ (11.48)

where $2d$ is the grain size, n the order of diffraction, λ the wavelength of the laser beam used (for He–Ne laser, $\lambda = 0.6328$ nm), D the distance between the source and the screen and x_m the distance between the central bright spot and the m^{th} fringe.

Experimental Procedure The experimental set-up used for particle size determination using laser is shown in Fig. 11.24. Sprinkle lycopodium powder having an average grain size is in the order of few micrometers on a glass plate. Insert this glass plate in between a laser beam and a screen is placed nearly tens of centimetres (nearly 20–50 cm) away from the source. A diffraction pattern as shown in Fig. 11.25 is obtained.

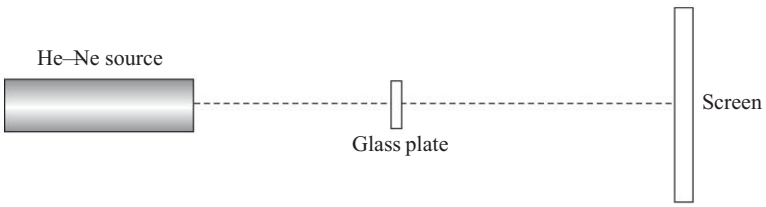


Fig. 11.24 *Experimental set up for Fraunhofer diffraction*

The distance between the central spot and the m^{th} minima positions is measured in the screen directly using a scale. The distance between the source and the screen (D) is also measured. Different sets of readings are taken by varying the m^{th} minima positions (say 1, 2, 3, 4, 5, …). The experiment is also repeated by varying the distance between the source and the screen. The observed readings are tabulate in Table 11.3.

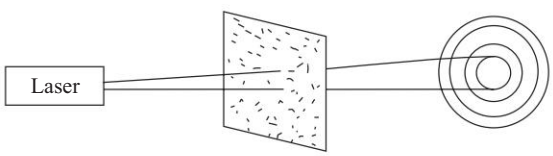


Fig. 11.25 *Diffraction pattern*

Table 11.3 *Particle size determinaion usig laser*

S.No.	*Distnce between the source and the screen (D)m*	*Order of diffraction n*	*Distnce between the central spot and m^{th} fringe*	*Particle size 2d (m)*
1		1		
2		2		
3		3		
4		4		
.		.		
.		.		
10		10		
			Mean value	

Results

The grain size of the lycopodium powder is _____ m.

11.10 APPLICATIONS OF LASERS

Laser beams have found wide application due to their high intensity, high monochromaticity, high coherence and high directionality.

The McGraw·Hill Companies

11.10.1 Three Dimensional Profiling

The process of storing and retrieving the three-dimensional image of an object by using laser scanner and computer software is known as three dimensional (3D) profiling. The principle involved in storing the three dimensional images is same as that of holography.

When a highly coherent, monochromatic and directional light such as a laser beam is incident on an object, it produces secondary wavelets. Each point on the object acts as a source for secondary wavelets. These waves spread over in all the directions and form an interference pattern. The interference pattern characterises the object. The observed interference pattern is recorded using a suitable technique. The recorded image is again illuminated by using another identical coherent laser source. Thus, the three dimensional image of the object is reproduced due to the interference of the two lights which incorporate the variation in amplitude and intensity.

The block diagram of a simple 3D profiling using laser beam is shown in Fig. 11.26.

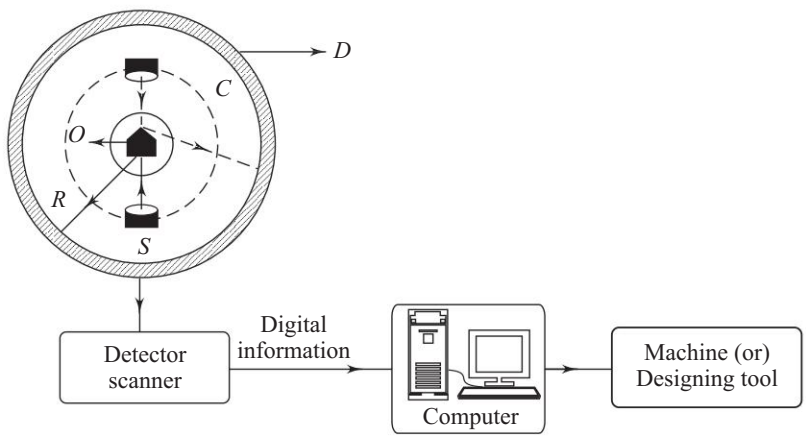


Fig. 11.26 *Laser—3D profiling*

The following are the important parts of the 3D profiling system:

(i) Laser source
(ii) Detector array and scanner
(iii) Computer

A high-intensity and monochromatic laser beam is generated using a laser source (*S*). The laser source is movable on circular scale (*C*) and it can be positioned at any desired angle over 360°. The laser beam from the source is made to fall on the object (*O*) which is kept fixed at the centre of the orbit of the laser source as shown in Fig. 11.26. Thus, the reflected light (*R*) from the surface of the object spreads in all directions.

The reflected light from the source is collected by using light-sensing diodes. These diodes are known as detectors. These diodes produce an equivalent amount of current depending on the intensity of the light incident on it. These diodes are arranged in the form of an array known as detector (*D*) to receive the reflected light from the source.

In order to record the reflected light from the source in all directions, a detector scanner is fitted in such a way that it moves along the three planes. The received signals

by the scanner are converted into digital form using an analog to digital convertor unit which is attached with the scanner. The received signal carries both amplitude and phase variation which reflects the nature of the object. Thus, the optical signal is converted into electrical signal and is stored in the microcomputer.

The information stored in the computer is in the form of matrix. Using the signal processor software available with the computer, one can reconstruct the 3D image of the object from the stored data. The resolution of the 3D images can be increased by using the software controlled by the computer.

11.10.2 Computer Peripherals

The application of a laser beam in computer peripherals has found greater revolution in recent years. The optical disks have very high storage capacity than conventional magnetic disks. Similarly, the data transferring from one computer to another computer through optical wave guides play an immense role. The application of laser beam in storing data in optical disks is discussed below, while that in the memory transfer using optical fibres is discussed in the previous chapter.

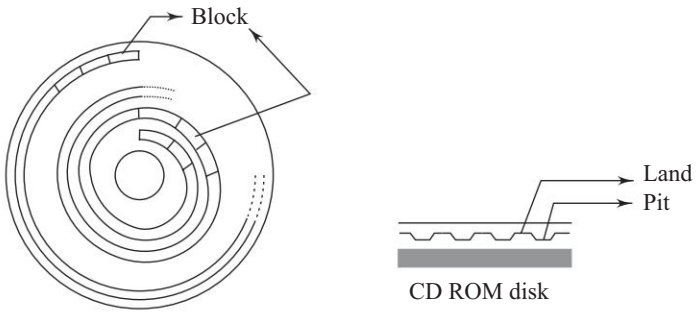


Fig. 11.27 *CD ROM*

An optical disk has very high storage capacity. For example, the storage capacity of a 5.25 inch disk is 550 MB. An optical disk is made up of a resin such as polycarbonate. One side of the disk is coated with highly reflective materials like aluminium. When a high intensity light beam such as laser beam is incident on it, it will produce some changes on the coating: a 0.1-mm wide tiny pit is formed along the trace (to represent 1). The non-pitted areas between the pits in the circular disk is represented as lands as shown in Fig. 11.27. The information in CD ROM disk is stored in a long spiral track. The track is divided into blocks of the same size as shown in Fig. 11.27. The information is read/write on the optical disk using laser beam, which replaces the read/write arm in the conventional magnetic disk memories. A high-intensity laser beam is used to store the data, while a low intensity laser beam is used to read the data.

The experimental set-up used for the read-out mechanism for CD ROM is shown in Fig. 11.28. A low-intensity laser beam from the source S is made to incident on the CD ROM disk through an adjustable mirror M and lens L_2 after passing through the beam splitter B. The intensity of the laser beam used for reading the data is about 4 mW. The laser beam scans the tracks and receives the reflected lights from the pits and lands. The mirror M and lens L_2 act as read arm and can move front and back to

The McGraw·Hill Companies

read the data on different tracks. The incident laser beam on pits spreads, while on lands gets reflected. A photodiode sensor *D* receives the reflected light after passing through the lens L_3. The photodiode receives a little light from the pits and a larger light from the lands. Thus, the change in the reflected light is sensed and converted into electrical signals for data reading. In this system, the speed of the optical disk is adjusted such that the track passes under the read/write head at a costant velocity.

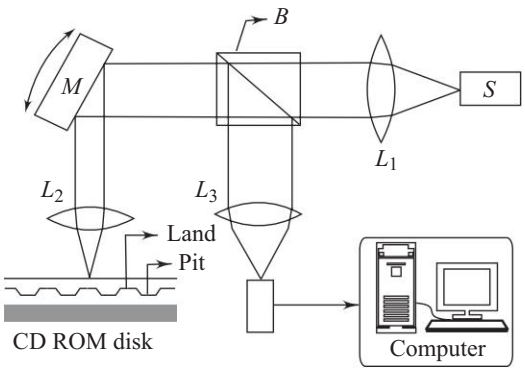


Fig. 11.28 *Optical read-out mechanism for CD ROM*

There are three types of optical disks, namely, compact disk which read only memory (CD ROM), write once read many (WORM) and erasable optical disk (EOD). The main advantage of CD ROM is its high storing capacity, mass copy of information stored removable from the computer, etc. CD ROM disks are used for archival storages. The main disadvantages are longer access times when compared with magnetic disks and it cannot be updated, as it is a read only memory. However, in recent years, read/write CDs are also available in the market. The RW CDs overcome all the drawbacks of the CD ROM and hence the optical disk storage plays a vital role in computer memory storages.

11.10.3 Material Processing

A high-intensity laser beam is used to study the surface defects in materials such as ICs, aircrafts, automobile tyres, etc. The laser beam is reflected from the surface of the material under investigation and also directly from the source as a reference. The information about the material is obtained simply by forming the interference between the two laser beams. There are two methods to study the surface defects, namely, diffraction and opto-acoustic image processing. In both the techniques, a laser beam is used to study th surface defect.

The block diagram of the experimental set-up used in diffraction method is shown in Fig. 11.29. A high intensity laser beam from the source *S* is made to fall on the converging lens L_1. The lens L_1focuses the laser beam on the object *O*. The optical diffraction pattern of the image is focused on the photographic plate (*P*) or photodiode (*D*). The photodiode senses the light and converts the light energy into electrical signals. Using the signal analyser along with the necessary software, the image of the defect can be obtained. Thus, using the high intensity laser beam, the surface defects on the materials can be studied effectively.

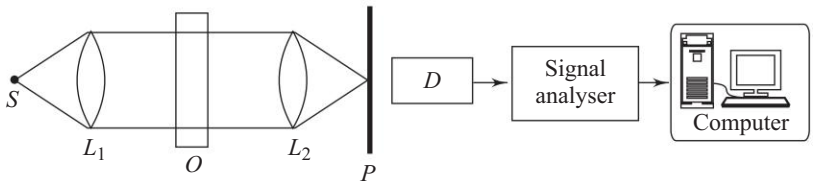


Fig. 11.29 *Surface defect detection*

11.10.4 Drilling

Drilling is the first production application of laser light. The first commercial application developed in drilling is the drilling of gemstones for wire dies, watches and jewels. A high-energy laser beam is used to obtain small holes in materials at high speed with good quality. When a high-intensity laser beam is used to drill the material, a large number of physical processes, ablation mechanisms, etc., take places depending on materials and irradiation condition.

The schematic representation of a typical laser drilling process is shown in Fig. 11.30. The high-energy laser beam radiation is made incident on the surface of the material. When the laser radiation hits the material surface, the material is removed by means of two processes. In the first process, when the radiation is incident on the material surface, the temperature of the material is raised and it reaches the vaporisation temperature. When the material is vaporised, it creates a cavity in the material. When the process continues, the vaporisation increases and hence, the pressure of vapour also increases. When the vapour pressure reaches the required pressure level, say 100 Kpa, the second process namely *melt expulsion,* takes place. The high-pressure vapour escapes from the area after producing a high recoil pressure on the molten materials, which are present at the bottom of the hole. After acquiring the required pressure, the molten materials overcome the surface tension, and hence, they are forced out to the surface of the hole. These molten materials, i.e., the liquid metal, is ejected vertically out of the hole through its sides as shown in Fig. 11.30. The energy required to remove material by melt expulsion is very less when compared with vaporisation. Thus, it is clear that the recoil pressure and vaporisation temperature are more essential parameters in laser drilling.

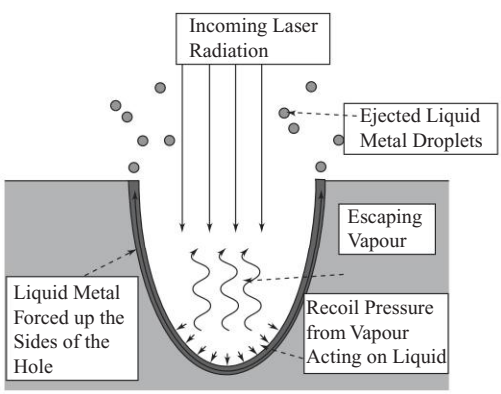


Fig. 11.30 *Laser drilling*

The McGraw·Hill Companies

Based on the applications, the laser drilling parameters are selected using the relationship between the drilling parameters and hole parameters. For any industrial applications, the relationships between the above parameters are given in Table 11.4.

Table- 11.4 *Drilling parameter, hole parameter and application requirements*

S. No.	*Drilling parameter*	*Hole parameter*	*Application requirements*
1	Pulse energy	Exit hole diameter	Aero dynamics
2	Beam diameter	Entrance hole diameter	Flow characteristics
3	Pulse duration	Hole profile	Cooling effect
4	Wavelength	Surface roughness	Optical effect
5	Focusing	Hole depth	Materials property
6	Gas instability	Pitch	Tolerances
		Tolerances	

There are different drilling methods for any particular drilling applications. The available drilling methods are single pulse, percussion, parallel percussion, trepanning, helical trepanning, imaged and angled. Lasers such as Nd–YAG and CO_2 are used for industrial drilling purposes due to their high energy output as well as continuous operations.

Advantages

The following are the advantages of laser drilling:

- It is a non-contact processing method.
- It is used to drill micro holes and small holes with large aspect ratios.
- It generates very low heat in the material during drilling.
- It is possible to drill at different angles.
- It is highly flexible to drill over a wide range of materials.
- Its accuracy and consistency are very high.

Applications

The following are the application of laser drilling in industries:

- It is used in the electronic industry to make the printed circuit boards.
- In aerospace, it is used to make cooling holes in aircraft engine blades and combustors.
- It is used to drill the inject nozzles, fuel injector nozzles, aerosol nozzles, etc.
- It is used to drill diamonds to remove imperfections.
- It is used to drill rocks.

11.10.5 Welding

Laser beam welding is a thermal technique, which is used to join two or more metal parts. The joining of metals is established by solidification of metal at a common

The McGraw·Hill Companies

melting point. The applications of laser welding are ever increasing due to its deep weld penetration and minimizing heat inputs. This method also facilitates to automate the process techniques.

Consider that the given two metal plates are to be welded using laser beam welding. The two metal plates are held in contact at their edges as shown in Fig. 11.31. The focal spot for the welding on the surface of metal plate edges are identified. The high intensity laser beam (Nd–YAG or CO_2 or diode laser) is focused on the spot employing the optical arrangements. At the surface of the metal plate, the highly concentrated laser beam, i.e., high energy is converted into thermal energy and thus it heats the spot in the metal plates. Therefore, the surface of the metal plate starts melting and then progresses through it by surface conductance. Thus, the metal plates fuse together at the points of contacts. The energy of the beam required to melt the metal plates are kept well below the vaporisation temperature of the workpiece metal. When the intensity of the laser beam is increased, the workpiece metal undergoes several phase changes, namely, solid to liquid and then again to vaporisation. The laser welding station requires few basic components, namely, good laser source, beam guiding, workpiece and manipulating and workstations.

A simple experimental set-up used for laser beam welding is shown in Fig. 11.31. The laser light output from the laser source is passed through the optical instruments which helps to control the energy, focus the laser beam as well as for automation. The optical assembly helps to get a fine laser beam. In addition, at the laser focusing point, a shield gas is passed through the shielding gas jet. The shielding gas is used to remove the molten material to favour vaporisation. Further, it provides the necessary cooling effect at the spot which protects the optical arrangements against environmental effects such as fumes. It also helps to increase the adsorption of energy by the sample. The manipulating and work-station on the laser welding systems helps to provide a relative motion between the laser beams and the workpiece, either in two or three dimensions at the required speed and accuracy.

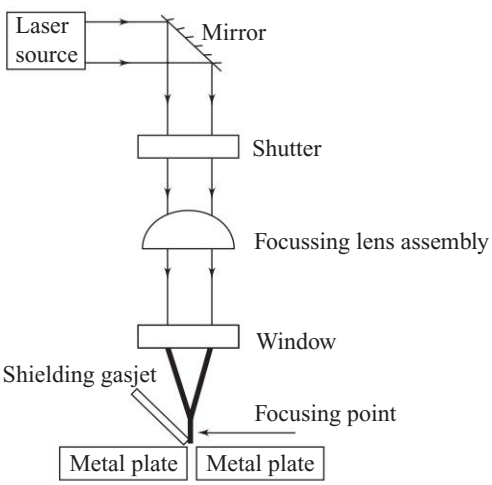


Fig. 11.31 *Laser beam welding*

The McGraw·Hill Companies

Types of Laser Beam Welding

Even though quite a variety of laser beam weldings are available, the most common types used for industrial applications are Nd–YAG, carbon dioxide lasers and diode lasers. The wavelength of laser light produced by Nd–YAG and CO_2 laser is respectively 1.06 μ m and 10.6 μ *m*.

The CO_2 laser source requires beam guiding systems to propagate the laser beam from mirror to mirror. On the other hand, flexible glass fillers can be used for the Nd–YAG laser source. The laser beam welding is used for materials whose thickness is less than 0.80 inches.

Advantages

The following are the advantages of using laser drilling in industries:

- It is a non-contact processing method.
- It has the ability to weld smaller and thinner components.
- It gives a very low heat input into the materials.
- It provides an excellent metallurgical quality on the established weld.
- It has the flexibility to drill over a wide range of materials.
- It is used for doing deep and narrow welds.
- It has high accuracy and consistency.
- It is used to drill submicron and small holes with huge aspect ratios.
- It is used to drill at different angles.
- It can be carried to difficult terrains.
- It creates minimal heat affected zones during welding.

The main disadvantage of laser beam welding is its high cost.

Applications

The following are the applications of laser welding in industries:

- It is used to weld complex-shaped contours in a controlled manner.
- In automobile industries, it is mainly used for car body constructions, space frames, etc.
- Further, it finds application in areas like aerospace, defence, electronics, petro-chemicals, medicine, etc.

11.10.6 Laser heat treatment

Laser Heat Treatment

Laser heat treatment is initially validated for surface hardening. However in recent years, it is also used to change the metallurgical and mechanical properties of materials. Laser heat treatment is used for many practical applications like hardness, increase in strength, increases in fatigue life and surface hardening etc.

A simple schematic representations used for surface hardening using laser source is Fig. 2.18. It consists of laser source, beam, analysis & guidance, gas shield cnc table and work piece.

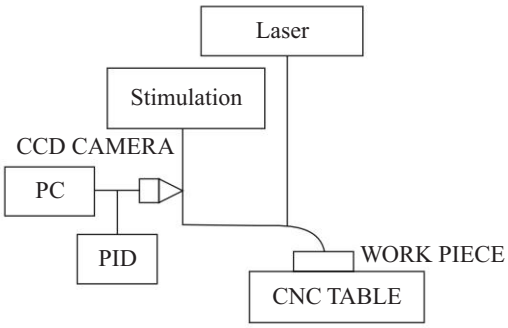


Fig. 11.32 *Experimental arrangement for laser heat treatment*

The principle behind the hardening of materials like iron and steel is the temperature dependence of structural changes. For example, there are three different structures of iron namely body centered cubic (bcc) structure (α-iron) below 1085 K, a face centre cubic (fee) structure between 1085 K to 1667 K and body centered cubic (bcc) structure (γ-iron) beyond 1667 K as a function of temperature. When the high intense of electromagnetic radiation of laser beam incident on a surface of a metal, it transfers heat to atoms in the surface of metal. When the laser beam incident on the metal, the photons from the laser beam interact with the free electrons and the materials and thereby raise their energy states to the conduction band. This effect is known as inverse Breamsstrahliing effect. Thus, the incident laser beam hardens the surface of the metal. The important parameters to be considered for the laser treatment are laser power, beam diameter, absorptivity of the surface and the transverse speed of laser. The laser beam generated bv the source is focused to fall on the work piece which is kept on CNC table. The whole process is controlled by the computer and the process is monitored by CCD camera. The PUD controller is used to fix the temperature of the laser operation to be performed on the work piece.

Advantages

The following are the advantages of the laser heat treatment in surface hardening.

It is a computer controlled process method.
It is very simple to control the reaction depth.
It does not require any post process. « It is free from contamination.
It is also used to clad the nanomaterials.

11.10.7 Laser Cutting

Laser cutting is carried out using the laser machining process. The principle behind laser machining is a thermal mechanical process. During the thermal mechanical process, the material is removed either by phase changes like melting and vapourization or hydrothermal motions.

Working The schematic diagram for laser cutting assisted by a gas jet is shown in Fig, 2.19. It consist of nozzle with laser beam focusing system and gas inlet to assist the laser cutting. The high intense laser beam is focused through the lens L via the nozzle. In order to provide the necessary mechanical force to eject the melt from the

The McGraw·Hill Companies

cut nozzle, gas is circulated through the nozzle. The passed gas is also used to cool the cut zone by forced convection process. The nozzle of the laser cutting kit is brought to the top surface of the workpiece. The assisting gas used for laser welding develops an additional exothrmic energy through chemical reactions. The chemical reaction takes place between assist gas and the molten material. The additional energy is more useful to cut thick piece. On the other hand, one can also use inert gas in order to reduce the oxidation during laser machining. The important points to be considered for laser cutting are nozzle design, jet alignment, and effect of pressure and gas purity.

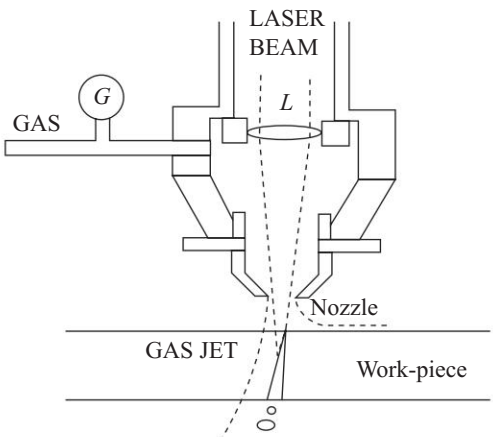


Fig. 11.33 *Laser cutting*

Advantages

The following are the advantages of gas cutting

- It is used to cut materials of any thickness with high precision
- It is very fast and accurate.
- It is very simple and cost effective.

Disadvantages

- The efficiency and the overall cutting quality of laser welding strongly depend on the interaction of gas with the workpiece.

11.10.8 Medicine

The following are the some of the applications of laser in medicine.

(i) It is used to perform microsurgery and bloodless operations, to cure cancer and skin tumours in human beings and animals.
(ii) It is used for the treatment of a detached retina.
(iii) With the help of optical fibers, the argon ion laser is used to control gastric haemorrage. The block diagram of the experimental set-up namely, laser endoscopy used for this purpose is shown in Fig. 11.32.

The McGraw·Hill Companies

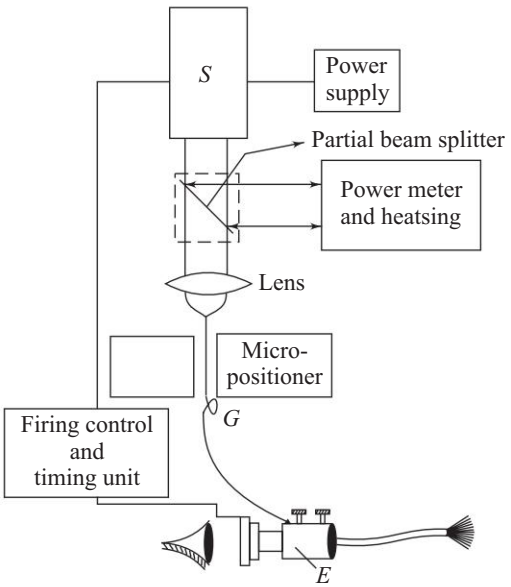


Fig. 11.34 *Laser endoscopy*

The high-energy (13 *W*) argon ion laser (*S*) is used as an optical source and an endoscope (*E*) is used as the delivery unit. The argon ion laser output is delivered into the required spot of the gastric bleeding using fiber optic endoscopes which are connected through the lens system using encapsulated quartz fibre guide (*G*), as shown in Fig. 11.32. The laser beam can be moved into any direction and position with the help of the optical flexible fibre guide (*G*).

11.10.9 Holography

Holography was invented by Dennis Gabour in 1948 with an idea to improve the resolution of the images obtained from an electron microscope. The method of producing three dimensional image of an object due to the interference phenomena of coherent light waves on a photographic plate is known as *holography.* The term *holo* means whole or complete, i.e., the recorded interference pattern gives all the information about an object. A hologram has no resemblance to the original object; however, it gives all information about the object through optical codes. The salient feature of holography is the three dimensional nature of the images produced, whereas in photography, two dimensional images are produced.

(i) Basic principle When an object is illuminated by light from a coherent source, each point on the object acts as a source of secondary waves. These secondary waves spread out in all directions and it characterises the object through an interference pattern. This interference pattern is recorded using a suitable method. The recorded interference pattern is again illuminated with coherent light from another source. This results in reproduction of the original object in three dimensions.

Let us consider an object *AB* illuminated by light waves from a coherent source as shown in Fig. 11.33. As a result, each point on the object results in a spherical waves

The McGraw·Hill Companies

with coherent nature. These spherical waves interfere with the reference beam and hence produces interference or diffraction pattern at the plane of the photographic plate *P*. The pattern is developed and it contains light and dark partially absorbing fringes and this pattern is the characteristic of the object. This pattern is called as *Gabour Zone plate*.

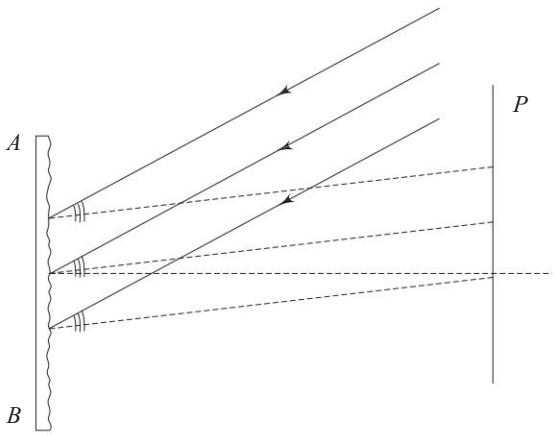


Fig 11.35 *Principle of holography*

(ii) Production of a Hologram Photographic plates are considered the best medium to record holograms. Recording depends only on intensity variation and not on amplitude variation. Therefore, in order to convert the amplitude variation into intensity variation, Gabour superimposed the reflected waves with a reference beam having the same wavelength. As a result, constructive and destructive interference fringes were obtained between the reflected waves and the reference beam. Thus, an increase in intensity at the constructive point and a decrease in intensity at the destructive point were noted. This change of intensity was recorded on the photographic plate.

In order to produce a high-quality hologram, the light source used must be highly monochromatic and coherent. Monochromatic laser beams are used to produce an intense beam of coherent light. Thus, the spatially coherent laser beam is divided into wavefronts or amplitudes. One part of the beam falls directly on the photographic plate and the other falls on the object whose image is to be recorded.

The experimental arrangement for the production of a hologram using the laser beam is shown in Fig. 11.34.

A laser beam from the source is made to fall on an optical device called the beam splitter. A part of the beam from the beam splitter is made to fall on the silver coated mirror M_1. After reflecting from the mirror M_1, the reference beam is passed through the lens L_1 and finally falls on the photographic plate. The focal length of lens L_1 is smaller than that of the lens L_2.

The other part of the beam from the beam splitter is made to fall on the mirror M_2 through the lens L_2 of larger focal length. The beam is reflected from the mirror M_2 and is made to fall on the object *AB*. The reflected waves from the surface of the object *AB* are also made to fall on the photographic plate. Thus, the interference pattern of

the reflected beam and reference beam is formed. The resultant pattern (hologram) is recorded on the photographic plate. The hologram of the object can be made by developing the plate using standard techniques.

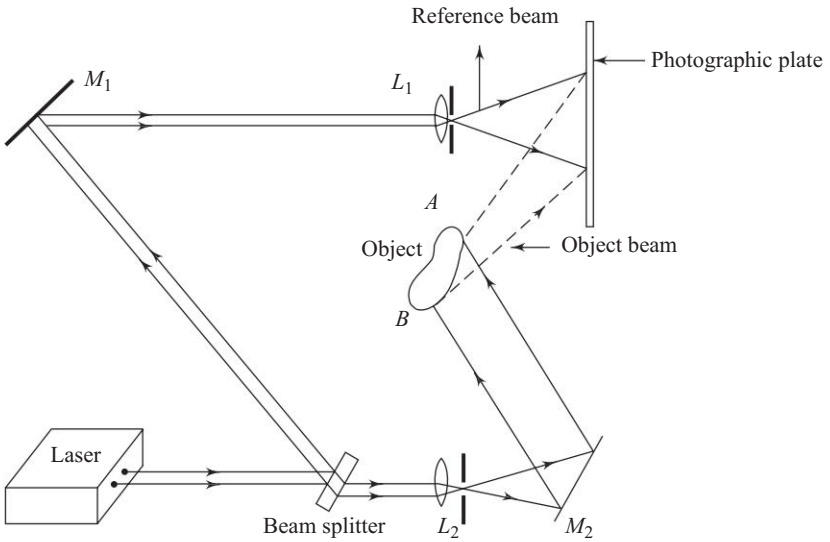


Fig. 11.36 *Experimental arrangement for production of a hologram of a three dimensional image*

In order to obtain a high resolution hologram, care should be taken to provide spatial filtering, the pattern lengths of the sample and the reference beam. The experimental arrangement of such a special laser holography is shown in Fig. 11.35.

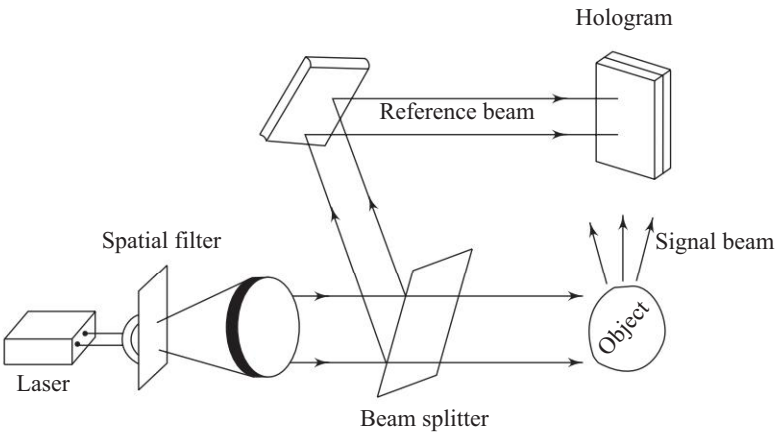


Fig. 11.37 *Experimental arrangement for a laser hologram*

The McGraw·Hill Companies

(iii) Reconstruction of Images The experimental arrangement used for the reproduction of three dimensional images of an object is shown in Fig. 11.36.

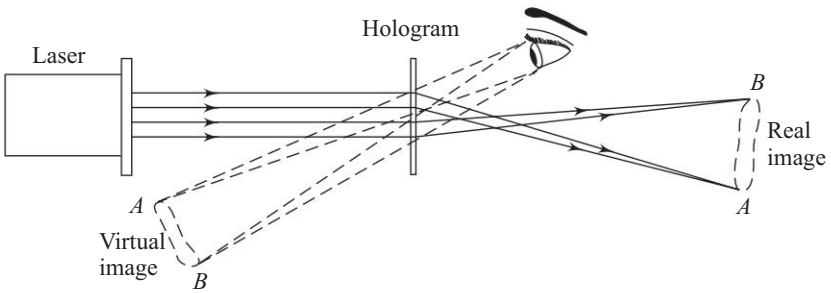


Fig. 11.38 *Reconstruction of a hologram of three-dimensional image*

For the reconstruction of the image, the recorded photographic plate is illuminated by a monochromatic laser beam. The illumination results in two images. The first image is formed behind the light source and forms a virtual image of the object. The second image is formed due to the diffraction of the beam at a distant place. The diffracted image is a real image and it can be photographed. The image obtained on the photographic plate is simply a two dimensional image of the original object.

(iv) Applications of Holography Some of the applications of holography in various fields are given below:

1. Three-dimensional images produced by holograms are used in educational and technical fields and also in advertising, artistic displays, etc.
2. For imaging the objects
3. Interferrometric imaging in astronomy
4. Ultrasonic imaging is used in the field of medicine, industries, etc.

11.10.10 Other Applications

In Communication

(i) In communication systems, the laser beams are used to transmit thousands of television programmes and simultaneous telephone conversations at a time.

(ii) The communication between the planets has been made possible using laser beams.

(iii) The laser light waves are not absorbed by water and hence it can be successfully employed to establish underwater communication between submarines.

In Computers

(i) By using optical fibres as light guides, the entire content of memory can be transmitted from one computer to another computer by a laser beam.

(ii) Laser is used in printers for computer printouts.

The McGraw·Hill Companies

In Meteorology A laser interferometer is used to measure length to a very high accuracy.

In Chemistry It is used to accelerate some chemical reactions to create new chemical compounds by destroying atomic bonds between the molecules.

In Photography Laser beams are used to obtain a three dimensional lensless holography (hologram).

In Weapons

(i) A powerful laser beam can be used to destroy big size objects like aeroplanes, missiles, etc., by pointing the laser beam on them, and hence it is called death ray.

(ii) The laser beam can be used to determine precisely the distance, velocity, direction, size and the nature of reflected light in radar. It is known as *LIDAR (light detection and ranging).*

In Industry

(i) The high power laser is used to weld or melt materials. It can also be used to make very small holes in hard materials like diamond, hard steel, etc., which cannot be made by mechanical drilling.

(ii) It can be used to test the quality of the material.

Keypoints to Remember

- The laser beam is strictly monochromatic, i.e, of single frequency and the bandwidth of laser beam is narrow $\Delta\gamma = 0$.
- Absorption is the process of absorbing energy from the incident light and their by excitation into higher energy state.
- Spontaneous emission is the process of returning the atoms eventually from the higher state to the ground state by emitting their excess energy.
- Stimulated emission is the process of stimulating an atom in the higher energy state to make a transition to the lower state with the creation of a second photon due to the interaction of photons of energy E.
- According to Einstein, the total emission is a sum of spontaneous and stimulated emission of radiation.
- Population inversion is the process of increasing population in the upper energy level so that it is greater than the lower energy level, to achieve the laser action.
- Different methods used to achieve population inversion are pumping, direct electron excitation, inelastic atom–atom collision and chemical reaction.
- Different types of lasers are solid state laser, gas laser, liquid laser, dye and chemical laser, and semiconductor laser.

The McGraw·Hill Companies

- The wavelength of laser radiation from a ruby laser is 6934Å.
- Nd–YAG laser radiates the laser light at a wavelength of 1.064 μm with a power output of 70 W.
- The active medium in helium–neon laser is helium and neon with a ratio of 10:1. The laser output is 6328Å.
- When the carbon atom is fixed in its position and each oxygen atom vibrates in the opposite direction symmetrical to the carbon atom in a straight line, it is known as symmetrical mode of vibration.
- In the bending mode, the oxygen atom and carbon atom-vibrate at right angles to the line passing through the centre of gravity.
- In a symmetric mode of operation, two oxygen atoms vibrate about the central C atom asymmetrically.
- Homojunction is a *p-n* junction made up of the same material with two regions of different conductivities.
- Heterojunction is an interface between two adjoining single crystal semiconductors with different band-gap energies.
- Two different types of heterojunctions are isotype and anisotype heterojunctions.
- The method of producing a three-dimensional image of an object due to the interference phenomena of coherent light waves on a photographic plate is known as holography.

Solved Problems

Example 11.1 The first line of the principal series of sodium is the *D* line at 590 nm. This corresponds to a transition from the first excited state (3p) to the ground state (3s). What is the energy in electron volts of the first excited state?

Given Data

The wavelength of the sodium *D* line: $\lambda = 590$ nm

$$\lambda = 5900 \times 10^{-10}\ \text{m}$$

Solution The eergy emitted $E = h\nu = \dfrac{hc}{\lambda}$

$$= \frac{6.626 \times 10^{-34} \times 3 \times 10^{8}}{5.9 \times 10^{-7}} = 0.369 \times 10^{-19}\ \text{J}$$

$$= \frac{3.369 \times 10^{-19}}{1.602 \times 10^{-19}} = \text{eV} = 2.10299\ \text{eV}$$

The energy of the first excited state = 2.1 eV.

Example 11.2 What fraction of sodium atom is in the first excited state in a sodium vapour lamp at a temperature of 250°C?

Given Data

The temperature, $T = 250^\circ\text{ C} = 250 + 273 = 523\text{ K}$

Solution Let N_2 be the population of the first excited state and N_1 be the population of the goun stte.

Then,
$$\frac{N_2}{N_1} = \frac{e^{-E_2/kT}}{e^{-E_1/kT}} = e^{-(E_2 - E_2)/RT}\, e^{-(h\nu)/kT}$$

$$= \exp\left[-\frac{6.626\times10^{-34}\times3\times10^{8}}{1.38\times10^{-23}\times523\times5.9\times10^{-7}}\right]$$

$$= \exp[-46.68] = 5.335\times10^{-21}$$

The ratio between the atoms in the first excited state and the ground state $= 5.335\times10^{-21}$.

Example 11.3 What is the ratio of the stimulated emission to spontaneous emission at a temperature of 250°C for the sodium *D* line?

Given Data

The temperature, $T = 25^\circ\text{C} = 523\text{ K}$

Solution
$$\frac{\text{Stimulated emission}}{\text{Spontaneous emission}} = \frac{1}{e^{h\nu/kT} - 1} = \frac{1}{e^{hc/kT\lambda} - 1}$$

$$= \frac{1}{\exp\left[\dfrac{6.626\times10^{-34}\times3\times10^{8}}{1.38\times10^{-23}\times523\times5.9\times10^{-7}}\right] - 1}$$

$$= 5.329\times10^{-21}$$

The ratio between the stimulated emission and spontaneous emission $= 5.329\times10^{-21}$.

Example 11.4 Calculate the threshold condition for the ruby laser in which the appropriate parameters are as follows: $\nu_0 = 4.3\times10^{14}$ Hz; $\Delta\nu_0 = 1.5\times10^{11}$ Hz; $n_0 = 1.76$; $\tau_{sp} = 4.3\times10^{-3}$ s ; $\tau_{photon} = 6\times10^{-9}$ s.

Given Data

The refractive index of the ruby rod,	$n_0 = 1.76$
The frequency,	$\nu_0 = 4.3\times10^{14}$ Hz
The doppler broadening,	$\Delta\nu_0 = 1.5\times10^{11}$ Hz
The lifetime of spontaneous emission,	$\tau_{sp} = 4.3\times10^{-3}$ s
The lifetime of photon,	$\tau_{photon} = 6\times10^{-9}$ s

The McGraw·Hill Companies

Solution The difference between the population of the excited state and the grond stte,

$$N_2 - N_1 = \frac{4\pi^2 v_0^2 n_0^3}{c^3} \times \left(\frac{\tau_{SP}}{\tau_{photon}} \right) \Delta v_0$$

$$= \frac{4\pi^2 (4.3 \times 10^{14})^2 \times 1.76^3}{(3 \times 10^8)^3} \times \frac{4.3 \times 10^{-3}}{6 \times 10^{-9}} \times 1.5 \times 10^{11}$$

$$= 1.584 \times 10^{23} \text{ m}^{-3}$$

The difference between the excited state and the ground state populations $= 1.584 \times 10^{23}$ m^{-3}

Example 11.5 Using Einstein's theory, show that in the optical region, say, at $\lambda = 5000$ Å and $T = 300$ K, the amplification is not possible.

Given Data

The wavelength of the incident light, $\lambda = 5000$ Å. The temperature, $T = 300$ K.

Solution The frequency of theincident light,

$$v = \frac{c}{\lambda} = \frac{3 \times 10^8}{5000 \times 10^{-10}} = 6 \times 10^{14} \text{ Hz}$$

$$\frac{\text{Stimulated emission}}{\text{Spontaneous emission}} = \frac{1}{e^{hv/kT} - 1}$$

$$= \frac{1}{\exp\left[\frac{6.626 \times 10^{-34} \times 6 \times 10^{14}}{300 \times 1.38 \times 10^{-23}} \right] - 1} = \frac{1}{e^{96} - 1}$$

$$= 2.03 \times 10^{-42}$$

This shows that the spontaneous emission is more predominant than that of the stimulated emission. For stimulating emission, $N_2 >> N_1$ should exist. Therefore, there is no amplification possibility. But, subsequent development in maintaining population inversion by pumping the atoms from lower level to higher level optically or electrically led to the discovery of lasers.

Example 11.6 A He–Ne laser emits light at a wavelength of 632.8 nm and has an output power of 2.3 mW. How many photons are emitted in each minute by this laser when operating?

Given Data

The wavelength of the laser beam, $\lambda = 632.8$ nm
$= 6328 \times 10^{-10}$ m

The output power, $P = 2.3$ mW

The McGraw·Hill Companies

Solution The frequency of the photon emittedby the laser beam

$$= \frac{c}{\lambda} = \frac{3 \times 10^8}{6328 \times 10^{-10}} = 4.74 \times 10^{14} \text{ Hz}$$

The energy of a photon, $h\nu = 6.626 \times 10^{-34} \times 4.74 \times 10^{14}$

$$= 3.14 \times 10^{-19} \text{ J}$$

Energy emitted by the laser $= 2.3 \text{ mW} = 2.3 \times 10^{-3} \text{ J s}^{-1}$

$$= 2.3 \times 10^{-3} \times 60 \text{ J min}^{-1}$$

The number of photons emitted $= \frac{2.3 \times 10^{-3} \times 60}{3.14 \times 10^{-19}}$

$$= 4.3949 \times 10^{17} \text{ photons min}^{-1}$$

The number of photons emitted $= 4.3949 \times 10^{17}$ photons min^{-1}.

Example 11.7 A cadmium sulphide photodetector crystals irradiated over a receiving area of 4×10^{-6} m^2 by a wavelength of 0.4×10^{-6} m and a intensity of 200 W m^{-2}. Assuming that each quantum generates an electron hole pair, calculate the number of pairs generated per second.?

Given Data

Area of the photodetector $= 4 \times 10^{-6} \text{ m}^2$

Wavelength of the cadmium sulphate crystal $= 0.4 \times 10^{-6}$ m

Intensity of the light $= 200 \text{ W m}^{-2}$

Solution

We know that number of photons $= \frac{\text{Intensity} \times \text{area}}{\text{Energy of photon} \left(\frac{hc}{\lambda}\right)}$

Energy of each photon is given by $= \frac{hc}{\lambda}$

Substituting the values, we get, $= \frac{6.626 \times 10^{-34} \times 3 \times 10^8}{0.4 \times 10^{-6} m}$

$$= 4.9695 \times 10^{-19} \text{ J}$$

The number of pairs generated per second is

$$= \frac{200 \times 4 \times 10^{-6}}{4.9695 \times 10^{-19}}$$

$$= 1.609 \times 10^{15}$$

Therefore, each photon quantum generates an electron hole pair, the number of photons is equal to the number of electron hole pairs, and the number of pairs generate per second is 1.609×10^{15}.

Example 11.8 Calculate the wavelength of radiation emitted by an LED made up of semiconductivity materials with band-gap energy of 2.8 eV

Given Data

Band gap of given semiconductor is $E_g = 2.8$ eV

$= 2.8 \times 1.609 \times 10^{-19}$

$= 4.4856 \text{ x } 10^{-19}$ J

Solution

We know that wavelength of the emitted radiation is

$$\lambda = \frac{hc}{E_g}$$

Substituting the values, we get,

$$= \frac{6.626 \times 10^{-34} \times 3 \times 10^{8}}{4.4856 \times 10^{-19}}$$

$= 0.443 \text{ x } 10^{-6}$ m

The wavelength of radiation emitted by an LED is 0.443 μm.

Example 11.9 Calculate the long wavelength of a extrinsic semiconductor if the ionisation energy is 0.02 eV.

Given Data

The ionization energy $= 0.02$ eV

$= 0.02 \times 1.609 \times 10^{-19}$

$= 3.218 \times 10^{-21}$ J

Solution

We know that long wavelength of extrinsic semiconductor is

$$\lambda = \frac{hc}{E_g}$$

Substituting the values, we get,

$$= \frac{6.626 \times 10^{-34} \times 3 \times 10^{8}}{3.218 \times 10^{-21}}$$

$= 6.177 \times 10^{-5}$ m

The long wavelength of the extrinsic semiconductor is 61.77 μm.

Example 11.10 Calculate the number of photons from a green light of mercury ($\lambda = 4961$) required to do one joule of work.

Given Data

The wavelength of green light from mercury lamp $= 4961 \times 10^{-10}$ m

Solution

We know that the energy $E = \frac{hc}{\lambda}$

Substituting the values, we get,

$$= \frac{6.626\times10^{-34}\times3\times10^{8}}{4961\times10^{-10}}$$

$$= 4.007 \times 10^{-19} \text{ J}$$

Number of photons required to do one joule of work

$$= \frac{1 \text{ Joules}}{4.007\times10^{-19}}$$

$$= 2.495 \times 10^{18}$$

The number of photons from green light of mercury is 2.495×10^{18} m^{-3}.

Example 11.11 For a semiconductor laser the band gap is 0.80 eV. What is the wavelength of light emitted from it?

Given Data

Band gap of given semiconductor is $E_g = 0.8$ eV

$$= 0.8 \times 1.609 \times 10^{-19}$$

$$= 1.28 \times 10^{-19} \text{ J}$$

Solution

We know that wavelength emitted from the given semiconductor laser is

$$\lambda = \frac{hc}{E}$$

Substituting the values, we get,

$$= \frac{6.626\times10^{-34}\times3\times10^{8}}{1.28\times10^{-19}}$$

$$= 1.55 \times 10^{-6} \text{ m}$$

The wavelength of light emitted from the given semiconductor is 1.55 μm.

Example 11.12 Calculate the wavelength of emission from a GaAs semiconductor laser whose bandgap energy is 1.44 eV.

Given Data

Band gap of given GaAs semiconductor $E_g = 1.44$ eV

$$= 1.44 \times 1.609 \times 10^{-19}$$

$$= 2.3 \times 10^{-19} \text{ J}$$

Solution

We know that wavelength of emitted from given GaAs semiconductor laser is

$$\lambda = \frac{hc}{E_g}$$

Substituting the values, we get,

$$= \frac{6.626\times10^{-34}\times3\times10^{8}}{2.3\times10^{-19}}$$

$$= 8.6\times10^{-7}\ \text{m}$$

The wavelength of light emitted from a GaAs semiconductor is 0.86 μm.

Example 11.13 Calculate the wavelength of the emission from GaAs semiconductor laser whose band energy is 3 eV.

Given Data

Band gap of given GaAs semiconductor is $E_g = 3.0\ \text{eV}$

$$= 3.0\times1.609\times10^{-19}$$

$$= 4.8\times10^{-19}\ \text{J}$$

Solution

we know that, the wavelength of emitted from given GaAs semiconductor is

$$\lambda = \frac{hc}{E}$$

Substituting the values, we get,

$$= \frac{6.626\times10^{-34}\times3\times10^{8}}{4.8\times10^{-19}}$$

$$= 4.141\times10^{-7}\ \text{m}$$

The wavelength of light emitted from the GaAs semiconductor is 0.4141 μm.

Objectives Questions

11.1 The equation for angular spread is _____.
11.2 The angular spread for an ordinary light beam is_____.
11.3 The angular spread for a typical laser beam is _____.
11.4 The line width of laser emitted by ruby laser is_____ Å .
11.5 The energy of photon is equal to _____.
11.6 The rate of upward transmition after the absorption is equal to

a) $R_{12} = B_{12}\,\rho_\gamma\,N_1$ b) $R_{12} = \rho_\gamma\,N_1$

c) $R_{12} = B_{12}\,\rho_\gamma\,N_2$ d) $R_{12} = B_{12}\,\rho_\gamma\,N_1\,N_2$

11.7 The rate of spontaneous emission is equal to

a) $R_{21}(\text{sp}) = A_{21}N_2\,\rho$ b) $R_{21}(\text{sp}) = A_{21}N_1N_2$

c) $R_{21}(\text{sp}) = A_{21}\neg N_1\,\rho\gamma$ d) $R_{21}(\text{sp}) = A_{21}N_1N_2$

11.8 The rate of stimulated emission is equal to

a) $R_{21}(\text{ST}) = B_{21}N_1N_2$ b) $R_{21}(\text{ST}) = B_{21}\,\rho_\gamma\,N_1$

c) $R_{21}(\text{ST}) = B_{21}\,\rho_\gamma\,N_2$ d) $R_{21}(\text{ST}) = B_{21}\,\rho_\gamma N$

11.9 The mathematical expression for existence of stimulated emission was proposed by _____.

11.10 Which event is likely to take place when a photon of energy equal to the difference in energy between two levels is incident in a system?

a) Absorption b) Emission

c) Absorption and emission d) None of the above

11.11 The condition for population inversion is

a) $\frac{N_2}{N_1} = e(E_2 - E_1)/kT$ b) $\frac{N_2}{N_1} = e^{-(E_1-E_2)/KT}$

c) $\frac{N_1}{N_2} = e^{-}(E_1 - E_2)/kT$ d) $\frac{N_2}{N_1} = e^{-}(E_2 - E_1)/kT$

11.12 Complete the chemical reaction, which is used to obtain population inversion is chemical laser, $H_2^+ \rightarrow$ 2HF.

11.13 The population inversion takes place at _____ medium.

11.14 During the laser amplification process, total loss is not due to the following reason:

a) Transition, absorption and scattering by the mission

b) Absorption with in the medium

c) Scattering due to optical inhomogenity

d) Energy loss

11.15 Schawlow and Tones condition for population inversion to active laser action is equal to

a) $\left(\frac{n_o^3 w_o^2}{2\pi c^3}\right)\left(\frac{\tau_{sp}}{Y_{photons}}\right)\Delta w_o$ b) $\left(\frac{n_o^3 w_o^2}{2\pi c}\right)\left(\frac{1^{sp}}{Y_{photons}}\right)\Delta w_o$

c) $\left(\frac{n_o^3 w_o^2}{2\pi c^3}\right)\left(\frac{1sp}{Y_{photons}}\right)\Delta w_o$ d) $\left(\frac{n_o^3 w_o^2}{2\pi c^3}\right)\left(\frac{1_{sp}}{Y_{photons}}\right)\Delta w_o$

11.16 Examples for solid laser are _____ and _____.

11.17 In a ruby laser, the active medium consist of aluminum oxide doped with 0.05 wt. of _____.

11.18 In a ruby laser, the laser action is achieved by _____ method.

11.19 Ruby laser radiates an intense pulse laser of wavelength _____ Å.

The McGraw·Hill Companies

11.20 In the Nd–YAG laser, the active medium is made up of _____ and _____.
11.21 _____ method is used to achieve population inversion in an Nd–YAG laser.
11.22 The light source used for optical pumping is _____ or _____.
11.23 The wavelength of laser light radiated by Nd–YAG laser is _____.
11.24 The active medium of a helium–neon laser is made up of _____ ratio of helium and neon.
11.25 Atoms can stay in metstable states.
11.26 Complete the following reaction for helium-neon laser action: $He^* + Ne \rightarrow He + \rightarrow +$ _____.
11.27 The wavelength of laser light emitted from He–Ne laser is _____.
11.28 _____ is molecular gas laser.
11.29 The most powerful transition in CO_2 laser at normal operating temperature is at ____.
11.30 In a helium–neon laser, helium is used to decrease the _____ in the lower levels.
11.31 The power output from CO_2 laser is _____
11.32 Diode laser consists of a _____ doped in a single crystal.
11.33 LED is converted into a laser diode employing a _____.
11.34 The wavelength of light emitted from Ga As laser at room temperature is _____Å.
11.35 An example for injection laser is _____.
11.36 Injection lasers are prepared using _____ heterojunctions.
11.37 Achronium for LIDAR is_____.

KEY

11.1 λ/d
11.2 1m per 1m
11.3 1mm per 1m
11.4 5×10^{-4}
11.5 h
11.6 a)
11.7 a)
11.8 c)
11.9 Einstein
11.10 c)
11.11 d)
11.12 F_2
11.13 active
11.14 d)
11.15 a)
11.16 Ruby, Nd-YAG
11.17 Chromium oxide
11.18 Optical pumping
11.19 6934
11.20 Ytrium aluminum garnet, Neodymium
11.21 Optical pumping
11.22 Xerox, Krpton
11.23 1.064μm
11.24 10:1
11.25 False
11.26 Ne
11.27 6328 Å
11.28 Co_2
11.29 10.6 μm
11.30 Population
11.31 10 kw
11.32 *p-n* junction
11.33 high current
11.34 8874
11.35 Homojunction-Ga-As
11.36 Isotype
11.37 Light detection and ranging

Exercises

11.1 Calculate the value of stimulated emission coefficient during a transition from an excited state (2p) to the ground state (2s). The lifetime of the excited state (2p) for spontaneous emission is 1.6×10^{-9} s. The energy difference between the excited state (2p) and the ground state (2s) is 10.2 eV.

11.2 Calculate the condition for population inversion in a He–Ne laser using the following data: $\nu_0 = 4.74 \times 10^{14}$ Hz; $\Delta\nu_0 = 1300$ MHz; $\tau_{rad} = 7 \times 10^{-7}$ s; $\tau_{photon} = 10^{-7}$ s; $n_0 = 1$.

11.3 Examine the possibility of light amplifying device in the microwave region, say, at $\lambda = 10$ cm and at room temperature ($T = 300$ K).

11.4 A laser beam can be focused on an area equal to the square of its wavelength (λ^2). For a He–Ne laser, $\lambda = 6328$ Å. If the laser radiates energy at the rate of 1 mW, find the intensity of the focused beam.

11.5 Calculate the number of photons emitted per second by a 5 mW laser, assuming that it emits light of the wavelength 6328 Å.

11.6 If laser action occurs by the transition from an excited state to the ground state $E_1 = 0$, and if it produces light of 693 nm wavelength, what is the energy level of the excited state.

11.7 Transition occurs between a metastable state E_3 and an energy state E_2 just above the ground state. If emission is at 1.1 μm and $E_2 = 0.4 \times 10^{-19}$ J, find the energy of the E_3 state.

11.8 Imagine that we chop a continuous laser beam (assumed to be perfectly monochromatic, $\lambda_0 = 623.8$ nm) into 0.1 ns pulses using some sort of shutter. Compute the resultant line width $\Delta\lambda_0$, bandwidth an coherence length.

Short Questions

11.1 What does the term 'laser' stand for?
11.2 What is absorption of light?
11.5 What is meant by spontaneous emission?
11.6 What is stimulated emission?
11.8 Explain the principle of laser.
11.9 What are the conditions needed for laser action?
11.10 What is meant by population inversion?
11.11 What is meant by pumping of atoms?
11.16 What are the different types of pumping?
11.17 Write the Schawlow–Townes condition for laser action and explain the terms.

The McGraw·Hill Companies

11.18 What are the different types of laser?
11.19 Mention any two applications of laser beams.
11.20 What is three dimensional profiling?
11.21 Explain how laser beam is used to detect surface defect.
11.25 Explain how laser beam is used in laser endoscopy.
11.27 Why is optical pumping mechanism not suitable for in pumping CO_2, lasens?
11.28 Explain the application of laser beam in computer peripherals.
11.29 What is meant by hologram?
11.30 Explain the construction of hologram.
11.31 Explain the reconstruction of hologram.
11.32 What are the advantages of a hologram?
11.34 What is the reason for using a laser beam for the construction and reconstruction of a hologram?

Descriptive Questions

11.1 Derive Einstein's relation for stimulated emission, and hence explain the existence of stimulated emission.
11.2 Obtain Schawlow–Townes condition for laser oscillation.
11.3 Explain the different methods used for pumping of atoms.
11.4 Explain with a neat sketch the construction and working of ruby laser.
11.6 What is a gas laser? Explain the working of He–Ne laser with relevant diagrams.
11.8 What is a semiconductor diode laser? Explain the construction and working of a Ga–As laser.
11.10 Write short notes on the following.
(i) Three dimensional profiling
(ii) Computer peripherals
(iii) Material processing
11.11 Explain with a neat sketch the construction and reconstruction of a hologram using laser beam.
11.12 Discuss the applications of laser in welding, heat treatment and cutting.

Chapter

Fiber Optics

12

OBJECTIVES

- To understand the principle behind fiber optic communication
- To gain knowledge about the structure of optical fibers and the ways of propagation of optical pulses through fibers
- To study the classification of optical fibers based on materials, refractive index and number of modes propagated through it
- To study the different process methods of optical fibers
- To study the applications of optical fibers in all fields of science and technology

12.1 INTRODUCTION

Fiber optics is an overlap of science and technology which deals with transmission of light waves into optical fibers, their emission and detection. Optical cables are small in size and weight and are more efficient for communication than metal cables. The principle behind the transmission of light waves in an optical fiber is total internal reflection. In this chapter, the principle, preparation, characteristic properties and applications of optical fibers, particularly in the field of electronics and medicine, are discussed briefly:

12.2 OPTICAL FIBERS–PRINCIPLE

The fundamental optical parameter of an optical fiber is refractive index or index of refraction (n). The refractive index of the material is

$$n = \frac{c}{v} \tag{12.1}$$

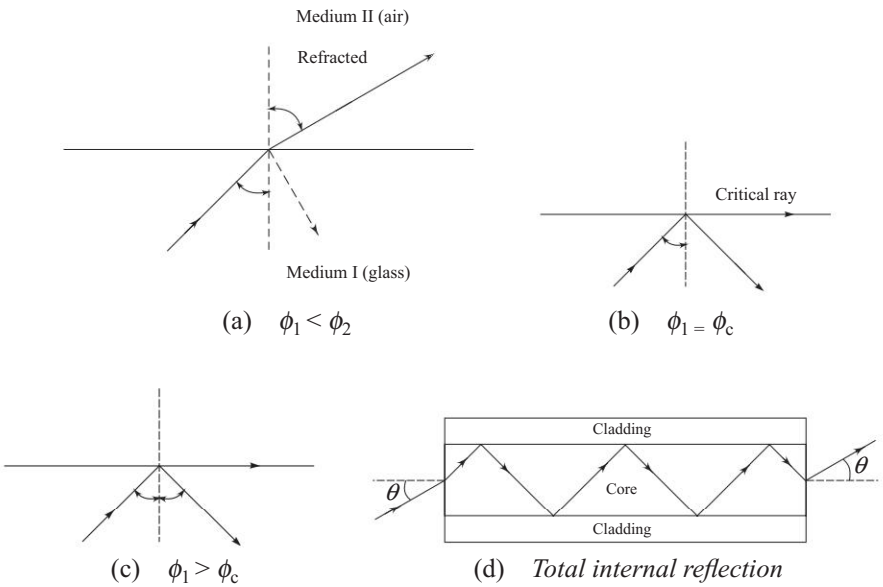
The McGraw·Hill Companies

where c is the speed of light in free space (3×10^8 m s^{-1}) and v, the speed of light in the medium. The refractive index of air, water and glass are respectively 1.00, 1.33 and 1.50. Consider that a high-refractive-index glass medium—I and low-refractive-index air medium—II are in contact as shown in Fig. 12.1. Consider an incident ray of light is passing through the medium of refractive index n_1. When the light ray is incident at the boundary of the two mediums, a part of the ray gets reflected back into the medium I, while a part is refracted into the medium II. The phenomena of reflection or refraction at the interface between the two mediums depend on the speed of light in the respective medium. According to Snell's law, angle of incidence is (ϕ_1) and angle of reflection is (ϕ_2). When the angle of incidence ϕ_1 is increased, the angle of refraction ϕ_2 increases (Fig. 12.1a). When the angle of incidence reaches the critical angle ϕ_c, the angle of the refracted ray approaches 90^o and almost emerges out parallel to the interface between the two media (Fig. 12.1b). When the angle of incidence exceeds the critical angle (Fig. 12.1c), the incident light ray gets reflected in the same medium. The reflection of light in the same medium at a particular angle of incidence is known as *total internal reflection* (Fig. 12.1d). Therefore, ϕ_c is known as the *maximum angle for total internal reflection*. The same can be obtained by considering Snell's law,

$$n_1 \sin \phi_1 = n_2 \sin \phi_2$$

or

$$\sin \phi_1 = \frac{n_2}{n_1} \sin \phi_2 \qquad (12.2)$$



(a) $\phi_1 < \phi_2$ (b) $\phi_1 = \phi_c$

(c) $\phi_1 > \phi_c$ (d) *Total internal reflection*

Fig. 12.1 *Fiber optics—principle*

When total internal reflection takes place, $\phi_1 = \phi_c$ and $\phi_2 = 90^o$.

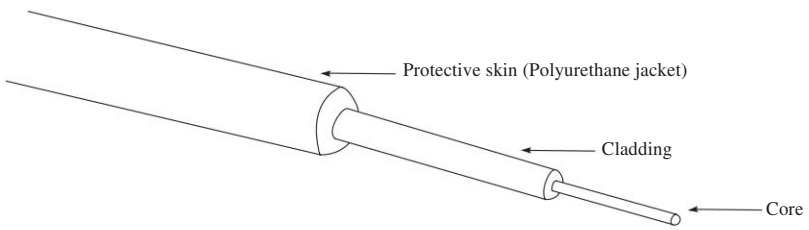
Therefore,

$$\sin \phi_c = \frac{n_2}{n_1} \tag{12.3}$$

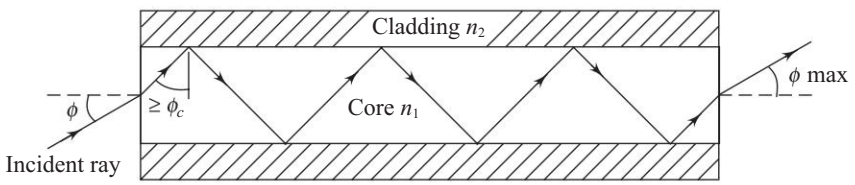
We know that sin $90^o = 1$. Therefore, Eq. (12.3) determines the critical angle for an optical fiber.

12.3 STRUCTURE OF OPTICAL FIBERS

An optical fiber, which is in the form of a cylinder, consists of an inner and outer shell. The inner shell is known as *core* and has high refractive index (n_1), while the outer shell is known as *cladding* and has low refractive index (n_2). In order to achieve total internal reflection within the core, the refractive index of the core is made slightly higher than the cladding material. Generally, the thickness of the core is nearly 50 μm while for the cladding, it is 125–150 μm. Therefore, the overall fiber diameter of an optical fiber is 125–150 μm. A typical optical fiber with refractive index $n_1 > n_2$ is shown in Fig.12.2(a). These fiber cables are bundled together and used for transmission applications. A typical three dimensional view of the optical fibre is shown in Fig. 12.2(b)



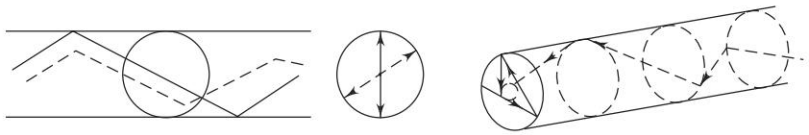
(a) *Three dimensional view*



(b) *Cross-sectional view*

Fig. 12.2 *Optical fiber—structure*

Light Rays of Propagation In optical fibers, the light rays are transmitted in two ways, namely, meridional and skew. The schematic representation of the meridional and skew ray's propagation is shown in Fig. 12.3. The ray which passes through the axis of the core is known as *meridional ray*. On the other hand, the ray which never passes through the core of the axis is known as a *skew ray*.



(a) *Meridional* (b) *Skew rays*

Fig. 12.3 *Optical fiber—propagation of light rays*

12.4 ACCEPTANCE ANGLE AND CONE

The angle of acceptance or maximum/total acceptance angle determines the propagation of light rays through optical fibers by total internal reflection. The angle of acceptance and acceptance cone for launching of light rays is shown in Fig. 12.4. Consider that an incident ray *OA* is incident on the fiber core at an angle ϕ_a.

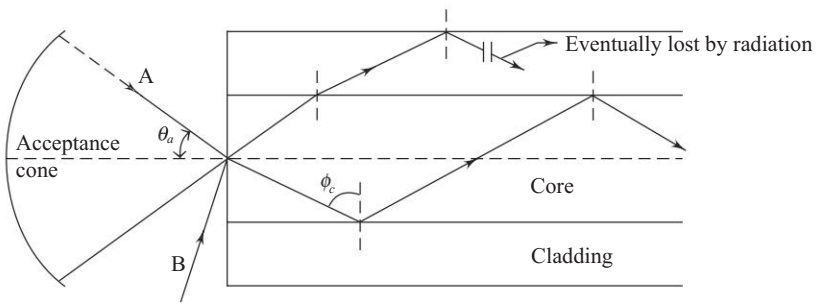


Fig. 12.4 *Optical fiber—acceptance angle and cone*

The incident rays enter into the core and are incident at the core–cladding interface, and hence, are transmitted into the cladding. Thus, they do not undergo total internal reflection at the interface. On the other hand get refracted at the core–cladding interface into the cladding. The refracted ray propagates further and is lost due to radiation. When the ray is incident at an angle greater than ϕ_a, it gets transmitted at an angle less than ϕ_c into the core–cladding interface without undergoing total internal reflection. The ray *BC* enters at an angle greater than θ_a is and lost in propagation due to radiation. On the other hand, the ray *OA* which enters at an angle less than θ_a ia incident on the fiber core at an angle greater than ϕ_c, and hence, propagates into the core through total internal reflection. Therefore, the angle ϕ_a, i.e., maximum angle at or below which the light rays undergo total internal reflection, is known *as acceptance angle*. The core $OO^!$ is known as the *acceptance cone.*

12.5 NUMERICAL APERTURE AND ACCEPTANCE ANGLE

The amount of light rays accepted by the fiber is known as *Numerical Aperture* (*NA*). The angle of acceptance of the incident rays and the refractive index of media involved in optical fibers, namely, air, core and cladding are used to obtain the numerical aperture. The parameters like refractive index of core, cladding and air are used to

The McGraw·Hill Companies

define the numerical aperture of the optical fiber. Let n_1, n_2 and n_0 be respectively the refractive index of the core, cladding and air. Consider that an incident light is incident at an angle ϕ_1 which is smaller than the acceptance angle ϕ_a. The incident ray from the air medium of refractive index n_o enters into the core of the optical fiber of refractive index n_1. The refractive index of air is lower than the cladding.

Let *OA* be the fiber axis, which is normal to the fiber core. Hence, the incident ray *AB* undergoes refraction at the air–core interface as shown in Fig. 12.5. According to Snell's law, the angle of refraction is

$$n_o \sin \phi_1 = n_1 \sin \phi_2 \tag{12.4}$$

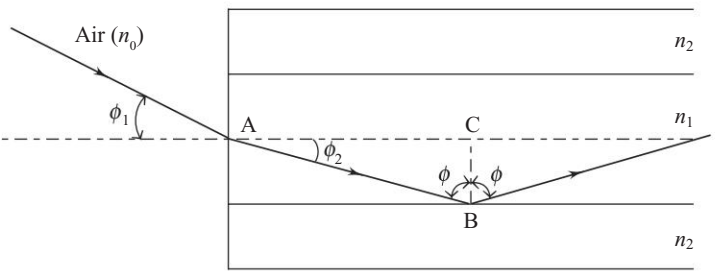


Fig. 12.5 *Optical fibers—numerical aperture*

The refracted ray *BC* emerges out at an angle of ϕ_2 as shown in Fig.12.5.

According to Snell's law, when the rays are incident and emerge at angles of respectively ϕ_1 and ϕ_2,

$$n_o \sin \phi_1 = n_1 \sin \phi_2 \tag{12. 5}$$

We know from triangle *BCE,*

$$\phi = \frac{\pi}{2} - \phi_2$$

$$\therefore \quad \sin \phi_2 = \cos \phi$$

Substituting the value of $\sin \phi_2$, in Eq. (12.5), we get,

$$n_0 \sin \phi_1 = n_1 \cos \phi \tag{12. 6}$$

We know that, $\sin^2 \phi + \cos^2 \phi = 1$

Therefore, Eq (12.6), can be written as

$$n_0 \sin \phi_1 = n_1 (1 - \sin^2 \phi)^{1/2}$$

The light rays are transmitted into optical fibers through total internal reflection and hence,

$$n_1 \sin \phi = n_2 \sin 90$$

$$n_1 \sin \phi = n_2$$

Substituting the above condition in Eq. (12. 6), we get,

$$n_0 \sin \phi_a = n_1 (1 - \sin^2 \phi)^{1/2}$$

or

$$= (n_1^2 - n_2^2)^{1/2}$$

Therefore,

Numerical aperture (NA) = $n_o \sin \phi_a$ (12.7)

We know that refractive index of air is 1 and hence, $n_o = 1$.

Therefore, NA = $\sin \phi_a = \left(n_1 - n_2^2\right)^{\frac{1}{2}}$

or $$NA = \left(n_{core}^2 - n_{cladding}^2\right)^{1/2} \quad (12.8)$$

12.5.1 Refractive Index (or) Core–Cladding Index Difference

The numerical aperture $$NA = \left(n_1^2 - n_2^2\right)^{1/2} \quad (12.9)$$

$$NA = n_1 (2\Delta)^{1/2} \quad (12.10)$$

where $$\Delta = \frac{n_1^2 - n_2^2}{2n_1^2}$$

When the core–cladding index difference is less than 1 (i.e., $\Delta << 1$)

we get, $$\Delta = \frac{n_1 - n_2}{n}$$

12.5.2 Number of Modes through Optical Fibers

Let N be the number of optical modes propagated through an optical fiber.
Therefore, the number of modes (N)

$$N = 4.9 \left(\frac{d\ \text{NA}}{\lambda}\right)^2 \quad (12.11)$$

where d is the diameter of the fiber core; λ the wavelength of light propagated; and NA, the numerical aperture.
The diameter of the core, when the optical fibers transmit a single mode is

$$d \leq \frac{0.76}{\text{NA}} \lambda \quad (12.12)$$

Similarly, the diameter of the core, when the optical fibers transmit in multimode is

$$d > \frac{0.76}{\text{NA}} \lambda\lambda \quad (12.13)$$

Equations. (12.12) and (12.13) are respectively used to determine the diameter of the core for single mode and multimode fibres.

12.6 TYPES OF OPTICAL FIBERS

Optical fibers are classified into three major categories based on the materials used for making optical fibers, number of modes transmitted into the fibers and the refractive index profile of the fibers. these are further classified into different categories as given in Fig. 12.6.

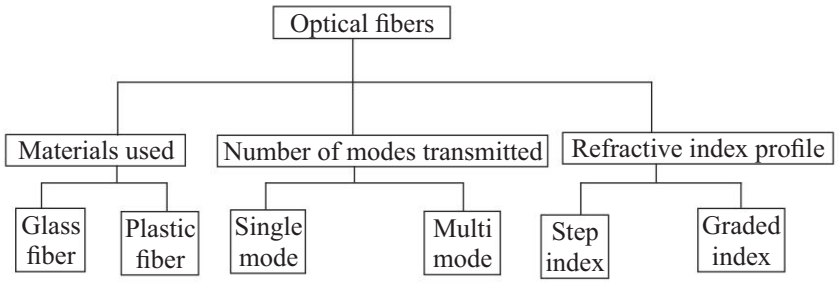


Fig. 12.6 *Classification of optical fibers*

Let us discuss the classification of optical fibers in brief in the following heading:

12.6.1 Classification Based on Materials Used

Generally, optical fibers are made up of materials like silica and plastic. Therefore, there are two different types of fibers, namely, glass and plastic fibers. The most important properties required for the selection of optical fibers are that the material should be an efficient guide for light waves; it should have low scattering cross-section and low loss of optical energy due to the absorption, attenuation and dispersion during the propagation of waves.

(a) Glass Fiber In glass fibers, the core and cladding are made of glass. Glass fibers are obtained by fusing metal oxides and silica glasses. In glass fibers, the material is made of a random network as that of non-crystalline materials. The examples of different materials used for glass-fiber core and cladding are given in Table 12.1.

Table 12.1 *Core and cladding materials used in glass and plastic fibers*

Fiber nature	*Glass fiber*		*Plastic fiber*	
	Materials	*Refractive index*	*Materials*	*Refractive index*
Core	GeO_2–SiO_2 P_2O_5–SiO_2	$n_1 > 1.458$	Polystyrene	$n_1 = 1.60$
Cladding	SiO_2	$n_2 = 1.458$	Methyl-methacrylate	$n_2 = 1.49$

Generally, silica is the most common material used in a glass fiber. The refractive index of silica is 1.458 at 850 nm. In order to produce a slight difference in core and cladding media, metal oxides like B_2O_3, GeO_2 or P_2O_5 are added to SiO_2. Glass fibers are made using low-melting silicates like soda-lime silicates, germanesilicates and various borosilicate glasses.

(b) Plastic Fiber Plastic fibers are made up of plastics. they are easy to handle without any care due to their toughness and durability. Plastic fibers provide high numerical aperture (~0.58) and larger angle of acceptance (~70^o) than glass fibers. This is mainly due to high refractive index differences between the core and cladding.

The McGraw·Hill Companies

A typical composition of core and cladding material used in plastic fiber is shown in Table 12.1.

12.6.2 Classification of Optical Fibers Based on Number of Modes

Based on the number of modes transmitted into fibers, optical fibers are classified into two types, namely,

(a) Single mode and

(b) Multimode.

(a) Single-Mode Fibers Single-mode or monomode fibers are fibers which transmit a single mode during transmission. The diameter of a single mode fiber is very small and it is in the order of 5–10 μm. The diameter of the cladding is very high (~125 μm) than the core. The typical structure of single mode fibers is shown in Fig. 12.7. Generally, single-mode fibers are step-index fibers. At a particular wavelength, say 0.8 μm, the core radius is 3 μm while the numerical aperture is 0.1. The signal is transmitted into the fiber without any dispersion.

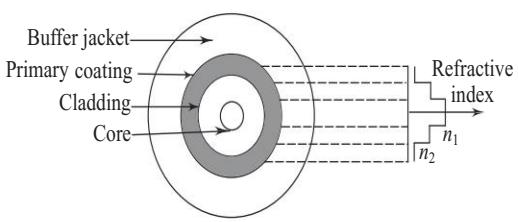


Fig. 12.7 *Single-mode fiber-structure*

Advantages

The following are the advantages of single-mode fibers:

- They facilitate to transmit information to longer distances due to negligible dispersion.
- They are free from intermodal dispersion. The intermodal dispersion is the time difference between the entry of pulse at one end of optical fiber and arrival of the pulse at the other end of the fiber.
- They are used in long-distance communication due to higher bandwidth.

Disadvantages

The following are the disadvantages of single-mode fibers:

- It is difficult to launch the light through the fiber.
- It is difficult to achieve end-to-end connection of similar fibers.

(b) Multimode fiber The optical fiber which is used to transmit more than one mode is known as a multimode fiber. The multimode fibres are preferable mainly due to its larger core radii than single-mode fibers. The structure of a multimode fiber is shown in Fig. 12.8.

The McGraw·Hill Companies

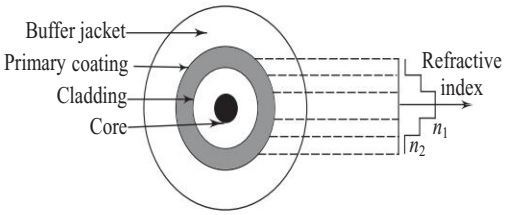


Fig. 12.8 *Multimode fiber-structure*

Disadvantages

The following are the disadvantages of multimode fibers:

- Multimode fibers suffer due to intermodal dispersion and hence, dispersion takes place during long transmission.
- They can be used to transmit signals only for short distances.

The differences between single and multimode fiber are shown in Table 12.2 for comparison.

Table 12.1 *Difference between single and multimode fibers*

Single mode fiber	*Multimode fiber*
It is used to transmit one single mode and hence, known as single-mode fiber.	It is used to transmit more than one or multi modes and hence, known as multimode fiber.
The core diameter is very small.	The core diameter is very large.
The difference in the refractive index of core and cladding is very small.	The difference in the refractive index of core and cladding is very large due to larger core diameter.
Optical light power using LEDS cannot be connected.	Optical light power using LEDS can be easily connected into fibers.
It is very difficult to connect two fibers (i.e., end-to-end connection)	End-to-end connections are very easy.
There is no intermodal dispersion, i.e., no time difference between the entry and the arrival of an optical signal in a fiber.	It suffers due to intermodal dispersion. Thus, it gives a time delay between the entry and arrival of an optical signal in a fibre.
It is used for long distance transmission applications.	It is used for short distance transmission applications.

12.6.3 Classification of Optical Fibers based on the Refractive Index Profile

Optical fibers are further classified into two types based on the refractive index profile of core and cladding as

(i) step index, and

(ii) graded index.

(i) Step Index Fibers The refractive indices of air, core and cladding varies step by step and hence, it is known as step-index fiber. The structure of a step-index fiber is shown in Fig. 12.9.

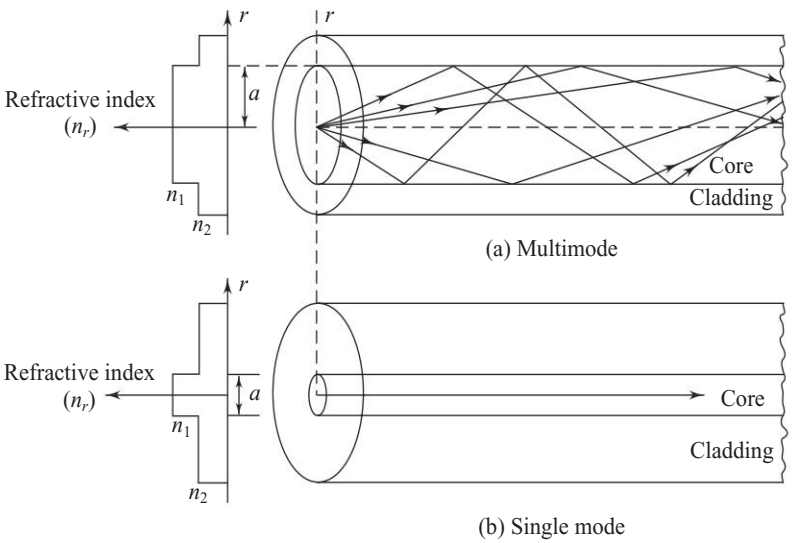


Fig. 12.9 *Structure of step-index fiber*

One can conclude the refractive index (n) profile based on radial distance (r) from the fiber axis as
For core,

$$n(r) = n_1;\ r < a \tag{12.14}$$

For cladding,

$$n(r) = n_2;\ r \geq a \tag{12.15}$$

Step index fibers are used both for single and multimode signals as shown in Fig. 12.9(a). The core diameter facilitates to divide the core of the fiber either as single or multimode. When the core diameter is 50 μm or greater then it is used for multimode transmission (Fig. 12.9a). On the other hand, if the core diameter is in the order of 2–10 μm, then it is used for a single-mode fiber (Fig. 12.9b).

The structure of a single mode step index fiber is shown in Fig. 12.10. In order to avoid the loss due to absorption, the core and cladding materials are made using high purity glasses. The diameter of the cladding is 10 times higher than the core. The increase in diameter of the core helps to improve the easy handling. The core is then covered by primary coating by a less purity glass material. A buffer jacket covers the primary coating and prevents the fiber from damages.

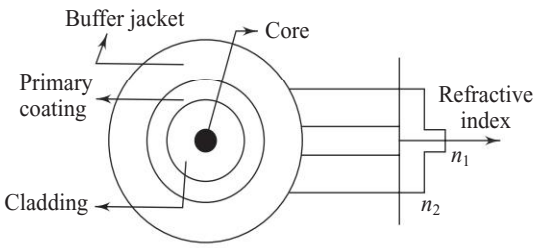


Fig. 12.10 *Wave guide structure—single mode fiber*

The McGraw·Hill Companies

Let n_1 and n_2 respectively be the refractive index of the core and cladding. Then,

$$n_2 = n_1 (1-\Delta) \tag{12.16}$$

where Δ is the core–cladding index difference. It is also known as core index difference.
One can select the value of n_2 so as to keep Δ in the order of 0.01. Thus, the core refractive index is larger than the cladding index and helps to propagate electromagnetic energy into the fiber through total internal reflection at the core–cladding interface.

(ii) Graded Index Fiber In case of a graded index fiber, the refractive index of the core varies gradually with increase in radial distance from the axis of the fiber. The refractive index is maximum at the fiber axis and decreases gradually with increase in radial distance and reaches a minimum at the core–cladding interface. Thus, it is called graded index fiber. The refractive index is equal at the core–cladding interface. The structure of the graded index fiber is shown in Fig.12.11. Therefore, the refractive index *(n)* profile of the core and cladding from the axis of the fiber is
For core,

$$n(r) = n_1(1-2\Delta(r/a)^{\alpha})^{1/2}: r<a \tag{12.17}$$

For cladding,

$$n(r) = n_1 (1-2\Delta)^{1/2} = n_2; r \geq a \tag{12.18}$$

where α is the characteristic refractive index profile of the fiber core. It is also known as *profile parameter of optical fiber.*
The core–cladding index difference for the graded index fiber is

$$\Delta = \frac{n_1^2 - n_2^2}{2n_1^2} \simeq \frac{n_1 - n_2}{n_1} \tag{12.19}$$

From Eqs. (12.15) and (12.16), the nature of the fiber can be explained based on the refractive index profile as shown in Fig. 12.11. When α = ∞, the fiber will be step index.

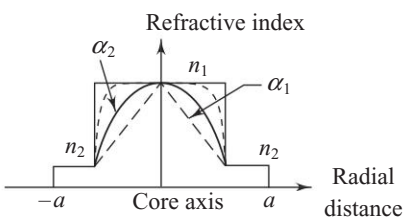


Fig. 12.11 *Graded index fiber wave guide-structure*

When α = 2, the fiber will be parabolic. When α = 1, the fiber will be triangular. A graded fiber with a parabolic refractive index profile (α = 2) is used to propagate multimode signals as shown in Fig. 12.11. In a graded index fiber, the refractive

index is maximum at the fiber axis and it decreases from the fiber core. When the ray enters into the core, it undergoes much refraction as shown in Fig. 12.12. As a result, the intermodal dispersion is less in case of multimode graded index fiber than in a multimode step index fiber.

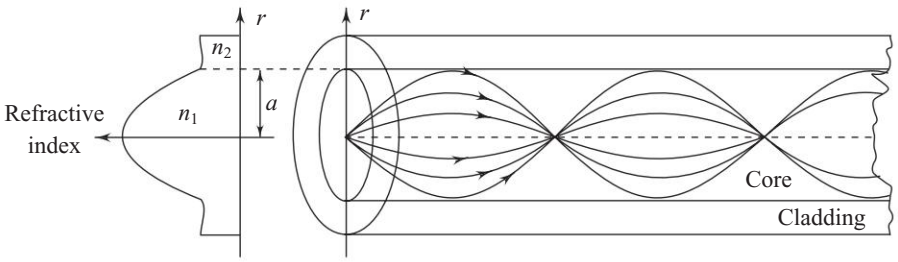


Fig. 12.12 *Multimode graded fiber– ray transmission*

The structure of a multimode graded index fiber is shown in Fig. 12.13. High pure materials are used for the processing for an optical fiber to avoid losses due to absorption. The thickness of the cladding will be very small, of the order of 10 wavelengths. The cladding is mainly used as a barrier to prevent the impurity ions from entering into the core material. Generally, for the NA of 0.2–0.3, the diameter of the core is of the order of 50–60 μm. A graded index fiber with high numerical aperture is used for high bandwidth and medium half bandwidth applications.

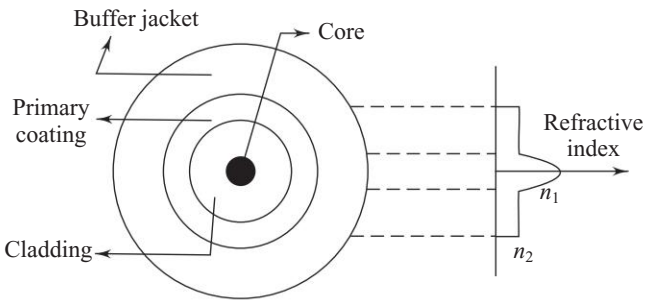


Fig. 12.13 *Wave guide structure—multimode graded index*

The difference between step index and graded index fibers is shown in Table 12.3 for comparison.

Table 12.3 *Difference between step index and graded index fibers*

Step index fiber	*Graded index fiber*
The refractive index of core, cladding and air varies in a step-by-step manner.	The refractive index of core, cladding and air varies with radial distance.
The refractive index difference at the core–cladding interface is larger.	The refractive index difference at the core–cladding interface is smaller.

The McGraw·Hill Companies

It is used to transmit single and multimode signals.	It is mainly used for multimode signals.
Intermodal dispersion is larger.	Intermodal dispersion is less due to variation in the refractive indices with radial distance.
For multimode transmission, the bandwidth is lower due to low numerical aperture.	For multimode transmission, the bandwidth is higher due to high numerical aperture.
Light rays are propagated as meridional rays since during propagation; the light ray crosses the axis of the core.	Light rays are propagated as skew rays since during propagation; the light ray never passes through the axis of the core.
Light rays are propagated into the core in a zigzag manner.	Light ray are propagated into the core either spherical or helical
It is used for long-distance transmission applications.	It is used for short distance transmission applications.

12.7 FABRICATION OF OPTICAL FIBERS

Optical fibers are prepared employing glass and plastic materials. Glass fibers are prepared from SiO_2 by doping it with other metal oxides like GeO_2, PbO, B_2O_3, etc. On the other hand, plastic fibers are prepared using polystyrene and methylmethacrylate respectively as core and cladding materials. The following are the different methods used for the fabrication of optical fibers:

(i) Modified chemical vapour deposition (MCVD)

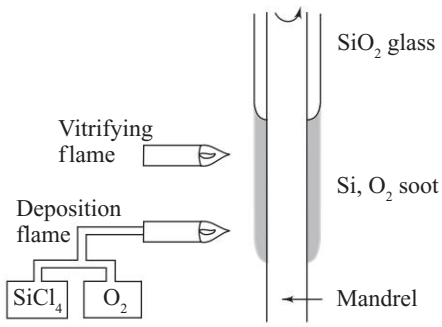
(ii) Fiber pulling from preform, and

(iii) Double crucible

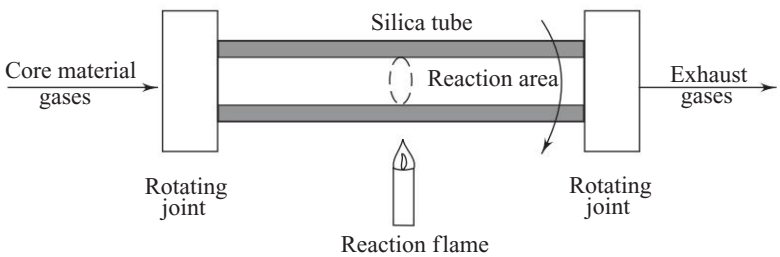
The processing of optical fibers employing the above methods is discussed briefly under the following headings:

12.7.1 Modified Chemical Vapour Deposition (MCVD)

Modified chemical vapour deposition is a method used to produce the glass preform, which is known as *final fiber.* The refractive index profile of the preform is same as that of the fiber. The experimental MCVD consists of two different stages, namely, the formation of silica cladding and then the deposition of the core glasses. The experimental set-up used for the two stages is shown in Fig.12.14. In the first step, SiO_2 is used as a material to produce the SiO_2 tube or substrate which is known as the final fiber. In order to get a difference in refractive index between the core and cladding, a dropout material, known as *soot*, is added. The tube or substrate is obtained by vapour deposition of SiO_2 and soot on a rotating mandrel or ceramic former. When the temperature reaches 1673 K, the soot vitrifies and thus forms clear glass.



(a) *Formation of silica cladding tube*



(b) *Deposition of core glasses*

Fig. 12.14 *MCVD—Schematic diagram*

After forming the clear tube, the SiO_2 tube is flirted with this tube. The core material which is in the form of glass is passed into the tube as shown in Fig. 12.14. The core particles are sintered to obtain a clear glass. The required heating is given to the tube so as to obtain the glass preform. which is then used to obtain the optical fiber. The process of drawing optical fibers from the glass preform is discussed in the following method.

12.7.2 Fiber Pulling from a Preform

In this method, optical fibers are obtained by pulling the glass preform and coating with a suitable material. The experimental set-up used for fiber pulling is shown in Fig.12.15. The fiber drawing furnace, a gauge controller, casting funnel, curing furnace and take-up drum are arranged in a single line as shown in Fig.12.15. The glass preform is fixed at the top of the pulling tower using a clamp. The top of the glass preform is softened using the circular drawing furnace. The filament drawn from the top of the glass preform is fixed at the take-up drum. When the take-up drum rotates, it draws the fiber from the top of the glass preform. Thus, the rotation of the take-up drum helps to control the diameter of the fiber. Further, the diameter of the fiber is controlled using a non-contact thickness or diameter gauge. The gauge controller in-turn controls the rotation of the take-up drum through a feedback loop.

The McGraw·Hill Companies

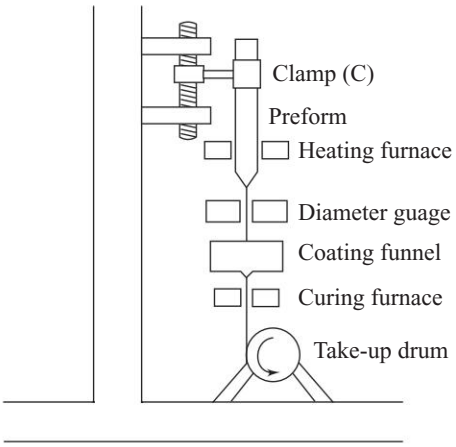


Fig. 12.15 *Schematic diagram of a fiber pulling tower*

The diameter controlled fiber is then passed through the coating funnel which consists of plastic coating material. The plastic material which is coated on the surface of the fiber is converted into solid by passing through the curing furnace. The plastic coating on the fiber prevents damages due to impurities and structural changes.

Advantages

The following are the advantages of this method:

- It is possible to produce several kilometres of 125 μm diameter fibers.
- It is easy to control the diameter of the fiber through non-contact method.

Disadvantages

The following are the diadvantages of this method:

- It is not possible to produce a continuous fiber.

12.7.3 Double Crucible

In the double-crucible or direct melt technique, the fiber is drawn directly from the melt of the core and cladding materials. This method is used to produce fiber continuously at low cost. Thus, it overcomes the drawback of the fiber pulling from a glass preform.

The experimental arrangement used in the double-crucible method is shown in Fig.12.18. It consists of two crucibles, namely, inner and outer crucibles. The core material is taken in the inner crucible, while the cladding material is taken in the outer crucible. The materials in the respective crucibles are melted by keeping the crucibles in the furnace. The fiber is drawn from the melts using the same procedure as that performed in the preform method. One can obtain the fiber continuously by feeding the high purity core and cladding materials to the respective crucibles. Platinum crucibles are used to avoid contamination double crucible method is more suitable in producing plastic step index fibers.

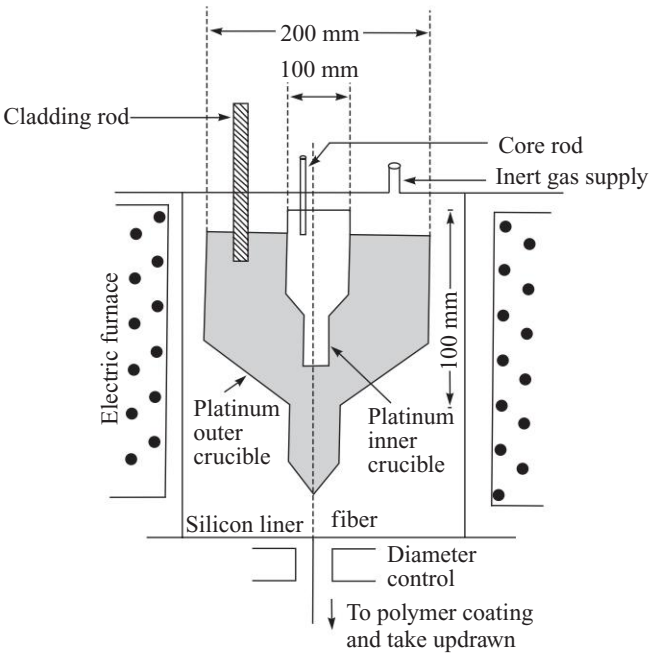


Fig. 12.16 *Double-crucible technique*

Advantages

The following are the advantages of the double-crucible technique:

- It is used to produce fiber at low cost.
- It is possible to produce continuous fiber.
- Contaminations are avoided by using platinum crucible.

12.8 LOSS IN OPTICAL FIBERS

The selection of the right fiber for the transmission of optical signals requires the transmission characteristics of optical fibers. The transmission characteristics are the factors which affect the performance of optical fibers during the transmission of signals. The following are the factors which affect the transmission of light waves through optical fibers:

i) Attenuation

ii) Dispersion, and

iii) Bending

12.8.1 Attenuation

The loss of energy during transmission of signals through an optical fiber is known as *attenuation*. When the optical signal is transmitted into the fiber, energy is lost mainly due to electron absorption, Rayleigh scattering, material absorption and impurity

The McGraw·Hill Companies

absorption. The losses of energy due to the first three absorptions are known as *intrinsic absorption mechanism*. On the other hand, the loss of energy due to impurity is known as *extrinsic absorption mechanism.* Intrinsic absorption depends on the glass fiber, while extrinsic absorption depends on impurity.

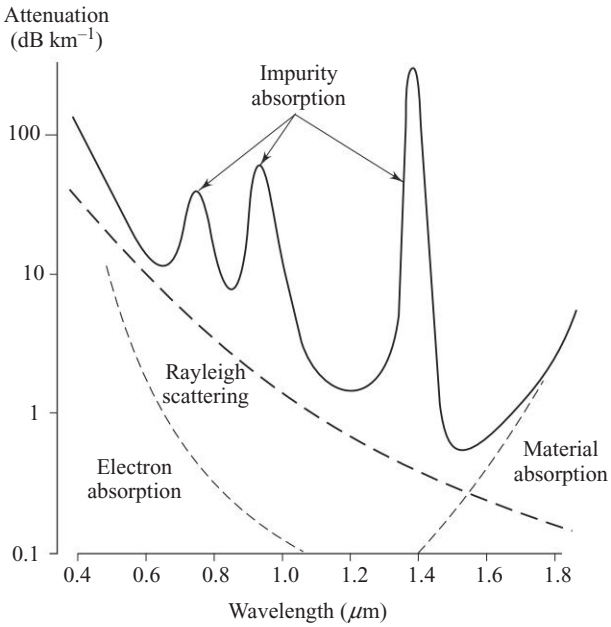


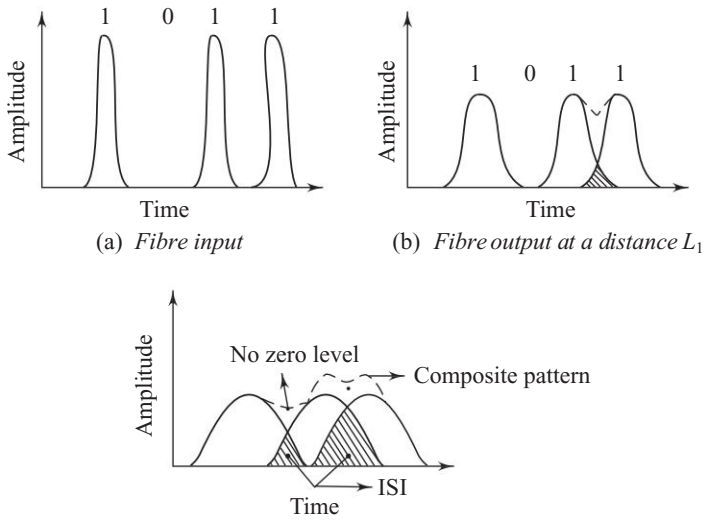
Fig. 12.17 *Attenuation as a function of wavelength of a silica-based glass fiber*

The attenuation of optical signals in a fiber due to the above mechanism is shown in Fig. 12.17. It is clear that the attenuation due to impurity is higher than core/cladding materials. In intrinsic absorption, the scattering of light due to small irregularities in the core structure leads to Rayleigh scattering. The irregularities in the core are due to fluctuation in density which arises during the manufacturing of the fiber. The Rayleigh scattering is more prominent only when the wavelength of the incident light and dimension of the scattering mechanisms are of the same order. One can reduce the loss due to Rayleigh scattering, by operating the fiber at a longer wavelength. When longer-wavelength operation is used, the atomistic bonds associated with core material absorb energy. The loss of energy due to core materials is known as *material absorption.* Therefore, it is good to operate at wavelengths lower than 1.55 μm to avoid significant loss in attenuation. The atom in the core material absorbs energy from the incident light, and hence, it is excited to a higher energy state. This absorption is known as *electron absorption*. The electron absorption takes place only in the lower wavelength region and is shown in Fig. 12.17.

12.8.2 Dispersion

When the light is transmitted through the optical film, broadening of the light pulse takes as it travels a longer distance. The broadening of the light pulse during the long

transmission results loss of energy. This is known as *dispersion*, shown in Fig. 12.18a. Let 1011 be the digital input pulse which is shown in Fig.12.18b. When the light pulse is transmitted through the fiber, pulse broadening and overlapping take place and hence, the output pulse is indistinguishable, as shown in Fig. 12.18c. It is known as Inter Symbol Interface (ISI) and is more effective in digital communications than analog communications.



(a) *Fibre input*

(b) *Fibre output at a distance* L_1

(c) *Fibre output at a distance* $L_1 > L_2$

Fig. 12.18 *Brodening of light pulses*

12.8.3 Bending

When an optical fiber is bent, radiative losses occur which are known as *bending losses*. Bending losses are classified into two types, namely, microbending and macrobending losses.

(i) Microbending Losses The radiative losses which arise when the fiber axis is bent during the cabling of the fiber are known as microbending losses. During the bending of the optic axis, the light rays are propagated into the cladding and hence, lost due to radiation. The losses are mainly due to winding of the fiber cable over spools or the pressure entered on the fiber by the adjacent fiber cables. The existence of thermal contraction between the cladding and core materials also leads to a small amount of loss. One can overcome the above losses by careful installation and manufacturing.

(ii) Macrobending Losses A small distribution or imperfection in the fiber cable leads to losses known as macrobending losses. The improper cabling of the fiber may lead to high radius bends which are larger than the fiber diameter. Similarly, an external force also causes a small bending in the fiber due to deformation of the cable jacket. The above bends lead to loss of light radiation and hence, are known as macro bending losses.

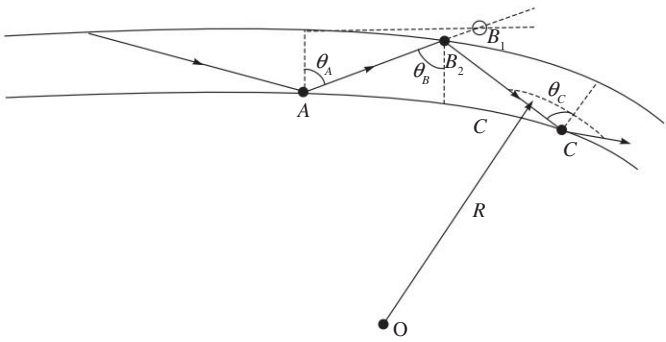


Fig. 12.19 *Propagation of light in bending fiber*

The schematic representation of bending losses is shown in Fig. 12.19. Consider that a ray *AB* reaches the point *B* at an angle ϕ_A as guided waves. If there is no bend in the fiber, it reaches the point *D*, and undergoes total internal reflection within the core. On the other hand, the ray reaches the core cladding interface at the point *C* at an angle ϕ_B which is smaller than the critical angle ϕ_C. Hence, the ray does not undergo total internal reflection. Hence, the ray is lost at the point *C*.

One can determine the optical power loss by knowing the bend radius *R* and critical radius R_c. ϕ_B Optical power loss

$$\alpha \exp(-R/R_c) \tag{12.20}$$

where $R_c \approx a/(NA)^2$, and the radius of the core.

Equation.(12.20) shows that the optical power loss depends on the radius of the bends. Thus, one should take care to avoid even small microbends while bending and cabling the fibers.

12.9. FIBER OPTICAL COMMUNICATION

Fiber optical communication systems consist of similar components like any other communication systems. They consist of an information source, transmitter, transmission medium i.e., optical fibers and receiver as shown in Fig.12.20.

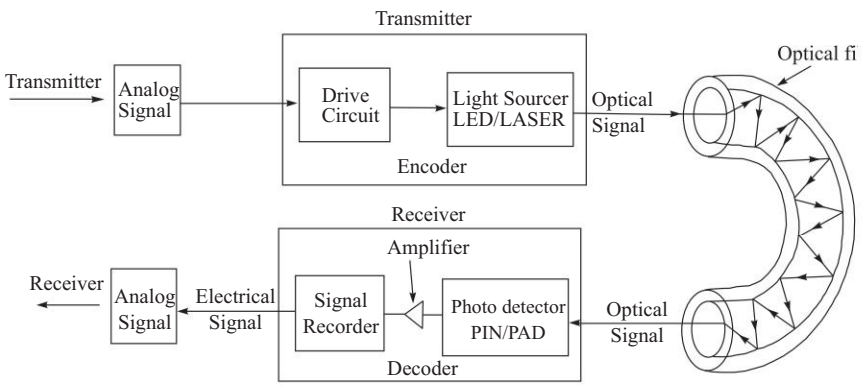


Fig. 12.20 *Block diagram of general communication system*

Information Signal Source The signals available at the information source are either voice or music or video signals. The original available signal at the signal source is analog and it is communicated to the optical systems. In order to communicate through optical fiber, the analog signals are converted into electrical signals. The converted electrical signals are passed through the transmitter.

Transmitter The transmitter is a modulator device used to receive the energy from the information source, and then modulate the electrical signal into a digital pulse for propagation into an optical fiber. The modulator consists of a driver and a light source as shown in Fig. 12.20. The driver receives the analog electrical signals and then converts then into digital pulses. These digital pulses are converted into optical pulses after passing through a light source. The light source, either light emitting diodes (LEDs) or a semiconductor laser, is used for the conversion of electrical pulses into optical pulses. The optical pulses are then focused into the optical fiber as shown in Fig.12.20.

Optical Fibers (or) Transmission Medium The optical fiber is used as a transmission medium to propagate the received optical signals to the other end. The optical signal is transmitted using the principle of total internal reflection.

Receiver The receiver is a demodulator device used to receive the optical pulses from the optical fiber and then convert into electrical pulses. The demodulator consists of photodetector, an amplifier and a signal resorder. The optical pulses which are emerging from the optical fiber are received by the photodetector. The photodetector converts the digital optical pulses into electrical pulses. The electrical signals are then amplified by the amplifier and the amplified analog electrical signals are converted into digital form. The optical analog signal contains the same information available at the information source. A single optical fiber is used to transmit several optical signals using a device known as *multiplexer*. The optical fibers are used to communicate information from one place to another place without any loss of information on the signals. The following are the advantages of optical fibers in communication.

Advantages of Optical Fibers in Communication

(i) A large number of telephone signals, nearly 15,000, can be passed through optical fibers at a particular time without any interference; whereas in an ordinary copper cable, only 48 signals are passed at a particular time.

(ii) The signal leakage is nil due to the total internal reflection, and hence, there is no cross-talk.

(iii) It is one of the most ideal means for communication in explosive environments.

(iv) It gives foolproof communication during wartime.

(v) The cost of these cables is very low when compared to metal cables.

(vi) Optical fibers have immunity to adverse temperature, moisture and chemical reactions

12.10 SPLICING

In optical communication, the fiber link is one of the most important factors to be considered. The splicing links between the two fibers are made in two ways, permanent

The McGraw·Hill Companies

bond or an easily demountable connection. The process of making permanent links between the two fibers is known as *splice.* On the other hand, the processes of making demountable links between fibers are known as *connectors.* Thus, the permanent links between the fibers are made using splice for making long optical cables. The three different methods used for optical fiber splicing are fusion, V-groove mechanical and elastic-tube splicing.

(i) Fusion Splice In fusion splicing, the connections are made by thermal bonding. The schematic representation of this method is shown in Fig. 12.21. The ends of the fibers are first prealigned and later butted together. The above process is done either through grooved fiber holder or under viewing through a microscope positions.

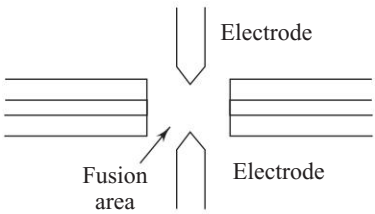


Fig. 12.21 *Fusion splicer*

The two ends of the fiber are positioned correctly and then heated with an electron arc or a laser pulse. Thereafter, the fiber ends are melted and hence fused together. The joint is then verified through a microscope. The loss due to fusion of fibers by splice method is less than 0.06 dB.

(ii) V-groove Mechanical Splice In this method, a V shaped channel is grooved using materials like silicon, plastic, ceramic or metal substrate. The fiber ends are first butted together employing the V-shaped groove as shown in Fig. 12.22.

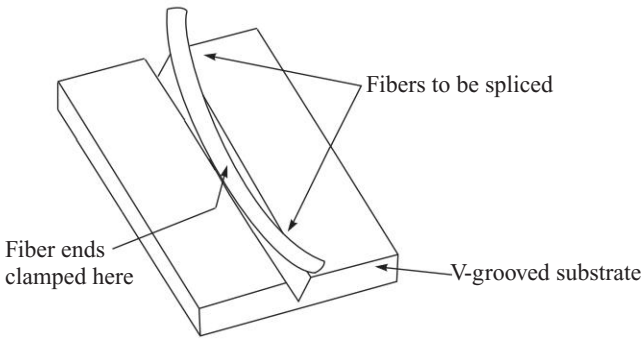


Fig. 12.22 *Splicing – V – groove mechanical method*

The two ends are bonded either using an adhesive material or cover plate. The loss in this method purely depends on the fiber size and eccentricity of the fibers.

(iii) Elastic-tube Splice It is a unique method compressed to the other two methods. In this method, the lateral, longitudinal and angular alignment ends are made

automatically. The experimental arrangement used for elastictube splice is shown in Fig. 12.23.

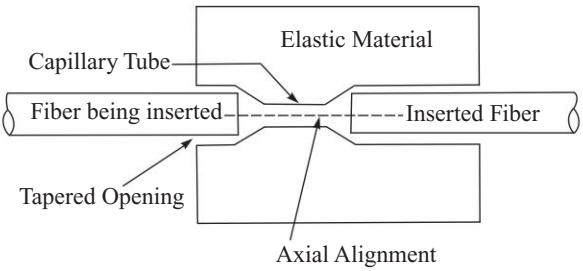


Fig. 12.23 *Splicing—elastic tube*

It consists of a tube made of an elastic material. The central part of the elastic tube splice is in the form of a capillary tube while the two ends of the tubes are tapered so as to insert the fiber in an easy way. When the fiber ends are inserted into the tube from either ends, the capillary-tube diameter expands. Thus, in turn, the elastic material exerts a symmetrical force on the fiber. An automatic alignment of the axes of the two fibers is achieved due to the symmetry features of the elastic materials. The loss which arises in this method is same as that of the fusion splice method. However, the elastic tube splicing method is used to splicing of varying diameters of optical fibers.

12.11 LIGHT SOURCES FOR FIBER OPTICS

The light source is one of the most important devices used in optical communication. It is used as a signal emitter, i.e., as a transmitter in communication systems. Generally, the light source for fiber-optic communication requires certain characteristic features and hence, an ordinary light source cannot be used for optical communication. The following are the salient features required for a light source used in optical communications:

- The emitted light should be highly monochromatic and coherent. Thus, a narrow bandwidth will reduce the dispersion, and hence improve the transmission band width.
- The intensity of the light should be high which will help it to travel longer distances.
- The light source should be small, compact and easily modulated with the fiber.
- The light source should be highly dispenable and of low cost.
- A good compactability of the emission characteristics of the light source and optical fibers should exist

The light sources which satisfies the above conditions are the Light Emitting Diode (LED) and Injection Laser Diode (ILD). The above light sources work based on the principle of electroluminescence, i.e., conversion of electrical energy into light energy. The details of the light emitting and injection laser diodes are given below in brief.

The McGraw·Hill Companies

12.11.1 Light Emitting Diode

A Light Emitting Diode (LED) is a junction diode which emits light when it is forward biased. The principle behind LED is electroluminescence. The process of injecting electrons and holes respectively into the *n*-type and *p*-type materials is known as *injection electroluminescence.*

Principle Electrons are charge carriers in a semiconductor, which absorbs energy when an electric current is applied. The total energy of the charge carrier namely electrons increases and hence, they are excited to the higher energy state (E_2). The excited electrons in the higher states stay only for few seconds. The electrons in their higher energy states after the mean lifetime eventually return back to the ground state energy level (E_1). During this process, a spontaneous emission of the radiation of light takes place. The energy of the emitted photon ($h\nu$) is equal to the energy of band gap (E_g) of materials, i.e.,

$$E_g = h\,\nu \tag{12.21}$$

where h is the Planck's constant and ν, the frequency of emitted radiation.

Substituting the value of $\nu = \frac{c}{\lambda}$ in Eq. (12.21), we get,

$$E_g = \frac{hc}{\lambda} \tag{12.22}$$

where c is the velocity of light and λ, the wavelength of light.
Equation. (12.22) can be written as,

$$\lambda = \frac{hc}{E_g} \tag{12.23}$$

Equation. (12.23) gives the wavelength of the emitted photon.
Substituting the values of $h = 6.62 \times 10^{-34}$ j s and $c = 2.998 \times 10^8$ m s^{-1} in Eqn. (12.23),

we get,

The wavelength of the emitted photon $\lambda = \frac{6.62\times10^{-34} \times 2.998\times10^{8}}{E_g}$

$$\lambda = \frac{12400}{E_g(eV)}\,\text{A}^\circ \tag{12.24}$$

Equation. (12.24) indicates that the wavelength of emitted photons depends on the energy gap in the semiconductor. Thus, the energy gap of the semiconductor plays a major role in selecting a suitable material for the LED applications. The type of material is also an important factor to be considered for the material selection.

Operation LED is a forward biased *p-n* junction (Fig. 12.24). When it is forward biased, suitably, it emits visible light. During the forward biasing, the charge carriers, namely, electron and holes, are injected respectively into the anode and cathode regions. The recombination of the charge carriers, namely, the electron from the *n*-side and the holes from the *p*-side will take place at the junction.

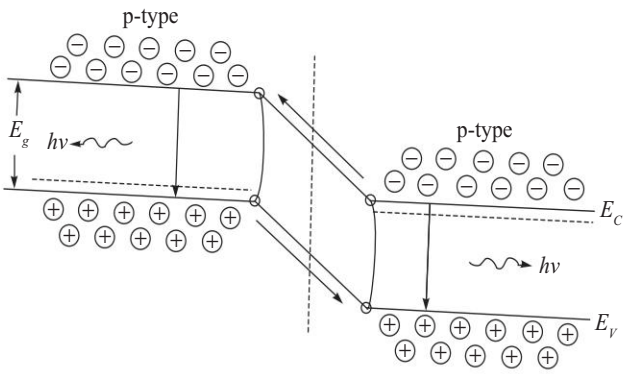


Fig. 12.24 *p–n junction spontaneous emission*

During the recombination, the difference in the energy is given up in the form of heat and light radiation, i.e., photons. The energy of light radiation depends on the strength of recombination. Thus, the diode current controls electroluminous efficiency of the LED. The emitted light is very small in intensity and is of the order of microampere range. The emitted light colour depends on the types of materials used. For example, materials like GaAs, GaP and GaAsP are used to produce infrared, red or green and red or yellow colours. The general characteristic properties of common LED materials are given in Table 12.4.

Table 12.4 *Properties of LED materials*

Materials	*Band gap Type*	*Band gap (eV)*	*Peak wave length (nm)*	*Colour*
GaP:ZnO	Indirect	1.78	700	Red
GaAsP	Direct	1.99	650	Red
GaAsP: N	Indirect	1.95	630	Red
AlGaAs	Direct	1.91	650	Red
AlGaP	Direct	2.00	620	Red
AlInGaP	Direct	2.08	595	Amber
GaAsP:N	Indirect	2.10	585	Yellow
AlInGaP	Direct	2.18	570	Yellow – Green
GaP:N	Indirect	2.20	565	Yellow – Green
GaP	Indirect	2.26	555	Green
SiC	Indirect	2.9 – 3.05	480	Blue
GaN	Direct	3.5	450	Blue

The McGraw·Hill Companies

12.11.2 Surface Emitting LED

The surface emitting LEDs (S-LED) are widely used in optical communication more than normal LEDs. The schematic diagram for a high radiative surface emitting LED is shown in Fig. 12.25.The major difference between the surface emitting LED and junction diode is the relatively large plane surface on the two sides and a narrow edge on the other side of the S-LED. Therefore, the emission of light takes place from the upper surface and hence, it is known as surface emission LED (S-LED).

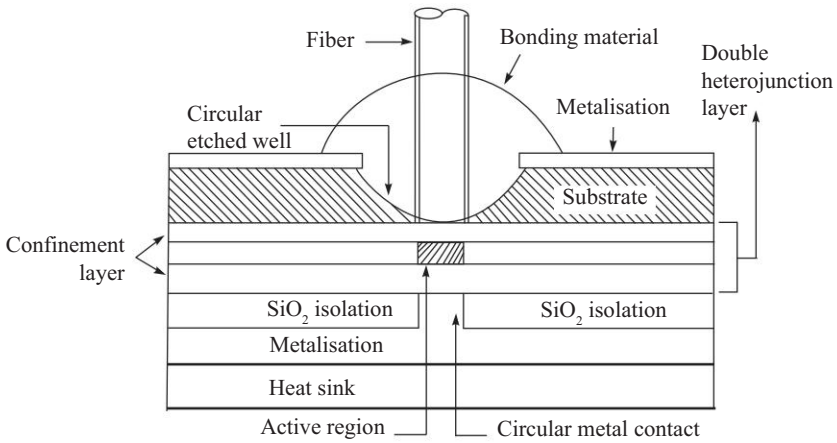


Fig. 12.25 *S-LED for optical communication*

The top surface of the S-LED is etched to form a well. Thus, it facilitates to improve the light launching capability of the diode. The diameter of the light emitting area of the S-LED is of the order of 20–50 μm.

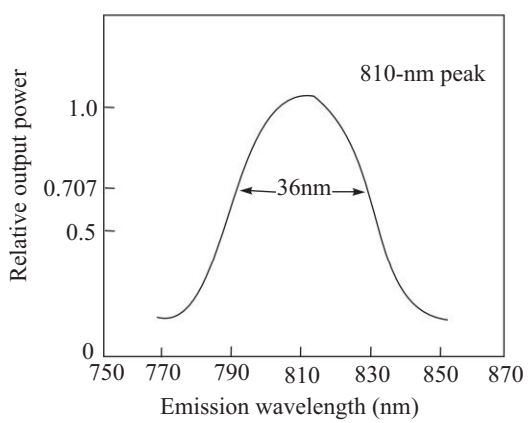


Fig. 12.26 *Emission spectra of a GaAlAs LED*

The smaller light emitting area of the S-LED facilitates to couple with an optical fiber with a diameter higher than 100 μm. The emission spectra of a GaAlAs LED shows a peak output power at 810 nm, as shown in Fig. 12.26.The spectral width of the S-LED is 36 nm which is slightly broader for optical applications.

12.11.3 Injection Laser Diode (ILD)

The injection laser diode is a semiconductor laser. It is more suitable for optical communications due to their salient characteristic properties when compared to other light sources. Properties like high monochromativity, high intensity, high power, narrow bandwidth and smaller light emitting area are the key factors for ILD. These properties help us to use ILD for many applications in optical fibers injected of a LED source.

The following are the essential features required for an ILD:

- It should give a continuous output at room and at higher temperatures.
- It should have a low threshold and high modulator bandwidth.
- It requires a highly stable and linear optical outputs.
- It requires a small light emissive area to improve the coupling.
- It requires a high coherency with narrow spectral width.

An ordinary LED can be converted into injection laser diode by providing an additional arrangement known as *optical resonant cavity*. The optical resonant cavity is used to amplify the light produced by the LED.

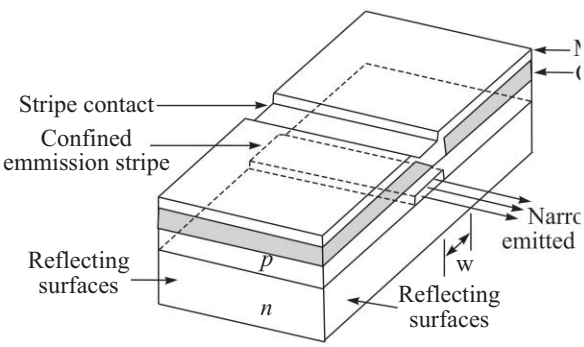


Fig. 12.27 *Injection laser diode*

The schematic representation of an ILD is shown in Fig. 12.27.The optical resonant cavity is provided in the LED by providing a space between the two highly polished ends to reflect the light. It is important to keep the dimension between the reflecting ends and the wavelength of light in the same order. This will help to achieve the multiple reflection of the light within this cavity and hence, the resonance condition. One can achieve the reflecting planes by cleavageing the LED at an appropriate plane. These cleavage planes are used as effective mirrors for light amplification to achieve the resonance condition.

Working When the diode is forward biased, the emission of photons takes place due to the recombination of electron-hole pairs. During the forward bias, the injection of

The McGraw·Hill Companies

electrons into the *p*-side and holes into the *n*-side leads to recombination of charge carriers in a laser diode. The recombination results in emission of light. The population inversion is required to achieve a high intense monochromatic laser beam. It is achieved by producing a large current density at the junction between the electrons and holes.

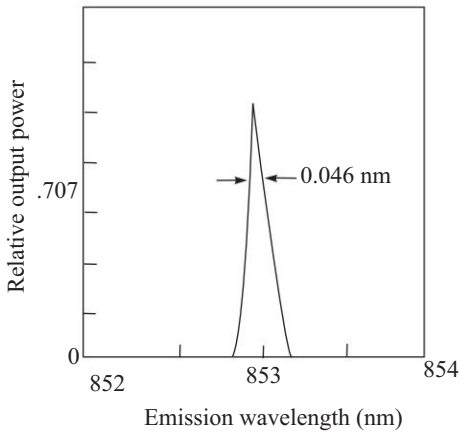


Fig. 12.28 *Light emitting spectra for a cavity laser*

The light emitted from the diode gets reflected from the cleavage plane back and then, gets reflected into the first plane, and so on. This process stimulates the recombination of more electron-hole pairs and hence, it produces an intense light radiation. The resultant radiations are highly coherent and monochromatic, and hence are more suitable for optical communications. The emission spectra produced by a simple stripe constant laser are shown in Fig.12.28. It is evident from the spectra that a peak power occurs at 853 nm with a spectral width of 0.046 nm. The observed spectral width is very narrow than that of a LED spectral width. The comparison between LED and ILD is given in Table 12.5.

Table 12.5 *Comparison between LED and ILD*

LED	*ILD*
The emitted light is a little spatial or temporal coherent.	The emitted light is highly coherent.
The power output is very less, i.e., of the order of several hundred microwatts.	The power output is very high, i.e., ofthe order of several millwatts.
The radiation level is very low.	The radiation level is very high.
It is difficult to obtain a narrow spectral width which is required for communications.	The spectral width is very narrow, and hence, more suitable for optical communications.

Due to the larger light emitting area, the light energy transmitted from LED into the fiber is less.	Due to smaller light emitting area, the light energy transmitted from the ILD into the fiber is high.
It requires a driving circuiting.	It doesn't require a driving circuiting.
It is used in many applications using multimode fibers.	It is more suitable for single-mode fiber applications.

12.12 PHOTODETECTORS

The device which is used to convert light signals into electrical signals is known as photodetector. Photodetectors are required at the receiving end of an optical communication link. When light energy is incident on the photodetector, it converts the varying optical energy easily into an electrical current. The essential requirements for a detector are high sensitivity, high reliability, high stability, short response time, low bias voltage and large electrical response. The optical detectors are of three types, namely, photo-emissive, photoconductive and photovoltaic. Generally, photo-emissive detectors are used in optical communication. In photo-emissive detectors, when a light is incident on the detector, it emits electrons.

12.12.1 Photoconductors

The principle behind the working of a photodetector is conductivity. A simple photoconductor is shown in Fig. 12.29. It is made up of semiconductor material. When a photon of energy $h\gamma$ is incident on the material, the incident photon generats electron-hole pairs. The electrons and holes are moved to the opposite charge contacts and hence, generate a photocurrent. The simplest photodetector is the photo conductor.

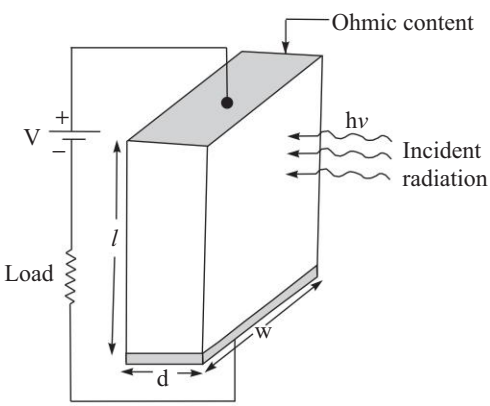


Fig. 12.29 *Photoconductor*

A photon incident on the surface, having an energy greater than or equal to the band gap E_g of the semiconductor material, will excite an electron from the valence band into the conduction band. This process generates a hole in the valence band. Thus, an electron-hole pair is generated by the optical photon. These photogenerated carriers increase the conductivity, which leads the current in the external circuit.

The McGraw·Hill Companies

12.12.2 Photodiode

A reverse-biased *p-n* junction is a photodiode. The schematic diagram of a *p-n* photodiode is shown in Fig. 12.30. It consists of a thin *p*-type layer which is deposited on an *n*-type substrate. The incident light enters through the *p*-type layer. When a reverse bias is applied to the diode, the majority carriers are attracted at the region where the *n* and *p* type region joins.

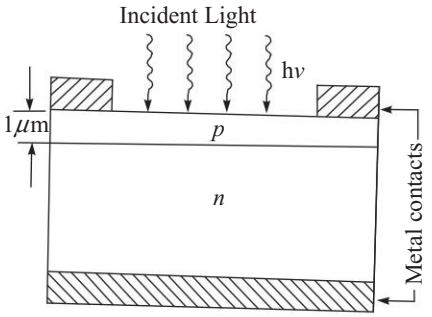


Fig. 12.30 *p-n junction—Photodiode*

Thus, it creates a thin region around the junction, known as the *depletion region*. The applied reverse-bias voltage to the diode is available as a forward bias to the minority carriers. Therefore, the minority carrier, namely, electrons in the *p*-type and holes in the *n*-type are attracted towards the junction. The motion of minority carriers forms a current known as *leakage current*. This leakage current is also known as *dark current*. The dark current in a photodiode depends on the reverse-bias voltage, ambient temperature and the series resistance in the circuit.

When a light is incident on the diode, a photon of energy $h\gamma$ is absorbed in the deflection layers. Thus, it emits an electron from the valence band to the conduction band and creates a hole in the valance band. As a result, an electron-hole pair is generated due to the optical photon. These electron-hole pairs are known as photocarriers. The photocarrier produces an electronic current in the external circuit, in addition to the already existing dark current. Therefore, an increase in the reverse-bias current takes place due to optical excitations. In order to avoid this, one can increase the width of the depletion region so as to absorb a large quantity of incident light.

The important characteristic features of a photodetector are, quantum efficiency and response speed. The number of electron-pairs generated per incident photon of energy hυ is known as *quantum efficiency and is given by*

$$\eta = \frac{\text{number of electron} - \text{hole pairs generated}}{\text{number of incident photons}} \qquad 12.25)$$

Where I_p/I_q is the number of electron-hole pairs generated and N, the number of incident photons.

Therefore, $$\eta = \frac{\left(I_p / I_q\right)}{N} \qquad (12.26)$$

The photocurrent generated per unit optical power is known as *responsibility R* and is given by

$$R = \frac{I_p}{P_0} \tag{12.27}$$

where I_p is the output photocurrent and P_0, the incident optical power.

The performance of the photodiode is evaluated based on the responsibility. For a particular wavelength, the responsibility is directly proportional to the quantum efficiency.

Drawbacks

- Many of the incident photons do not contribute to the photocurrent, since when it enters the depletion region, it passes through the n-region without creating electron-hole pairs. Hence, the responsibility or the conversion efficiency of the diode is low.
- It has a limited frequency response due to limited junction capacitance.
- The lifetime of the electron-hole pairs is very short. Thus, the response of the diode to the incident light is very fast.

12.12.3 *p-i-n* Photodiode

In order to improve the sensitivity of a photodiode, a very lightly doped intrinsic region is introduced between the *p*- and *n*- type regions in a *p-n* diode. The intrinsic region doped on *p*- and *n*-type *p-n* diode is known as *p-i-n* photo diode. The cross-sectional view of a *p-i-n* photodiode is shown in Fig.12.31.

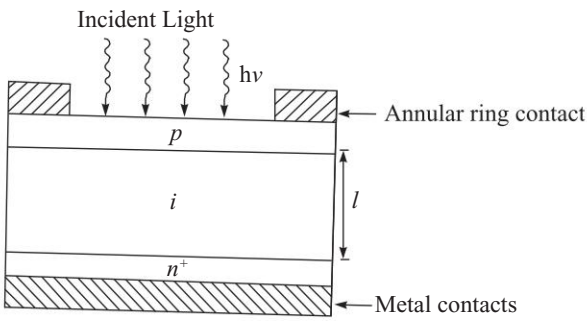


Fig. 12.31 *p-i-n photodiode*

When the energy of the incident photon is equal or greater than the band gap energy, the energy of the photon is used to excite an electron from the valence band to the conduction band. This process creates an electron-hole pair carrier. These carriers are generated in the depletion region, where in the most of the incident lights are absorbed. An increase in the width of the depletion layer region increases the absorption of photons. As a result, it generates a large photocurrent leading to increase the sensitivity of the diode. A typical energy band diagram for a *p-i-n* photodiode is Fig. 12.32.

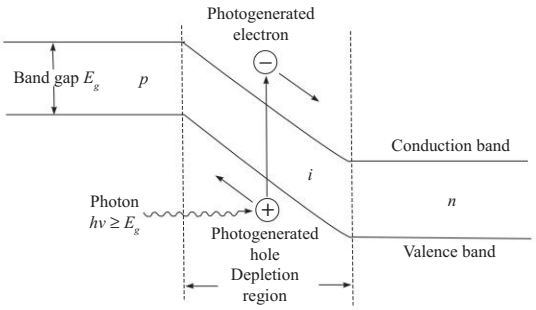


Fig. 12.32 *p-i-n photodiode—energy-band diagram*

Drawbacks

- In *p-i-n* photodiodes, the response speed is lower than that of a *p-n* photodiode.
- One can increase the quantum efficiency by increasing the thickness of the depletion region. An increase in the thickness of the deplection layer decreases the response speed. Therefore, an optimisation is required between the quantum efficiency and response speed.

12.12.4 Avalanche Photodetector

The Avalanche Photodiode (APD), a more sophisticated photodetector, overcomes the drawbacks of the *p-i-n* diode. The avalanche photodetector includes an internal gain mechanism and hence, the photoelectric current is amplified within the detector. A photodiode which is fabricated with special structures and materials is known as an avalanche photodiode. The schematic representation of the avalanche photodetector is shown in Fig. 12.33.

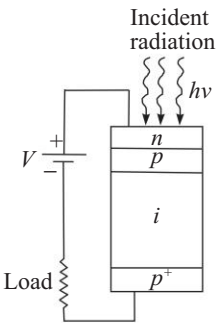


Fig. 12.33 *Avalanche photodetector*

The avalance photodetector is very useful to detect light rays of very low intensity. When the photodetector is operated at the reverse bias voltages, an increase in the voltage leads the diode to reach a threshold value. The photogenerated electrons in the depletion layer acquire sufficient energy from the internal field to liberate secondary electron-hole pairs. The process of generation of secondary electron-hole pairs is known

as *impact ionisation*. The secondary electron-hole pair produces further ionization due to drift in opposite directions along with primary carriers. In the presence of field intensity, the number of electrons generated multiplies exponentially. This phenomenon is known as *avalanche effect.*

Drawbacks

- Limitation in the operation speed is due to high transit of the carriers.
- Existence of noise in the output signal is due to the random variation in the avalanche multiplication factor.
- It is not possible to use an APD in high-intensity light optical communication.

12.13 FIBER OPTICAL SENSOR

An optical sensor is a device which converts one form of physical parameter into another form of physical parameter that can be conveniently and accurately measured. Fiber-optic sensors consist of an optical source, waveguides, modulator and optical detector. A typical fiber-optical sensor is shown in Fig 12.34.

Fig. 12.34 *Block diagram of a fiber-optic sensor*

The optical fibers are used as waveguides to guide the optical signals, and hence are known as waveguides. The optical source which produces the optical signal is coupled with the transmitting fiber. The optical signals are received by the transmitting fiber and then they are connected to the modulator. The modulator modulates the optical signals for any one of the properties such as wavelength, intensity and phase, depending on the requirements. The modulated signals are received by the receiving fiber and then connected to the detector. The modulated optical signals are detected by the optical detector. Fiber-optic sensors find wide applications in different fields when compared with normal sensors. Generally, LED or LASER is used as an optical source for sensor applications.

12.14. CLASSIFICATION OF OPTICAL SENSORS

The fiber-optic sensors are classified into two types based on modulation, namely,
(i) active or intrinsic sensors, and (ii) passive or extrinsic sensors.

12.14.1 Active or Intrinsic Sensor

In an active or intrinsic sensor, the physical parameters to be sensed act directly on the fiber with respect corresponding modulation in the transmitted physical parameter. The change into the optical propagation is measured. Simultaneous sensing and measurement are possible in active sensors. Based on modulation properties, active sensors are further classified as intensity-modulated, phase-modulated, polarisation-

The McGraw·Hill Companies

modulated and wavelength modulated sensors. The intensity-and phase-modulated sensors find wide applications than any other optical sensors.

(i) Intensity-Modulated Sensors In intensity-modulated sensors, the change in intensity of light due to absorption or emission of light or change in refractive index are measured. These sensors are used to measure physical parameters like pressure, displacement, refractive index, temperature and level of liquids.

(ii) Phase-Modulated Sensors In phase-modulated sensors, the change in phase of light signals due to the interference pattern produced between the transmitted and reference signals is measured. Sensors such as hydrophone, gyroscope, and magnetometers are based on this principle.

(iii) Polarisation-Modulated Sensors The principle behind a polaristion-modulated sensor is the polarisation state of guided signal by the variable. The application of these sensors includes measurement of magnetic field and the current measurement in high-voltage transmission lines.

(iv) Wavelength-Modulated Sensors The principle behind this sensor is the spectral dependent variation of absorption and emission by the variable. These sensors are used to measure temperature.

12.14.2 Passive or Extrinsic Sensor

In extrinsic optical fiber sensors, separate sensing elements and optical fibers are used as guiding media, i.e., to transmit radiation to and fro from the point or region to be sensed. The modulation of the light transmitted through the fiber is carried out externally either by induced or environmental changes. Thus, the modulated signal is transmitted through a second fiber. In a passive sensor, the fibers are used only as transmission media.

12.14.3 Pressure or Temperature Sensor

The pressure sensor is an example of active sensors. The principle behind the pressure sensor is interference, i.e,, the path difference between the reference beam and the beam passing through the environment. A schematic representation of a pressure sensor is shown in Fig.12.35. It consists of a laser source, beam splitter, reference fiber, test fiber and optical components.

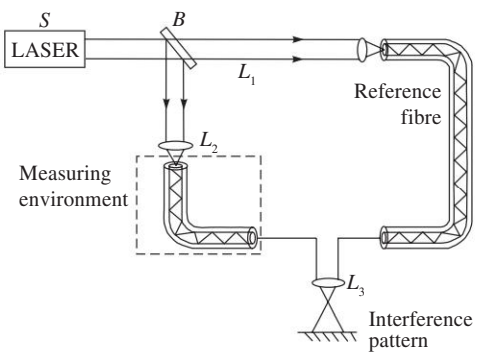


Fig.12.35 *Fiber optic sensor—pressure sensor*

High-energy laser light from the source *S* is made incident on the beam splitter *B*, which is kept at an angle of 45° to the incident light. The beam splitter splits the incident light into two components. The light beam emerging out of the beam splitter is made to pass through the reference fiber with the help of a converging lens L_1. The reference fiber is protected from environmental changes. Thus, the incident light emerges out of the reference fiber at the other end without any change in phase. The beam which is reflected from the beam splitter is made to pass through the test fiber using the converging lens L_2. The test fiber is exposed to the environment. Any change in the environment, i.e., pressure or temperature makes a corresponding change in the phase of the laser light which is passes through the test fiber.

The laser light emerging from the test fiber is made incident on the lens L_3. Therefore, a path difference between the light emerging from the reference and test fibers is observed. The observed path difference between the reference and test beam is due to the change in parameters such as pressure or temperature on the test-fiber environment. The path difference between the two beams produces an interference pattern after passing through the lens L_3. The observed interference pattern is used to measure the change in pressure or temperature accurately.

12.14.4 Displacement Sensor

The displacement sensor is an example of a passive sensor. In this method, two separate sensors are used one for transmitting the light on the moving object and the other, to receive the reflected light from the object. A schematic representation of a displacement sensor is shown in Fig 12.36.

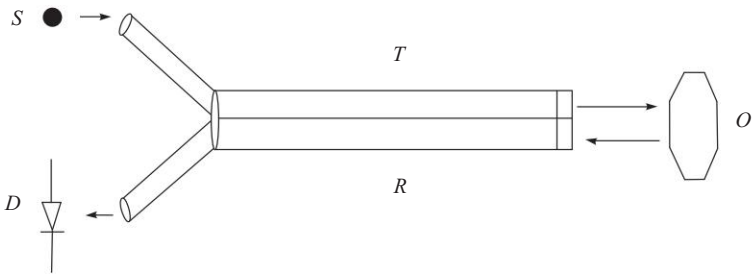


Fig. 12.36 *Fiber optic sensor—displacement sensor*

The displacement sensor consists of a light source *S*, a transmitting fiber *T*, a receiving fiber *R*, an object *O* and the detector *D*. The transmitting and receiving fibers consist of a bundle of fibers which are coupled properly to the source and detector. The light energy from the He–Ne laser source is made incident on the moving object after passing through the transmitting fiber. The light reflected from the moving object *O* is received by the receiving fiber and the same is detected at the other end by the detector. The change in the magnitude of the light detected by the detector is used to measure the movement of the object. For example, an increase in intensity of light is noticed when the object is moved towards the sensor. On the other hand, if the object is moving away from the sensor, the intensity of the light decreases. Thus, the change in intensity of the light is used to measure the displacement of the object.

The McGraw·Hill Companies

12.14.5 Advantages of a Sensor

(i) It is very sensitive and gives accurate results.

(ii) It is very small in size, volume and weight.

(iii) It is very safe in explosive environments and resistant to nuclear or ionizing radiations.

(iv) It is easy to couple with optical systems, and hence used in remote sensing applications.

(v) It is chemically inert, and hence it is used in chemical and biomedical instrumentations.

(vi) The sensed signal is immune to electromagnetic interference and radio frequency interference.

12.15 FIBER ENDOSCOPE

The principle behind the working of a fiber endoscope is the transmission of light to illuminate the inner structure of an object under study and to view or collect the image of the object. The structure of the flexible fiber endoscope is shown in Fig.12.37. It consists of bundles of flexible inner and outer fibers. The inner fiber is used to illuminate the structure of the object under study. Similarly, the outer fiber is used to collect the reflected light from the illuminated area of the object.

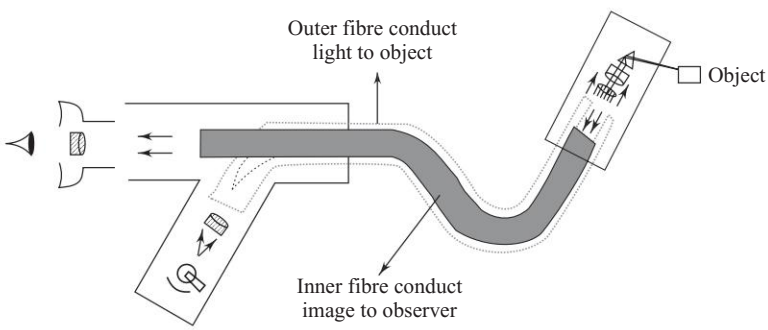


Fig. 12.37 *Flexible fiber endoscope*

The reflected light helps to view the structure of the objects to be viewed. One end of the optical fiber is focused an the object while the other end is used as a viewing end. An optical light source is attached at the viewing end of the endoscope. The light from the optical source is shone on into the object to be viewed. Then, the reflected lights from the surface of the object are viewed through the telescope attached to the end of the endoscope. The telescope is used to obtain a better image of the object.

Applications

- It is used to study the interior of the lungs and other parts of the human body.
- It is used to study tissues and blood vessels which are far below the skin.
- It is a non-destructive method.

12.16 ENGINEERING APPLICATIONS OF OPTICAL FIBERS

The following are the engineering applications of optical fibers:

(a) ***Telecommunications***

- It is used for long distance transmission or communication of sound signals.
- It is used to communicate a large number of telephone signals, more than 1,5000, without any interference; whereas an ordinary copper metal cable is used only for 48 signals at a particular time.
- During optical communication, the signal leakage is due to total internal reflection.
- Therefore, there is no cross talk in fiber optic communication.
- It is an ideal communication tool to use in explosive environments.
- It is highly immune to adverse temperature, moisture and chemical reactions.
- It is used as a secure communication medium during wartime.
- Optical cables are low in cost when compared to metal cables.

(b) ***Computer Networks***

- It is used as a communication device in Local Area Networks (LAN).
- The internal computers are connected to the server through optical cables to avoid the loss of data during transmission of signals.
- It is used to connect output devices like printers to the computer/server.
- Intra-and inter-net connectors are made through optical fibers.
- The operation of radars, digital cameras, etc., are monitored and controlled using optical fiber cables which are connected to the source computers.

(c) ***Cable Television***

- It is used to transmit the TV signals in digital form after proper modulation.
- It is used to transmit a large number of TV signals without overlapping.
- The loss of signal is very less during long transmission, and hence, it produces good quality TV signals without overlapping.

(d) ***Delay Lines***

- It is used in electronics to produce the required delay in microwave signal processing.
- It transmits the signal without any loss due to attenuation.

Keypoints to Remember

- The principle behind optical fiber communication is total internal reflection.
- According to Snell's law, the relationship between angles of incidence and refraction is

$$n_1 \sin \phi_1 = n_2 \sin \phi_2$$

The McGraw·Hill Companies

- When the angle of incidence is greater than the critical angle, the light rays are reflected within the originating medium. This phenomenon is known as total internal reflection.
- Different kinds of propagation of light rays takes place in optical fibers, namely, meridional and skew rays.
- The ray which passes through the axis of optical fibers is known as meridional rays.
- The rays which do not pass through the axis of optical fibers are known as skew rays.
- The angle of acceptance is the limiting angle for the incident rays, wherein the incident light rays undergo total internal reflection. When the angle of incidence is greater than the acceptance angle, total internal reflection will not take place.
- The inner glass cylinder of the optical fiber is known as core, which has high refractive index (n_1) than the cladding.
- The seal which covers the core is known as cladding. The refractive index of a cladding (n_2) is lower than that of the core material.
- The measurement of the amount of light rays which is accepted by the optical fiber is known as Numerical Aperture (NA).
- The number of modes propagated through an optical fiber is equal to 4.9 (d NA/λ)^2
- The condition to propagate in single mode through an optical fiber is $d \leq (0.76$ NA) λ.
- The condition to propagate in multimode through an optical fiber is $d > (0.76/$ NA) λ.
- There are two different types of fibers, namely, glass and plastic fibers.
- Glass fibers are obtained by fusing metal oxides and silica glasses.
- Plastic fibers are made of plastic materials.
- Optical fibers are classified as single and multimode fibers based on the number of modes of propagation.
- In a step index fiber, the refractive indices of air, core and cladding varies step by step.
- In a graded index fiber, the refractive index of the core varies radially from the fiber axis.
- Based on the refractive index profile, optical fibers are classified as step index and graded index fibers.
- The time difference between the entry at one end and the arrival at the other end is known as intermodal dispersion.
- The core cladding index difference is equal to $\Delta = \frac{n_1 - n_2}{n_1}$ The loss of light energy when it is propagated through an optical fiber is known as attenuation.
- The loss of optical energy when it is propagated through an optical medium due to the scattering, materials absorption and electron absorption is known as dispersion.

- A transmitter, or modulator, is used to convert the input analog electrical signals into electrical pulses and then to digital optical pulses.
- LED and ILD are used as sources in optical communication.
- The principle behind LED is electroluminescence.
- The process of injecting electrons and holes respectively into *n*- and *p*-type materials is known as electroluminescence.
- In a surface emitting diode, light emission takes place from the upper surface of the diode.
- Injection laser diode is an ordering laser diode added with an optical resonating cavity.
- Optical resonating cavity is used to amplify the intensity of light emitted by the laser diode.
- The loss in light radiation due to bending of optical fiber is known as bending losses.
- The radiative losses which arise when the fiber axis is bent during cabling is known as microbending loss.
- The *p-i-n* photodiodes and avalanche photodiodes are used as photodiodes in optical communication.
- The device which is used to convert light signals into electrical signals is known as photodetector.
- The leakage current or dark current in a photodiode is formed due to attraction of minority carriers towards the junction.
- An avalance photodetector is a photodetector which includes internal gain mechanism to amplify the photoelastic current.
- A receiver or demodulator is used to convert the digital optical pulses and then convert then into analog electrical signals.
- In optical communication, the LED and semiconductor lasers are used as optical light sources.

Solved Problems

Example 12.1 Estimate the critical angle when the core refractive index is 1.48 and the relative refractive index is 2%.

Given Data

The relative refractive index $\Delta = \dfrac{n_2 - n_1}{n_1} = 2\%$

The refractive index of core $n_1 = 1.48$

The McGraw·Hill Companies

Solution:

The relative refractive index $\Delta = \frac{n_2 - n_1}{n_1}$

$$= \frac{n_2}{n_1} - 1$$

Therefore, $\frac{n_2}{n_1} = 1\text{-}\Delta$

Substituting the value of Δ, we get,

$$= 1 - 0.02$$
$$= 0.98$$

We know that the critical angle ϕ_c is given

$$\phi_c = \sin^{-1}\frac{n_2}{n_1}$$

Substituting the value, we get,

$$= \sin^{-1}(0.98)$$
$$= 78.52^\circ$$

The critical angle of the fiber is 78.52°.

Example 12.2 A fiber has a diameter of 6 μm and its core refractive index is 1.47 and for cladding, it is 1.43. How many modes can propagate into the fiber if the wavelength of the source is 1.5 μm?.

Given Data

The diameter of fiber $d = 6 \times 10^{-6}$ m

The wavelength of laser source $\lambda = 1.5 \times 10^{-6}$ m

The refractive index of core $n_1 = 1.47$

The refractive index of clad $n_2 = 1.43$

Solution

The numerical aperture is $NA = \sqrt{n_1^2 - n_2^2}$

Substituting the values of n_1 and n_2, we get,

$$= \sqrt{1.47^2\, 1.43^2}$$
$$= 0.34$$

The number of modes propagating through the fiber

$$N = 4.9\left[\frac{d\,NA}{\lambda}\right]^2$$

Substituting the values, we get,

$$= 4.9\left[\frac{6\times10^{-6}\times0.34}{1.5\times10^{-6}}\right]^2$$

$$= 9.06$$

The number of modes that can be allowed through the fiber is 9.

Example 12.3 Find the numerical aperture of an optical fiber having a core refractive index of 1.55 and a cladding refractive index of 1.50.

Given Data

The refractive index of core $n_1 = 1.55$
The refractive index of clad $n_2 = 1.50$

Solution

The numerical aperture is $\text{NA} = \sqrt{n_1^2 - n_2^2}$

Substituting the values of n_1 and n_2, we get,

$$= \sqrt{1.55^2 - 1.50^2}$$

$$= 0.391$$

The numerical aperture of the fiber is 0.391.

Example 12.4 Find the numerical aperture of an optical fiber having a core refractive index of 1.6 and a cladding refractive index of 1.50

Given Data

The refractive index of core $n_1 = 1.6$

The refractive index of clad $n_2 = 1.5$

Solution

The numerical aperture is $\text{NA} = \sqrt{n_1^2 - n_2^2}$

Substituting the values of n_1 and n_2, we get,

$$= \sqrt{1.6^2 - 1.5^2}$$

$$= 0.55677$$

The numerical aperture of the fiber is 0.55677.

Example 12.5 A step-index fiber has a numerical aperture of 0.26, a core refractive index of 1.5 and a core diameter of 100 μm. Calculate the refractive index of the cladding, angle of acceptance and the maximum number of modes with a wavelength of 1 μm that the fiber can carry.

Given Data

The refractive index of core $n_1 = 1.5$
The numerical aperture NA = 0.26
The diameter of core in fiber $d = 100 \times 10^{-6}$ m
The wavelength of source $\lambda = 1 \times 10^{-6}$ m

Solution

We know that, the numerical aperture is $\text{NA} = \sqrt{n_1^2 - n_2^2}$

Substituting the values of n_1 and NA, we get,

$$0.26 = \sqrt{1.5^2 - n_2^2}$$

(Or)

$$0.0676 = 1.5^2 - n_2^2$$

(Or) $$n_2^2 = 2.1814$$

The refractive index of cladding is $n_2 = 1.4772$

The acceptance angle is given by $i_m = \text{Sin}^{-1}(NA)$

Substituting the values of NA, we get,

$$i_m = \text{Sin}^{-1}(0.26)$$

$$i_m = 15.07^{\circ}$$

The number of modes propagating through the fiber

$$N = 4.9\left[\frac{d \times NA}{\lambda}\right]^2$$

Substituting the values, we get,

$$= 4.9\left[\frac{100 \times 10^{-6} \times 0.26}{1 \times 10^{-6}}\right]^2$$

$$= 3312.4$$

The number of modes that can be allowed through the fiber is 3312.

Example 12.6 A signal of 100 mW is injected into a fiber. The outcoming signal from the other end is 40 mW. Find the loss in dB.

Given Data

Power injected into fiber $P_1 = 100$ mW
Power ejected into fiber $P_0 = 40$ mW

Solution

The loss in dB $= 10 \log(100/40)$

$= 3.979$ dB

The loss in dB is 3.978.

Example 12.7 Calculate the refractive indices of core and cladding materials of an optical fiber if its numerical aperture is 0.22 and relative refractive index difference is 0.012.

Given Data

The relative refractive index $\Delta = 0.012$

The Numerical aperture of fiber $NA = 0.22$

Solution

The refractive index of core $n_1 = \dfrac{NA}{\sqrt{2\Delta}}$

Substituting the values, we get,

$$= \frac{0.22}{\sqrt{2\times 0.012}}$$

$$= 1.4201$$

The refractive index of cladding $n_2 = n_1(1-\Delta)$

Substituting the values, we get,

$$= 1.4201 \times (1 - 0.012)$$

$$= 1.4031$$

The refractive index of the fiber core is 1.4201 and cladding is 1.4031.

Examples 12.8 Calculate the refractive indices of the core and gladding of a fiber from the following data. The numerical aperture is 0.027 and relative refractive index is 0.015.

Given Data

The relative refractive index $\Delta = 0.015$

The Numerical aperture of fiber $NA = 0.27$

Solution:

The refractive index of core $n_1 = \dfrac{NA}{\sqrt{2\Delta}}$

Substituting the values, we get,

$$= \frac{0.27}{\sqrt{2\times 0.015}}$$

$$= 1.558$$

The refractive index of cladding $n_2 = n_1(1-\Delta)$

Substituting the values, we get,

$$= 1.558 \times (1-0.015)$$

$$= 1.534$$

The refractive index of the fiber core is 1.558 and cladding is 1.534.

Examples 12.9 Calculate the total number of guided modes propagating in the multimode step index fiber having a diameter of 60 μm and numerical aperture of 0.25, operating at a wavelength of 2.7 μm.

Given Data

The numerical aperature $NA = 0.25$

The diameter of core in fiber $d = 60 \times 10^{-6}$ m

The operating wavelength $\lambda = 2.7 \times 10^{-6}$ m

Solution

We know that, the number of guided modes propagating through the fiber

$$N = 4.9\left[\frac{d \times NA}{\lambda}\right]^2$$

Substituting the values, we get,

$$= 4.9\left[\frac{60 \times 10^{-6} \times 0.25}{2.7 \times 10^{-6}}\right]^2$$

$$= 151.23$$

The number of guided modes that can be allowed through the fiber is 151.

Example 12.10 Calculate the total number of guided modes propagating in the multimode graded index fiber having the diameter of 50 μm and the numerical aperture of 0.2 and operating at a wavelength of 1μm.

Given Data

The numerical aperture $NA = 0.2$

The diameter of core in fiber $d = 50 \times 10^{-6}$ m

The operating wavelength $\lambda = 1 \times 10^{-6}$ m

Solution

We know that, the number of guides modes propagating through the step index fiber

$$N_{Step} = 4.9\left[\frac{d \times NA}{\lambda}\right]^2$$

Substituting the values, we get,

$$= 4.9\left[\frac{50 \times 10^{-6} \times 0.2}{1 \times 10^{-6}}\right]^2$$

$$N_{Step} = 490$$

The number of guide modes propagating through the graded index fiber is

$$N_{grad} = \frac{N_{step}}{2}$$

$$N_{grad} = 245 \text{ modes}$$

The number of modes that can be allowed through the graded index fiber is 245.

Example 12.11 Compute the numerical aperture, acceptance angle and the critical angle of the fiber having a core refractive index $n_1 = 1.50$ and the refractive index of the cladding $n_2 = 1.45$

Given Data

The refractive index of core $n_1 = 1.55$

The refractive index of clad $n_2 = 1.50$

Solution

We know that the relative refractive index $\Delta = \dfrac{n_2 - n_1}{n_1}$

Substituting the values, we get,

$$= \frac{1.5 - 1.45}{1.5}$$

$$\Delta = 0.033$$

The relative refractive index of the fiber is 0.033.

The Numerical aperture is given by, $NA = n_1\sqrt{2\Delta}$

Substituting the values, we get,

$$NA = 1.5 \times \sqrt{2 \times 0.033}$$

$$= 0.387$$

The acceptance angle is given by $i_m = \text{Sin}^{-1}(NA)$

Substituting the values of NA, we get,

$$i_m = \text{Sin}^{-1}(0.387)$$

$$i_m = 22.78^{\circ}$$

The acceptance angle of the fiber is 22.78°

We know that, the critical angle ϕ_c is given

$$\phi_c = \text{Sin}^{-1}\left(\frac{n_2}{n_1}\right)$$

Substituting the value, we get,

$$= \text{Sin}^{-1}\left(\frac{1.45}{1.5}\right)$$

$$= 75.2^{\circ}$$

The critical angle of the fiber is 75.2°.

Objectives Questions

12.1 The refractive index of the materials is equal to

a) c^2/v b) cv

c) c/v^2 d) c/v

12.2 According to Snell's law

a) $\frac{n_1}{n_2}=\frac{\sin\phi_1}{\sin\phi_2}$ b) $\frac{n_1}{n_2}=\frac{\sin\phi_2}{\sin\phi_1}$

c) $\frac{n_1}{n_2}=\sin\phi_1+\sin\phi_2$ d) $\frac{n_1}{n_2}=\frac{\sin^2\phi_1}{\sin^2\phi_2}$

12.3 Conditions to obtain the total internal reflection is that the refractive index of the core is slightly _____ than that of cladding material.

12.4 Numerical aperture of an optical fiber is equal to

a) $\frac{n_1-n_2^2}{n_1}$ b) $\frac{n_1-n_2}{n_1}$

c) $\frac{n_1+n_2}{n_1}$ d) $\frac{n_1^2+n_2^2}{n_1^2}$

12.5 The condition to propagate single mode through an optical fiber is

a) $d=\frac{0.76}{NA}\lambda$ b) $d<\frac{0.76}{NA}\lambda$

c) $d\geq\frac{0.76}{NA}\lambda$ d) $d\leq\frac{0.76}{NA}\lambda$

12.6 The number of modes passing through an optical fiber is

a) $N\approx 4.9\left(\frac{d\,NA}{\lambda}\right)$ b) $N\approx 4.9\left(\frac{d\,NA}{\lambda}\right)^{1/2}$

c) $N\approx 4.9\left(\frac{d\,NA}{\lambda}\right)^{2}$ d) $N\approx 4.9\left(\frac{d\,NA}{\lambda^2}\right)$

12.7 The condition to propagate multimode through an optical fiber is

a) $d=\frac{0.76}{NA}\lambda$ b) $d<\frac{0.76}{NA}\lambda$

c) $d>\frac{0.76}{NA}\lambda$ d) $d\leq\frac{0.76}{NA}\lambda$

The McGraw·Hill Companies

12.8 _____ is an example for glass fiber.
12.9 Single-mode fiber propagates the multimodes. (True/False)
12.10 _____ mode is propagated through the multimode fiber.
12.11 In a step index fiber, the refractive index of air, core and cladding varies step by step manner.
12.12 The core–cladding index is equal to

a) $\Delta = \frac{n_2 - n_1}{n_1}$ b) $\Delta = \frac{n_1 - n_2}{n_1} >$

c) $\Delta = \frac{n_2 + n_1}{n_1}$ d) $\Delta = \frac{n_2 + n_1}{n_2}$

12.13 In optical communications_____ is used as light source.
12.14 In optical communications _____ and _____are used as a photodiodes.

KEY:

12.1 d) 12.2 b)
12.3 higher 12.4 b)
12.5 d) 12.6 c)
12.7 c) 12.8 GeO_2 – SiO_2 – Core and SiO_2 – Cladding
12.9 False 12.10 Multi
12.11 True 12.12 b)
12.13 LED and semiconductor 12.14 *pin,* photodiode

Exercises

12.1 (i) Find the numerical aperture and the acceptance angle θ_a for a step index fiber for which $n_1 = 1.5$, $n_2 = 1.48$, and $n_0 = 1$.
(ii) Find the numerical aperture and the acceptance angle θ_a for a step index fiber for which $n_1 = 1.5$, $n_2 = 1.48$, and $n_0 = 1.5$.
12.2 The refractive index of core and gladding materials of an optical fiber are 1.54 and 1.5 respectively. Calculate the numerical aperture of the optical fiber.
12.3 A step-index fiber has a numerical aperture of 0.16, a core refractive index of 1.45 and a core diameter of 90 μm. Calculate the critical angle at the core–cladding interface.

The McGraw·Hill Companies

Short Questions

12.1 What is the basic principle of fiber optics?
12.2 Define total internal reflection.
12.3 List the conditions to be satisfied for total internal reflections.
12.4 Explain the classification of optical fibers.
12.5 Give an example for glass fiber along with the core and cladding compositions.
12.6 Define acceptance angle.
12.7 Define numerical aperture
12.8 What is meant by attenuation?
12.9 What is intrinsic absorption?
12.10 What is dispersion?
12.11 List any four applications of optical fiber
12.12 Explain how fiber is drawn from double crucibles.
12.13 What is a detector?
12.14 Mention the advantages of fiber optic sensor
12.15 Differentiate between active and passive fiber optic sensors.
12.16 What is a connector?
12.17 What is a plastic fiber?
12.18 Define step index fiber.
12.19 What is meant by a graded index fiber?
12.20 Distinguish between the single mode and multimode fiber.
12.21 Differentiate step index fiber and graded index fiber.
12.22 Define intermodal dispersion.
12.23 What is splicing?
12.24 List out different type of splicing
12.25 What is a connector?
12.26 What type of source is used for optical communications?
12.27 Why repeater is used in fiber optics?
12.28 What is the function of a multiplexer?
12.29 Mention any four salient features required for an optical light source.
12.30 What is LED?
12.31 Expand LED.
12.32 Explain injection electroluminescence.
12.33 What is the principle behind LED?
12.34 What is meant by surface emitting LED?
12.35 Explain microbending losses.
12.36 Explain macrobending losses.
12.37 Differentiate micro and macrobending loss in optical communications.
12.38 What is meant by optical resonant cavity?
12.39 What is meant by photoconduction?
12.40 Explain photoemission.
12.41 What is meant photodiode?

The McGraw·Hill Companies

12.42 Explain leakage current.
12.43 Define quantum efficiency?
12.44 Mention any four drawbacks of a *p-n* photodiode.
12.45 Explain *p-i-n* photodiode.
12.46 What are the drawbacks of *p-i-n* photo-diode?
12.47 What is meant by photodetector?
12.48 Mention any four drawbacks of an APD.
12.49 What are the advantages of optical fibers in communication over ordinary cable communication?
12.50 Sketch the block diagram of fiber optic communication system.
12.51 What is fiber endoscope?
12.52 Give the application of fiber optic sensor
12.53 What is meant by waveguide?
12.54 What is meant by light source in fiber optical communication?

Descriptive Questions

12.1 Write short notes on the following:
(a) Acceptance angle
(b) Numerical aperture
(c) Step index fiber
(d) Graded index fiber
12.2 Write an essay about the working of fiber optic communication system
12.3 Define numerical aperture and derive an expression for numerical aperture and angle of acceptance of fiber interms of refractive index of core and cladding.,
12.4 Write an essay on the basic principle, preparation and advantages of optical fibers.
12.5 Explain with basic principle, the construction and working of one type of optical fiber sensor
12.6 Describe the classification of optical fibers based on refractive index profile and propagation modes.
12.7 Give an account on fiber optic temperature sensor and fiber optic displacement sensor.
12.8 Describe the working of the displacement and Magnetic field optical fiber sensor.
12.9 Explain double crucible method of fiber manufactures?
12.10 Explain the construction and working of fiber optic medical endoscope.
12.11 Write an essay about the light source and detectors used in fiber optical communication.

Chapter

Crystal Structure

13

OBJECTIVES

- To develop knowledge in crystal structure and its properties
- To explain the fundamental terms in crystallography
- To discuss the various crystal systems according to Bravais lattices
- To discuss Miller indices in crystal planes and their applications
- To discuss the importance of packing factor in crystal structure

13.1 INTRODUCTION

Crystallography deals with the study of all possible types of crystals and determination of the actual structure of the crystalline solids by X-rays, neutron beams or electron beams. Solids are classified into two categories based on the arrangement of atoms or molecules. They are given below.

(i) Crystalline solids, and

(ii) Amorphous solids

13.1.1 Crystalline Solids

In crystalline solids, atoms are arranged in a regular manner, i.e., the atomic array is periodic. Each atom is at regular intervals along arrays in all directions of the crystal. The crystalline solids have directional properties and are also called as 'anisotropic substances'. The structure may be made up of metallic crystals or non-metallic crystals. The metallic crystals find wide application in engineering because of their strength, conductivity, reflection, etc.

Example of metallic crystals are copper, silver, aluminium, tungsten, etc.

13.1.2 Amorphous Solids (Non-metallic crystals)

In amorphous solids, the atoms or molecules are arranged randomly. The amorphous solids have no regular structure (no directional property) and hence they are known as isotropic substances. Such materials have no specific electrical property, but have only plasticity. Examples are glass, plastics and rubber.

13.2 FUNDAMENTAL TERMS OF CRYSTALLOGRAPHY

The structure of all crystals is described in terms of lattice with a group of atoms, each in a lattice point. The group is termed as basis. The basis is repeated in space to form the crystal structure. Let us now consider the various crystallographic terms in detail.

13.2.1 Lattice

A lattice is a regular and periodic arrangement of points in three dimensions. Figure 13.1 represents a two-dimensional lattice.

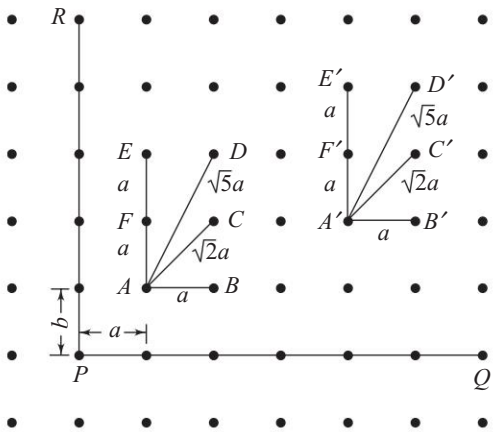


Fig. 13.1 *A two-dimensional lattice*

Consider the points P, Q and R. Let us join the points P and Q by a straight line, and the points P and R by another straight line. The line PQ is taken as an axis, say X-axis. Similarly, the line PR is taken as another axis, say Y-axis. The distance between any two successive lattice points in the X direction is taken as a. Similarly, the distance between any two successive lattice points along the Y direction is taken as b. Here, $\bar{a}$ and $\bar{b}$ are said to be lattice translational vectors. Consider a square lattice in which $a = b$.

Consider two sets of points A, B, C, D, E, F and A', B', C', D', E', F'. In these two sets, the surrounding environment looks symmetrical, i.e., the distances AB and $A'B'$, AC and $A'C'$, AD and $A'D'$, AE and $A'E'$ and AF and $A'F'$ are equal.

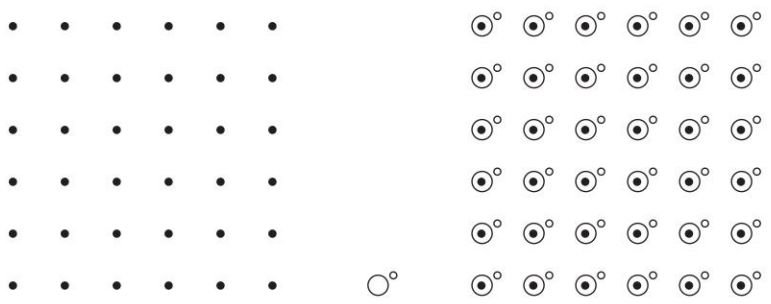
The McGraw·Hill Companies

The term lattice can be defined in another way. In an arrangement of points, if the surrounding environment looks the same when the arrangement is viewed from different lattice points, then that arrangement is said to be a lattice.

13.2.2 Basis

To construct a crystal structure, some basic arrangement is to be fixed at each and every lattice point. This basic arrangement is said to be a basis. Consider Figs 13.2a and 13.2b.

To obtain the structure shown in Fig. 13.2c, the arrangement shown in Fig. 13.2b is to be fixed in each and every lattice point. So, the arrangement shown in Fig. 13.2b is said to be a basis.



(a) Lattice (b) Basis (c) Crystal structure

Fig. 13.2 *Arrangement of basis and lattice point*

13.2.3 Crystal Structure

A crystal structure is obtained by arranging the basis in each and every lattice point. It can be written as,

Crystal structure = lattice + basis

The above equation is not a mathematical expression, but it is used to explain the formation of crystal structure. According to this equation, a crystal structure is formed by arranging the basis in each and every lattice point.

13.2.4 Unit Cell

In the construction of a wall, bricks are arranged one above the other. So, in the case of the wall, a brick is said to be a unit cell. Similarly, in the case of a crystal, a smallest unit is arranged one above the other. This smallest unit is known as unit cell. Thus, a unit cell is defined as a fundamental building block of a crystal structure. Figure 13.3 shows a cubic unit cell.

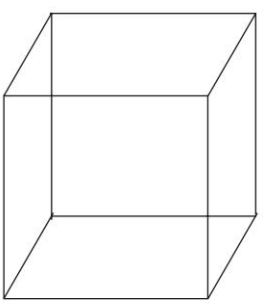


Fig. 13.3 *Unit cell*

13.2.5 Crystallographic Axes

Consider a unit cell consisting of three mutually perpendicular edges *OA*, *OB* and *OC* as shown in Fig. 13.4. Draw parallel lines along the three edges. These lines are taken as crystallographic axes and they are denoted as *X*, *Y* and *Z* axes.

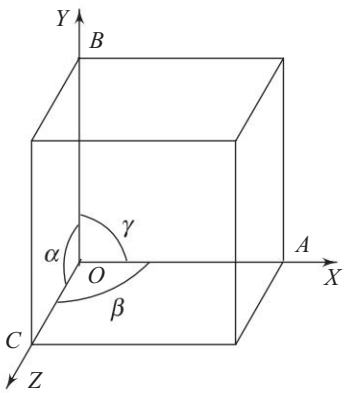


Fig. 13.4 *Crystallographic axes*

13.2.6 Primitives

Consider the unit cell shown in Fig. 13.5. Let *OA* be an intercept along the *X*-axis. Similarly, the intercepts made by the unit cell along the *Y* and *Z* axes are *OB* and *OC*, that is, *OA*, *OB*, and *OC* are the intercepts made by the unit cell along the crystallographic axes. These intercepts are known as primitives. In crystallography, the intercepts *OA*, *OB* and *OC* are represented as $\bar{a}$, $\bar{b}$ and $\bar{c}$.

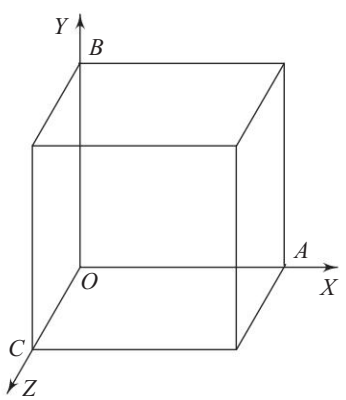


Fig. 13.5 *Primitives*

13.2.7 Interaxial Angles (or Interfacial Angles)

In a crystal, the angles between *X*, *Y* and *Z* axes are called as interaxial angles. The angle between *X* and *Y* axes is represented as γ. Similarly, the angle between *Y* and

The McGraw·Hill Companies

Z, and *Z* and *X* axes are denoted by α and β, respectively, as shown in Fig. 13.4. The angles α, β and γ are said to be interfacial angles.

13.2.8 Lattice Parameters

In order to represent lattice, the above three interfacial angles and the corresponding intercepts are essential. These six parameters are said to be lattice parameters.

13.2.9 Primitive Cell

It is the smallest unit cell in volume constructed by primitives. It consists of only one full atom as shown in Fig. 13.6.

If a unit cell consists of more than one atom then it is not a primitive cell. A simple cubic unit cell is said to be a primitive cell, whereas a body centred cubic unit cell is not a primitive cell.

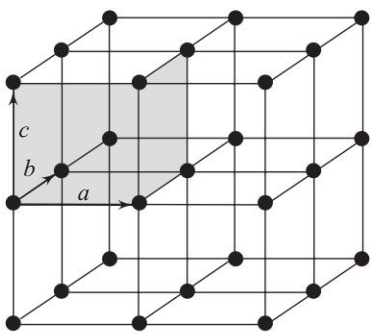


Fig. 13.6 *Primitives cell of a space lattice*

13.3 TYPES OF CRYSTALS

Crystals are classified into seven systems on the basis of the shape of the unit cell. These are classified in terms of lengths of unit cells and the angle of inclination between them. The seven systems are cubic, tetragonal, orthorhombic, monoclinic, triclinic, trigonal and hexagonal.

13.3.1 Bravais Lattices

Bravais in 1948, showed that there are 14 different types of unit cells under the seven crystal systems as shown in Fig. 13.7. They are commonly called Bravais lattices.

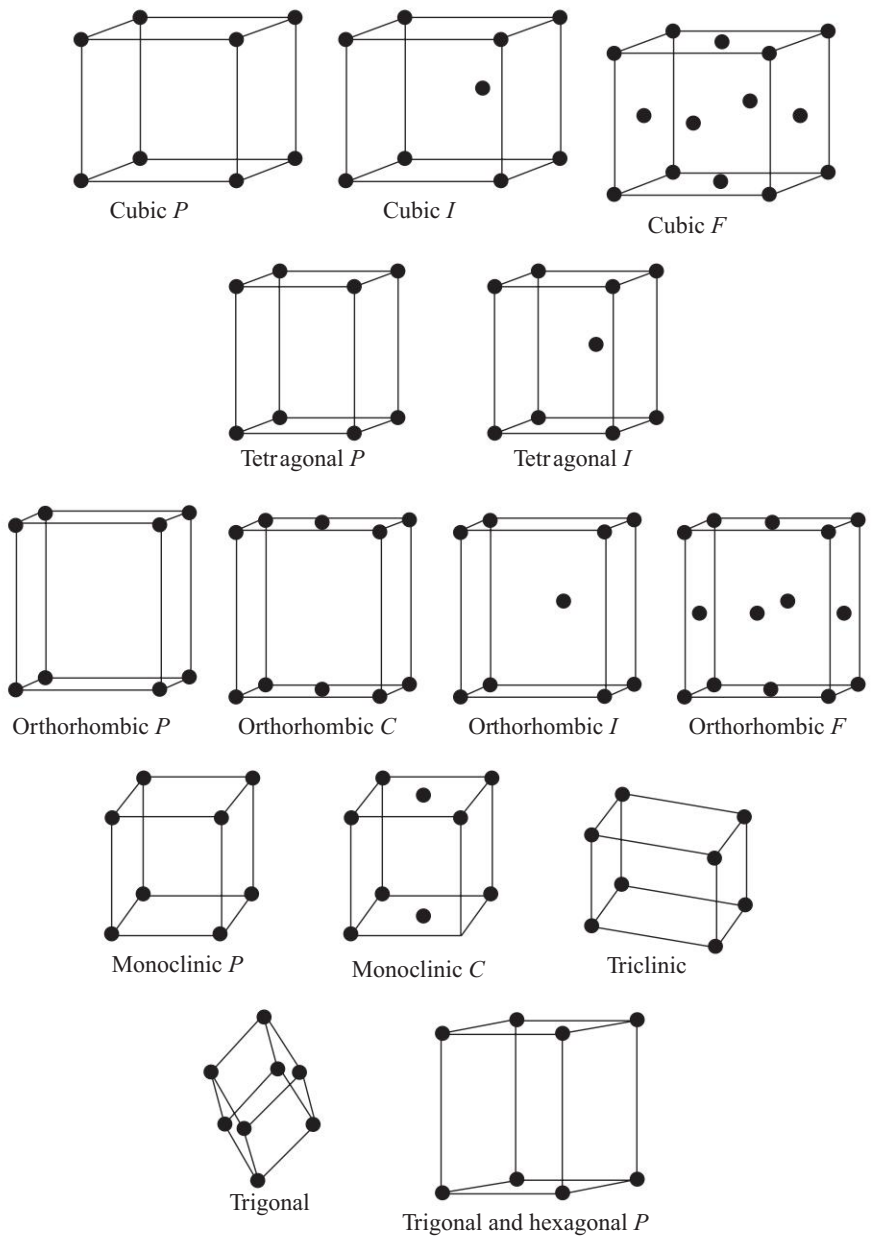


Fig. 13.7 *Bravais lattices*

Table 13.1 shows the seven crystal systems, the relation between primitives and angles, lattice symbols and the number of possible lattices. The lattice symbol *P*, *I*, *F* and *C* represents, respectively, primitive, body centred, face centred and base centred crystal structures.

The McGraw·Hill Companies

Table 13.1 *The seven crystal systems*

S. No.	*Name of the system*	*Relation between primitives and angles*	*Lattice symbols*	*No. of possible lattices*	*Examples*
1	Cubic	$a = b = c$, $\alpha = \beta = \gamma = 90°$	P I F	3	Po Na, W, α – Fe Ag, Au, Pb
2	Tetragonal	$a = b \neq c$, $\alpha = \beta = \gamma = 90°$	P I	2	TiO_2, SnO_2 KH_2PO_4
3	Orthorhombic	$a \neq b \neq c$, $\alpha = \beta = \gamma = 90°$	P I F C	 4	— $PbCO_3$, $BaSO_4$ KNO_3,K_2SO_4 α – S
4	Monoclinic	$a \neq b \neq c$, $\alpha = \beta = 90° \neq \gamma$	P C	2	$CaSO_4.2H_2O$ $K_2MgSO_4.6H_2O$
5	Triclinic	$a \neq b \neq c$, $\alpha \neq b \neq 90° \neq \gamma$	P	1	$K_2Cr_2O_7$
6	Trigonal	$a = b = c$, $\alpha = \beta = \gamma \neq 90°$ (but less than 120°)	P	1	Calcite, As,Sb,Bi
7	Hexagonal	$a = b \neq c$, $\alpha = \beta = 90°$ $\gamma = 120°$	P	1	SiO_2, AgI

13.3.2 Miller Indices

Miller devised a method to represent a crystal plane or direction. In this method, to represent a crystal plane, a set of three numbers are written within the parentheses. Similarly, crystal direction is represented as a set of three numbers written within the square brackets. Miller index is one in which the crystal plane is represented within the parenthesis.

Rules to Find the Miller Indices of a Plane To find the Miller indices for a given plane, the following steps are to be followed.

(i) The intercepts made by the plane along X, Y and Z axes are noted.
(ii) The coefficients of the intercepts are noted separately.
(iii) Inverse is to be taken.
(iv) The fractions are multiplied by a suitable number so that all the fractions become integers.
(v) Write the integers within the parentheses.

Notes

(i) While writing Miller indices, a comma or dot between any two numbers may be avoided.

(ii) The positive X axis is represented as (100), Y axis as (010) and Z axis as (001). Similarly, the negative X axis as $(\bar{1}00)$, negative Y axis as $(0\bar{1}0)$ and negative Z axis as $(00\bar{1})$.

(iii) The Miller indices for a plane (101) is read as 'one zero one' and not as one hundred and one.

For example, Miller indices of the plane shown in Fig. 13.8 can be found by the following method.

The given plane ABC makes intercepts $2a$, $3b$ and $2c$ along the X, Y and Z axes, respectively. Hence, the intercepts are $2a$, $3b$ and $2c$.

The coefficients of the intercepts are 2, 3 and 2. The inverse are 1/2, 1/3, 1/2.

The LCM is 6. Multiply the fractions by 6, so that they become integers as 3, 2, 3.

The integers are written within the parenthesis as (323). (323) represents the Miller indices of the given plane ABC.

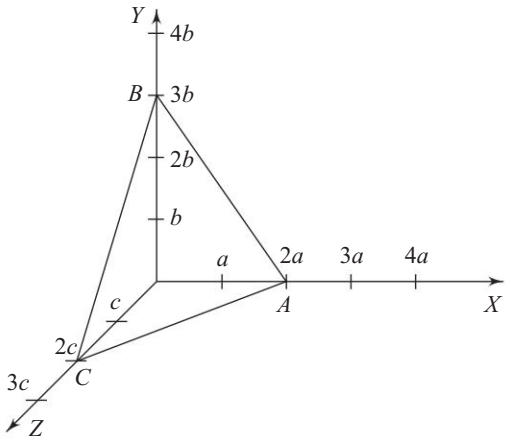


Fig. 13.8 *Miller indices of a plane*

Salient Features of the Miller Indices

(i) A plane parallel to one co-ordinate axis is taken as that plane will meet the axis at infinity. Therefore, the intercept is taken as infinity. The index number (Miller indices) for that plane in that co-ordinate axis is zero.

(ii) A plane passing through the origin is defined in terms of a parallel plane having non-zero intercepts.

(iii) Equally spaced parallel planes have same Miller indices.

(iv) Planes which have negative intercepts are represented by a bar, like $(\bar{1}\,00)$. The Miller indices $(\bar{1}\,00)$ indicating that the plane has an intercept in the negative X axis.

Important Features of Miller Indices The Miller index notation is especially useful for cubic systems. Its desirable features are listed below.

(i) The angle θ between any two crystallographic directions $[u_1\ \ v_1\ \ w_1]$ and $[u_2\ v_2\ w_2]$ can be calculated easily. The angle θ is given by Eq. (13.1),

The McGraw·Hill Companies

$$\cos\theta = \frac{u_1u_2 + v_1v_2 + w_1w_2}{(u_1^2 + v_1^2 + w_1^2)^{1/2}(u_2^2 + v_2^2 + w_2^2)^{1/2}} \tag{13.1}$$

(ii) The direction [*hkl*] is perpendicular to the plane (*hkl*)

(iii) The relation between the interplanar distance and the interatomic distance is given by Eq. (13.2),

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}} \text{ (for cubic)} \tag{13.2}$$

(iv) If (*hkl*) is the Miller indices of a crystal plane then the intercepts made by the plane with the crystallographic axes are given as *a*/*h*, *b*/*k*, *c*/*l*, where *a*, *b*, and *c* are the primitives.

Procedure to Find the Miller Indices of a Direction To find the Miller indices of a direction, choose a perpendicular plane to that direction. Find the Miller indices of that perpendicular plane. The perpendicular plane and the directions have the same Miller indices value. Therefore, the Miller indices of the perpendicular plane is written within a square bracket to represent the Miller indices of the direction.

To find the Miller indices of a line (direction), find the direction ratios of that line and then write them within the square brackets. It represents the Miller indices of that line.

The Miller indices of some of the important cubic crystal planes like (100) plane, (010) plane, (111) plane, (110) plane, (011) plane and (002) plane are shown in Fig. 13.9.

13.4 RELATION BETWEEN THE INTERPLANAR DISTANCE AND THE INTERATOMIC DISTANCE

Consider a cubic crystal with cube edge '*a*'. Let (*hkl*) be the Miller indices for the plane *ABC* as shown in Fig. 13.10. The intercepts made by the above plane are *a*/*h*, *a*/*k*, *a*/*l*. Consider another parallel plane to the above plane *ABC* passing through the origin *O*. A perpendicular line drawn from the origin *O* to the plane *ABC* represents the distance between two parallel planes. Let *ON* be the distance *d* between the two parallel planes.

Let α', β' and γ' be the angles made by the line *ON* with *X*, *Y* and *Z* axes. The direction cosines, cos α', cos β' and cos γ' are written as,

$$\cos\alpha' = \frac{ON}{OA} = \frac{d}{a/h} = \frac{dh}{a} \tag{13.3}$$

$$\cos\beta' = \frac{ON}{OB} = \frac{d}{a/k} = \frac{dk}{a} \tag{13.4}$$

$$\cos\gamma' = \frac{ON}{OC} = \frac{d}{a/l} = \frac{dl}{a} \tag{13.5}$$

The McGraw·Hill Companies

From the properties of the direction cosines of any line, we can write,

$$\cos^2 \alpha' + \cos^2 \beta' + \cos^2 \gamma' = 1$$

i.e. $$(dh/a)^2 + (dk/a)^2 + (dl/a)^2 = 1$$

$$\frac{d^2}{a^2}(h^2 + k^2 + l^2) = 1$$

or $$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}} \qquad (13.6)$$

Equation (13.6) gives the relation between the interatomic distance a and the interplanar distance d.

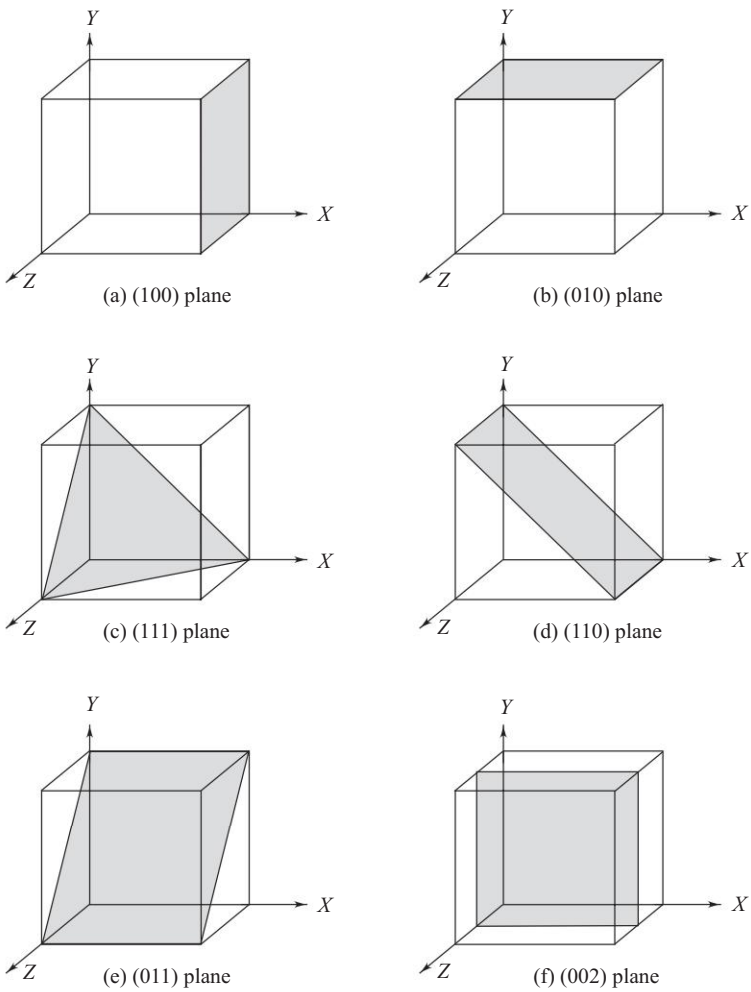


Fig. 13.9 *Miller indices of some cubic crystal planes*

The McGraw·Hill Companies

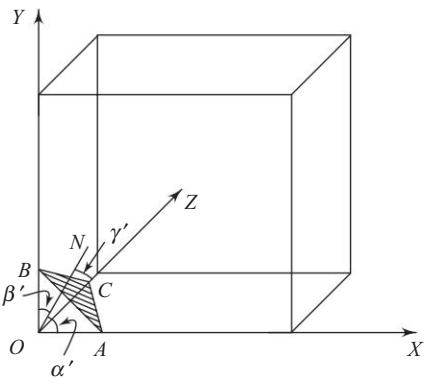


Fig. 13.10 *Spacing of planes in crystal latties*

13.5 CRYSTAL STRUCTURES OF MATERIALS

In this section, first consider some of the important parameters which are used to describe the crystal structure of the materials.

(i) Atomic Radius (r) It is half the distance between any two successive atoms. For a simple cubic unit cell, the atomic radius

$$r = \frac{a}{2} \tag{13.7}$$

where a is the interatomic distance.

(ii) Co-ordination Number It is the number of nearest neighbouring atoms to a particular atom. For a simple cubic unit cell, the coordination number is 6.

(iii) Density of Packing It is the ratio between the total volume occupied by the atoms or molecules in a unit cell and the volume of unit cell.

i.e.

$$\text{Density of packing} = \frac{\text{Total volume occupied by atoms in a unit cell}}{\text{Volume of the unit cell}}$$

$$= \frac{\text{Number of atoms present in a unit cell} \times \text{volume of one atom}}{\text{Volume of the unit cell}} \tag{13.8}$$

13.6 SIMPLE CUBIC CRYSTAL STRUCTURE

A simple cubic crystal structure (scc)unit cell consists of eight corner atoms as shown in Fig. 13.11. In actual crystals, each and every corner atom is shared by eight adjacent unit cells. Therefore, each and every corner atom contributes 1/8 of its part to one unit

cell. Hence, the total number of atoms present in a unit cell is $\frac{1}{8} \times 8 = 1$

Atomic Radius For a simple cubic unit cell, the atomic radius is given by, $r = \frac{a}{2}$

Coordination Number The coordination number of a simple cubic unit cell can be calculated as follows. Let us consider any corner atom. For this atom, there are four nearest neighbours in its own plane. There is another nearest neighbour in a plane which lies just above this atom and yet another nearest neighbour in another plane which lies just below this atom. Therefore, the total number of nearest neighbours is six and hence the coordination number is 6.

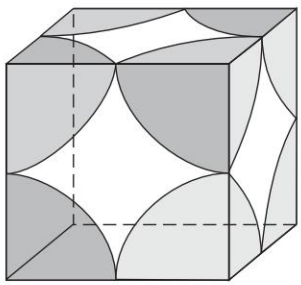


Fig. 13.11 *Arrangement of atoms in SC unit cell*

Packing Density The packing density of a simple cubic unit cell can be calculated as follows.

For simple cubic, the total number of atoms present is one. Therefore, from Eq. (13.8), the packing density of the simple cubic can be written as follows.

$$\text{Packing density} = \frac{1 \times (4/3) \times \pi r^3}{a^3} \tag{13.9}$$

Substituting the value $r = a/2$ in Eq. (13.9), we get,

$$\therefore \text{ Packing density} = \frac{1 \times (4/3) \times \pi \, (a/2)^3}{a^3}$$

$$\text{Therefore, packing density} = \frac{\pi}{6} = 0.52. \tag{13.10}$$

Thus, 52 per cent of the volume of the simple cubic unit cell is occupied by atoms and the remaining 48 per cent volume of the unit cell is vacant.

13.7 BODY CENTRED CUBIC STRUCTURE

A body centred cubic (bcc) structure has eight corner atoms and one body centred atom. The diagrammatic representation of a body centred cubic structure is shown in Fig. 13.12a. In a body centred crystal structure, the atoms touch along the diagonal of the

body. In a bcc unit cell, each and every corner atom is shared by eight adjacent unit cells. therefore, the total number of atoms contributed by the corner atoms is (1/8) 8 = 1.

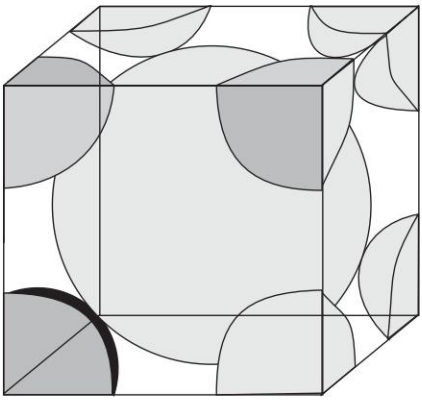


Fig. 13.12(a) *Arrangement of atoms in bcc unit cell*

A bcc unit cell has one full atom at the centre of the unit cell. Therefore, the total number of atoms present in a bcc unit cell is 2.

Atomic Radius For a body centred cubic unit cell, the atomic radius can be calculated from Fig. 13.12b as follows.

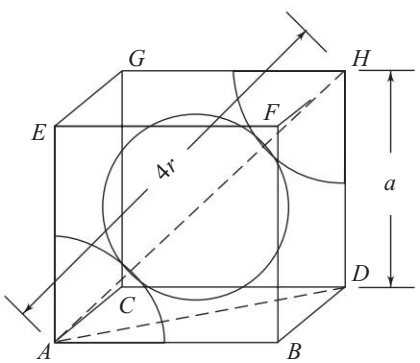


Fig. 13.12(b) *Calculation of atomic radius*

From Fig. 13.12b, $AH = 4r$ and $DH = a$
From the triangle AHD,

$$AD^2 + DH^2 = AH^2 \tag{13.11}$$

To find AD, consider the triangle ABD,
From the triangle ABD,

$$AB^2 + BD^2 = AD^2 \quad \text{or} \quad a^2 + a^2 = AD^2$$

i.e. $AD^2 = 2a^2$

Therefore, $AD = \sqrt{2}a$ (13.12)

Substituting *AD*, *AH* and *DH* values in Eq. (13.11), we get,

$$AD^2 + DH^2 = AH^2$$

or $2a^2 + a^2 = (4r)^2$

i.e. $16r^2 = 3a^2$

$$r^2 = \frac{3}{16}\, a^2$$

The atomic radius $r = \frac{\sqrt{3}}{4}\, a$ (13.13)

Coordination Number The coordination number of a body centred cubic unit cell can be calculated as follows. Let us consider a body centred atom. The nearest neighbour for a body centred atom is a corner atom. A body centred atom is surrounded by eight corner atoms. Therefore, the coordination number of a bcc unit cell is 8.

Packing Density The packing density of a body centred cubic unit cell can be calculated as follows:

The number of atoms present in a unit cell is 2.

Therefore, from Eq. (13.8), the packing density of the bcc unit cell can be written as follows.

Packing density $= \frac{2(4/3)\pi r^3}{a^3}$ (13.14)

Substituting the value of $r = (\sqrt{3}/4)\, a$ in Eq. (13.14), we get,

$\therefore$ Packing density $= \frac{2(4/3)\pi\left[\left(\frac{\sqrt{3}}{4}\right)a\right]^3}{a^3}$

$$= \frac{\sqrt{3}\pi}{8} = 0.68 \quad (13.15)$$

Therefore, the packing density, of a bcc unit cell = 0.68.

This shows that 68 per cent of the volume of the body centred cubic unit cell is occupied by atoms and the remaining 32 per cent volume of the unit cell is vacant.

The McGraw·Hill Companies

13.8 FACE CENTRED CUBIC STRUCTURE or CUBIC CLOSE PACKED STRUCTURE

A face centred cubic (fcc) unit cell consists of eight corner atoms and six face centred atoms. A face centred cubic or cubic close packed (ccp) unit cell is shown in Fig. 13.13a. The atoms in a face centred cubic unit cell touch along the face diagonal. The fcc unit cell consists of eight corner atoms and six face centred atoms. Each and every corner atom is shared by eight adjacent unit cells. Therefore, each and every corner atom contributes 1/8 of its part to one unit cell. The total number of atoms contributed by the corner atoms is (1/8) 8 = 1.

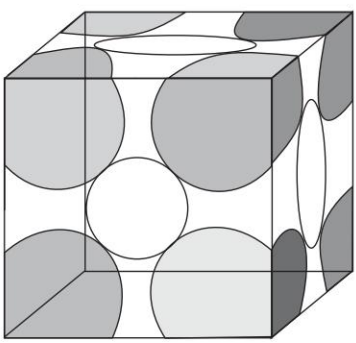


Fig. 13.13(a) *Arrangement of atoms in fcc unit cell*

Each and every face centred atom is shared by two unit cells. Therefore, a face centred atom contributes half of its part to one unit cell. The total number of atoms contributed by the face centred atom is (1/2) 6 = 3. Therefore, the total number of atoms present in an fcc unit cell is 4.

Coordination Number The coordination number can be calculated as follows. Let us consider a corner atom. In its own plane, that corner atom has four face centred atoms. These face centred atoms are its nearest neighbours. In a plane which lies just above this corner atom, it has four more face centred atoms as nearest neighbours. In a plane which lies just below this corner atom, it has yet four more face centred atoms as its nearest neighbours. Therefore, for an atom in an fcc unit cell, the number of nearest neighbours is 12.

Atomic Radius The atomic radius can be calculated from Fig. 13.13b as follows. From Fig. 13.13b, consider the triangle *ABC*,

$$AC^2 = AB^2 + BC^2$$

$$(4r)^2 = a^2 + a^2$$

$$16r^2 = 2a^2$$

The atomic radius $$r = \frac{a}{2\sqrt{2}} \quad (13.16)$$

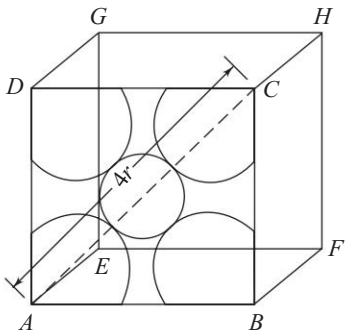


Fig. 13.13(b) *Calculation of atomic radius*

Packing Density The packing density of the fcc unit cell can be calculated as follows. The number of atoms present in an fcc unit cell is 4.

Therefore, from Eq. (13.8), the packing density of the fcc unit cell can be written as follows.

$$\text{Packing density} = \frac{4(4/3)\pi r^3}{a^3} \quad (13.17)$$

Substituting the value of $r = a/\left(2\sqrt{2}\right)$ in Eq. (13.17), we get,

$$\therefore \text{ Packing density} = \frac{4(4/3)\pi \times \left(a/2\sqrt{2}\right)^3}{a^3} = \frac{\pi}{3\sqrt{2}}$$

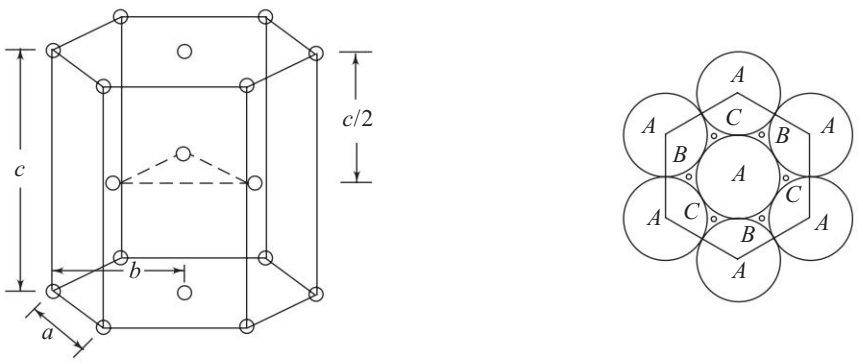
Therefore, the packing density of the fcc unit cell $= 0.74$ (13.18)

That is 74 per cent of the volume of an fcc unit cell is occupied by atoms and the remaining 26 per cent volume of the unit cell is vacant.

13.9 HEXAGONAL CLOSED PACKED STRUCTURE

The hexagonal closed packed (hcp) structure is shown in Fig. 13.14a. The hcp structure consists of three layers of atoms as shown in Fig. 13.14b. The bottom layer has six corner atoms and one face centred atom. The middle layer has three full atoms. The upper layer has six corner atoms and one face centred atom.

The McGraw·Hill Companies



(a) hcp unit cell (b) Arrangement of atoms

Fig. 13.14 *hcp structure*

The arrangement of atoms in a hexagonally closed packed structure as shown in Fig. 13.14 is explained as follows. In the bottom layer, the central atom is surrounded by six other atoms having equal radius. The bottom layer of atoms is denoted by the letter *A*. Three places are marked as *B* and the remaining three places are marked as *C* over the bottom layer. The second layer of atom is placed either over the point marked as *B* or the point marked as *C*. Consider that the second layer is placed in the places marked as *B*. The third layer can be arranged in two ways, (i) the third layer of atoms placed directly over the first layer, and (ii) the third layer placed over the points marked as *C*. The stacking sequence goes on as *ABABAB* ... for case (i) representing the hcp structure, and as *ABCABCABC* for case (ii) representing the *fcc* structure.

The total number of atoms present in the case of an hcp crystal structure is 6. To calculate this, first consider the bottom layer of atoms. The bottom layer consists of six corner atoms and one face centred atom. Each and every corner atom contributes 1/6 of its part to one unit cell. The total number of atoms contributed by the corner atoms is (1/6) 6 = 1. The face centred atom contributes 1/2 of its part to one unit cell. Therefore, the total number of atoms present in the case of the bottom layer is 1 + (1/2) = 3/2. Similarly, the upper layer also has 3/2 number of atoms. The middle layer has three full atoms. Therefore, the total number of atoms present in a unit cell is 3/2 + 3/2 + 3 = 6.

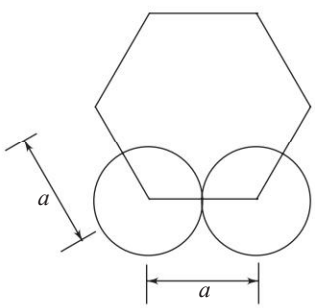


Fig. 13.15 *Calculation of atomic radius*

The McGraw·Hill Companies

Atomic Radius The atomic radius of an hcp crystal structure can be calculated as follows. Let us consider any two corner atoms. Each and every corner atom touches each other as shown in Fig. 13.15.
Therefore, from Fig. 13.15

$$a = 2r$$

i.e. $$r = \frac{a}{2} \tag{13.19}$$

The atomic radius $r = a/2$.

Coordination Number The coordination number of an hcp structure can be calculated as follows. Let us consider a face centred atom in the bottom layer. This face centred atom is surrounded by six corner atoms. These corner atoms are the nearest neighbours. The middle layer has three atoms which are nearest neighbours to the face centred atom. A unit cell, which lies below our reference unit cell also has three middle layer atoms. These three atoms are also the nearest neighbours for the face centred atom. Therefore, the total number of nearest neighbours is 6 + 3 + 3 = 12.

Packing Density The density of packing of an hcp crystal structure can be calculated as follows.

From Eq. (13.8), the packing density of the hcp unit cell can be written as follows.

Packing density $$= \frac{6(4/3)\pi r^3}{6\left(\sqrt{3}/4\right)a^2 c} \tag{13.20}$$

Substituting the value of $r = a/2$ in Eq. (13.20), we get,

$\therefore$ Packing density $$= \frac{6(4/3)\pi(a/2)^3}{6\left(\sqrt{3}/4\right)a^2 c} = \frac{2\pi a}{3\sqrt{3}c} \tag{13.21}$$

Substituting the value of $c/a = \sqrt{(8/3)}$ in Eq. (13.21), we get,

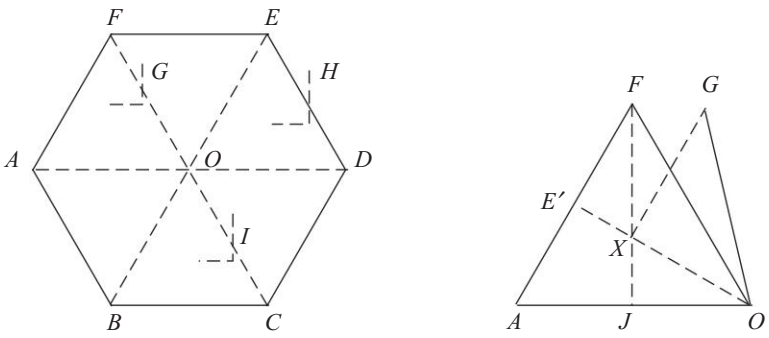
$\therefore$ Packing density $$= \frac{2\pi}{3\sqrt{3}}\sqrt{\frac{3}{8}} = \frac{\pi}{3\sqrt{2}} = 0.74 \tag{13.22}$$

Therefore, the packing density of hcp unit cell = 0.74.

That is, 74 per cent of the volume of the hcp crystal structure is occupied by atoms and the remaining 26 per cent volume is vacant.

Relation Between C and A Consider the bottom surface of a hexagonally closed packed structure. It has six corner atoms and one face centred atom. The face centred atom touches all the corner atoms. The arrangement of atoms is shown in Fig. 13.16a. The second layer of atoms are placed directly over the first layer. In Fig. 13.16a, *A*, *B*, *C*, *D*, *E*, *F* and *O* represent the bottom layer of atoms. *G*, *H*, *I* represents the middle layer of atoms. Consider the arrangement of the atoms *A*, *F*, *O* and *G* as shown in Fig. 13.16b.

The McGraw·Hill Companies



(a) Bottom surface of hcp unit cell (b) Triangle *AFO*

Fig. 13.16 *Arrangement of atoms–hcp structure*

In Fig. 13.16b, *A*, *F* and *O* represent the bottom layer of atoms and *G* represents one of the atoms in the second layer. Now, let us bisect the line *AF*, by drawing a normal from *O*. *OE′* bisects the line *AF* into two.

Consider the triangle *AE′O*. Since, the triangle *AFO* is an equilateral triangle, $\angle AOE' = 30°$. Draw another perpendicular line from *F* to the line *AO*. The line *FJ* bisects the line *AO* and hence $AJ = JO = a/2$.

From Fig. 13.16b, $\angle AOE' = 30°$ and $\angle JXO = 60°$

Consider the triangle *XJ′O*, the angles of the triangle are 30°, 60° and 90° and hence it faces are in the ratio of $1:\sqrt{3}:2$.

Therefore, in triangle *JXO*,

$$XJ : JO : OX = 1 : \sqrt{3} : 2 \tag{13.23}$$

$$OX = \frac{2JO}{\sqrt{3}} \tag{13.24}$$

Substituting the values of *JO* (= *a*/2) in Eq. (13.24), we get,

$$OX = \frac{2}{\sqrt{3}} \frac{a}{2}$$

$$OX = \frac{a}{\sqrt{3}} \tag{13.25}$$

In the triangle *XOG*,

$$OG^2 = OX^2 + GX^2 \tag{13.26}$$

Substituting *OG*, *OX* and *GX* values in Eq. (13.26), we get,

$$a^2 = \left(\frac{a}{\sqrt{3}}\right)^2 + \left(\frac{c}{2}\right)^2$$

$$a^2 - \frac{a^2}{3} = \frac{c^2}{4} \qquad \text{or} \qquad \frac{2a^2}{3} = \frac{c^2}{4}$$

$$\frac{a^2}{c^2} = \frac{3}{8}$$

$$\frac{c}{a} = \sqrt{8/3} = 1.633 \tag{13.27}$$

Hence, the axial ratios *c*/*a* for an hcp structure is $\sqrt{8/3}$.

Keypoints to Remember

- Based on the arrangements of atoms or molecules, solids are classified into two categories, namely. crystalline solids and amorphous solids.
- In crystalline solids, atoms or molecules are arranged randomly and have isotropic properties.
- The smallest unit cell in a crystal structure is known as a unit cell.
- A simple cubic unit cell is said to be a primitive cell.
- Types of seven crystal systems are cubic, tetragonal, orthorhombic, monoclinic, triclinic, trigonal and hexagonal.
- The different types of unit cells under the seven crystal structures are called Bravis lattice.
- Miller indices are a set of three numbers written within the parenthesis and is used to represent a crystal plane.
- The co-ordination number is the nearest number of neighboring atom to a particular atoms.
- Density of packing, or packing density is defined as the ratio of total volume occupied by atoms in a unit cell to the volume of the unit cell.
- The co-ordination number for simple cubic, bcc, fcc and hcp structures are respectively 6,8,12 and 12.
- The packing density for simple cubic, bcc, fcc and hcp structures are respectively 0.52, 0.68, 0.74 and 0.74.
- In the hcp structure, the atoms are arranged in stacking sequences.

Solved Problems

Example 13.1 Show that the crystal plane (*hkl*) is perpendicular to the direction [*hkl*].

Solution The Miller indices of the given plane = (*hkl*)

The McGraw·Hill Companies

The intercept made by the plane $= a/h, b/k, c/l$
The equation of the plane

$$\frac{x}{a/h} + \frac{y}{b/k} + \frac{z}{c/l} = 1$$

i.e.
$$\frac{xh}{a} + \frac{yk}{b} + \frac{zl}{c} = 1$$

For cubic unit cell $a = b = c$
Therefore, the equation of the plane can be written as,

$$xh + yk + zl = a \qquad \text{(i)}$$

The direction ratio of normal to the plane is h, k, l. The equation of a line passing through the points (x_1, y_1, z_1) and having direction ratios h, k, l is

$$\frac{x - x_1}{h} = \frac{y - y_1}{k} = \frac{z - z_1}{l} \qquad \text{(ii)}$$

Equation (ii) is an equation for a line perpendicular to the plane given by Eq. (i).

The direction ratios of Eq. (ii) is h, k, l. The Miller indices of the line given in this problem is $[hkl]$. Hence, the Miller indices are nothing but the direction ratios.

The angle between two lines having the direction ratios u_1, v_1, w_1 and u_2, v_2, w_2 is given by,

$$\cos\theta = \frac{u_1u_2 + v_1v_2 + w_1w_2}{[u_1^2 + v_1^2 + w_1^2]^{1/2}[u_2^2 + v_2^2 + w_2^2]^{1/2}} \qquad \text{(iii)}$$

where $u_1 = u_2 = h$; $v_1 = v_2 = k$ and $w_1 = w_2 = 1$

Equation (iii) can be written as,

$$\cos\theta = \frac{h^2 + k^2 + l^2}{[h^2 + k^2 + l^2]^{1/2}[h^2 + k^2 + l^2]^{1/2}} = 1$$

Therefore, $\theta = 0.$

That is, the normal to the plane and the direction $[hkl]$ are parallel. It means, the direction $[hkl]$ is perpendicular to the plane.

Example 13.2 Find the Miller indices of a set of parallel planes which make intercepts in the ratio 3a: 4b on the X and Y axes and parallel to the Z axis. a, b, c are primitive vectors of the lattice of the lattice. Also calculate the interplaner distance of the planes taking the lattice to be a cubic with a = b = c = 2 A°.

Given Data

The intercept along X axis	$= 3a$
The intercept along Y axis	$= 4b$
The intercept along Z axis	$= \infty$

The McGraw·Hill Companies

Solution

The intercepts made by the plane along the three axes are 3a, 4b,
The coefficient of the intercepts are 3, 4, ∞
The inverse of the coefficients are 1/3, 1/4, 1/ ∞,
The L C M of these above fractions is 12
Multiplying these fractions by 12, we get, 4, 3, 0
Therefore, the Miller indices of the plane is (430).

The distance between any two successive parallel planes is given by

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

Substituting the values for 430 planes, we get,

$$d_{430} = \frac{2 \times 10^{-10}}{\sqrt{4^2 + 3^2 + 0^2}}$$

$$= 0.4 \times 10^{-10} \text{ m.}$$

The lattice spacing for the plane 430 is 0.4×10^{-10} m.

Example 13.3 Sodium is a bcc crystal. Its density is 9.6×10^2 kg m^{-3} and atomic weight is 23. Calculate the lattice constant a for sodium crystal.

Given Data

Density of sodium $= 9.6 \times 10^2$ kg m^{-3}
Atomic weight of sodium $= 23$

Solution The mass of the unit cell of the sodium crystal

= density × vol ume of the unit cell
$= 9.6 \times 10^2 \, a^3$ (i)

where a is the interatomic distance.
The mass of the unit cell of the sodium crystal is given by another relation,

$$\text{Mass} = \frac{\text{Number of atoms in a unit cell} \times \text{atomic weight}}{\text{Avogardo constant}}$$

Since sodium is a bcc crystal, the number of atoms present in one unit cell is 2.

Therefore, mass of one unit cell $= \dfrac{2 \times 23}{6.022 \times 10^{26}}$

$= 7.638 \times 10^{-26}$ kg (ii)

From (i) and (ii), $9.6 \times 10^2 \times a^3 = 7.638 \times 10^{-26}$

$a = 4.30 \times 10^{-10}$ m

The lattice constant of sodium is 4.3 Å.

Example 13.4 CsCl crystallizes in simple cubic crystal structure. The atomic weight of Cs is 132.9 and that of Cl is 35.5. The density of CsCl is 4×10^3 kg m^{-3}. Determine the value of the Avogadro constant. Given that the lattice constant of Cesium chloride is 4.12 Å.

The McGraw·Hill Companies

Given Data

The density of CsCl $= 4 \times 10^3$ kg m^{-3}
The atomic weight of Cs $= 132.9$
The atomic weight of Cl $= 35.5$
The lattice constant a $= 4.12$ Å $= 4.12 \times 10^{-10}$ m

Solution The mass of the CsCl unit cell = density × volume of the unit cell

$$= 4 \times 10^3 (4.12 \times 10^{-10})^3$$
$$= 2.797 \times 10^{-25} \text{ kg} \quad \text{(i)}$$

Let N be the Avogadro constant, then the mass of the CsCl unit cell is given by,

$$\text{Mass} = \frac{\text{number of atoms in one unit cell} \times \text{molecular weight}}{\text{Avogardro's constant}}$$

Since, CsCl belongs to simple cubic, the number of atoms present in a unit cell is one.

Therefore, the mass of one unit cell of CsCl $= \dfrac{1(132.9 + 35.5)}{N}$

$$= \frac{168.4}{N} \quad \text{(ii)}$$

From Eqs (i) and (ii), we get,

$$\frac{168.4}{N} = 2.797 \times 10^{-25}$$
$$N = \frac{168.4}{2.797 \times 10^{-25}} = 6.0199 \times 10^{26} \text{ per kg mole}$$

The value of the Avogadro constant = 6.0199×10^{26} per kg mole.

Example 13.5 X-rays of 1.5418 A° wavelength are diffracted by (111) planes in a crystal at an angle of 30° in the first order. Calculate the interatomic spacing.

Given Data

The wavelength of X-rays $\lambda = 1.5418$ ***A*** °
The angle of diffracted angle $\theta = 30°$
Miller indices for diffracted plane $hkl = (111)$

Solution

We know that the interplanar distance d for the first-order diffraction is

$$d = \frac{\lambda}{2 \sin \theta}$$

We also know that, the interplaner distance between the planes is

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

Equating the above equation, we get the lattice constant is

$$a = \frac{\lambda \times \sqrt{h^2 + k^2 + l^2}}{2 \sin\theta}$$

Substituting the values of λ h, k, l, s, θ in the above equation, we get,

$$= \frac{1.5418 \times 10^{-10} \times \sqrt{3}}{2 \times (1/2)}$$

$$= 2.6705 \times 10^{-10} \text{ m.}$$

The lattice constant a is 2.6705×10^{-10} m.

Example 13.6 Calculate the value of d–spacing for (110) planes in a rock salt crystal of $a = 2.814$ A°.

Given Data

The lattice constant $a = 2.814 \times 10^{-10}$ m

Solution

The distance between any two successive parallel planes is given by

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

Substituting the values for (100) planes, we get,

$$d_{100} = \frac{2.814 \times 10^{-10}}{\sqrt{1^2 + 0^2 + 0^2}}$$

$$= 2.814 \times 10^{-10} \text{ m}$$

The lattice spacing for the plane (100) is 2.814×10^{-10} m.

Example 13. 7 Calculate the interplaner distance for the (321) planes in SCC lattice with inter-atomic spacing equal to 4.12 A°.

Given Data

The lattice constant $a = 4.12 \times 10^{-10}$ m

Solution

The distance between any two successive parallel planes is given by

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

Substituting the values for (321) plane, we get,

$$d_{321} = \frac{4.2\times 10^{-10}}{\sqrt{3^2+2^2+1^2}}$$
$$= 1.1011 \times 10^{-10} \text{ m}$$

The lattice spacing for the plane (321) is 1.1011×10^{-10} m.

Example13.8 Calculate the interplaner spacing for (101) and (221) planes in a simple cubic whose lattice constant is 0.42 nm.

Given Data

The lattice constant $a = 4.2 \times 10^{-10}$ m

Solution

The distance between any two successive parallel planes is given by

$$d = \frac{a}{\sqrt{h^2+k^2+l^2}}$$

Substituting the values for the (110) plane, we get,

$$d_{101} = \frac{4.2\times 10^{-10}}{\sqrt{1^2+0^2+1^2}} = 2.9698 \times 10^{-10} \text{ m}$$

Substituting the values for the (221) plane, we get,

$$d_{101} = \frac{4.12\times 10^{-10}}{\sqrt{2^2+2^2+1^2}} = 1.4 \times 10^{-10} \text{ m}$$

The lattice spacings for the planes (110) and (220) are 2.968×10^{-10} m and 1.4×10^{-10}m. respectively.

Example 13.9 (a) Calculate the next neighbour's distance in a body centred cubic crystal. (b) In a body centred cubic lattice, find the ratio of the nearest neighbour's distance to the next neighbour's distance.

Solution

(a) In a body centred cubic crystal, body centred atom is the nearest neighbour to a corner atom. The distance between these two is $\left(\sqrt{3/2}\right)$a , where a is the lattice constant. For a corner atom, the next nearest neighbour is another corner atom. The distance between the two corner atoms is a.

Therefore, the next neighbour distance in a bcc lattice is a.
(b) In a bcc lattice, the next neighbour distance is a. The nearest $\left(\sqrt{3/2}\right)$a neighbour distance is. The ratio between these two is

$$d_1 : d_2 = \sqrt{3}/2a : a$$

The McGraw·Hill Companies

Here d_1 is the nearest neighbour distance and d_2 is the next neighbour distance. Therefore,

$$d_1 : d_2 = \sqrt{3}/2 : 1 \qquad \text{i.e.} \qquad d_1 : d_2 = 0.866 : 1$$

The ratio between the nearest neighbouring distance and the next neighbour distance is $\sqrt{3}/2 : 1$.

Example 13.10 In a crystal, lattice planes cut intercepts of lengths $2a$, $3b$ and $4c$ along the three axes. Deduce the Miller indices of the plane.

Given Data

The intercept along X-axis	$= 2a$
The intercept along Y-axis	$= 3b$
The intercept along Z-axis	$= 4c$

Solution

The intercepts made by the plane along the three axes are $2a$, $3b$ and $4c$.

The coefficients of the intercepts are 2, 3, 4
The inverse of the coefficients are 1/2, 1/3, 1/4
The L.C.M of the above three fractions is 12.
Multiplying these fractions by 12, we get 6, 4, 3.

Therefore, the Miller indices of the plane is (643).

Example 13.11 In a crystal whose primitives are 1.2 Å, 1.8 Å and 2 Å. A plane (231) cuts an intercept of 1.2 Å along the X-axis. Find the lengths of intercepts along the Y and Z axes.

Given Data

The primitives are	1.2 Å, 1.8 Å and 2 Å
The Miller indices of the plane	= (231)
The intercept made by the plane along the X-axis	= 1.2 Å

Solution

Method I

Let l_1, l_2 and l_3 are the intercepts made by the plane along the X, Y and Z axes. Hence, the ratio of the intercepts is

$$l_1 : l_2 : l_3 \qquad \text{(i)}$$

The Miller indices of the plane	= (231)
The inverses of the Miller indices	= 1/2, 1/3, 1/1

Multiplying these fractions by 6, we get, 3, 2, 6.
Therefore, the intercepts are in the ratio of

$$3a : 2b : 6c \qquad \text{(ii)}$$

Equating Eqs (i) and (ii), we get,

$$l_1 : l_2 : l_3 = 3a : 2b : 6c$$

The a, b, c values are given as 1.2 Å, 1.8 Å, 2 Å
Therefore, $l_1 : l_2 : l_3 = 3.6$ Å : 3.6 Å : 12 Å i.e. $l_1 : l_2 : l_3 = 3.6 : 3.6 : 12$
The value of l_1 is given as 1.2 Å.

Therefore, $l_2 = \frac{1.2}{3.6} 3.6 = 1.2$ Å and $l_3 = \frac{1.2}{3.6} 12 = 4$ Å

The lengths of the intercepts made by the plane along Y and Z axes are 1.2 Å and 4 Å.

Method II
Let (hkl) be the Miller indices of a crystal plane. If Miller indices are known, then the intercepts made by the plane is given by

$$\frac{a}{h}, \frac{b}{k}, \frac{c}{l}$$

The values of a, b and c are given as

$$a = 1.2 \text{ Å}, b = 1.8 \text{ Å}, c = 2 \text{ Å}$$

Substituting the values of a, b and c we get the ratios of intercepts as

$$\frac{1.2\text{Å}}{l} : \frac{1.8\text{Å}}{3} : \frac{2\text{Å}}{1} = 0.6 \text{ Å} : 0.6 \text{ Å} : 2 \text{ Å}$$

Let these ratios be $x : y : z$

Therefore, 0.6 Å: 0.6 Å : 2 Å = $x : y : z$

Substituting, the x axis intercept, we get

0.6 Å : 0.6 Å : 2 Å = 1.2 Å : $y : z$

The y intercept is determined by considering x and y intercepts.

i.e. 0.6 Å : 0.6 Å = 1.2 Å : y

$y =$ 1.2 Å

The z intercept is, 0.6 Å : 2 Å = $x : z$

$z =$ 4 Å

The y and z intercepts are 1.2 Å, 4 Å.

Example 13.12 The Miller indices of a crystal plane in a simple cubic crystal is (110). Find the ratio of the intercepts on the three axes.

Given Data

The Miller indices of the plane = (110)

Solution

The Miller indices of the given plane = (110)
The inverses of the Miller indices = 1/1, 1/1, 1/0 i.e. 1, 1, ∞

The intercepts made by the plane in terms of lattice parameters is $1a$, $1b$, ∞c.

Let l_1, l_2 and l_3 be the lengths of the intercepts made by the plane along the *X*. *Y* and *Z* axes. The ratios of these three lengths is $l_1 : l_2 : l_3$. If $l_1 : l_2 : l_3$ are the ratios of the intercepts, then they are related to the lattice parameters as,

$$l_1 : l_2 : l_3 = 1a : 1b : \infty c$$

Since, the given crystal is a simple cubic $a = b = c$
Therefore, the ratio

$$l_1 : l_2 : l_3 = 1a : 1a : \infty c$$
$$= 1 : 1 : \infty$$

The intercepts made by the given plane along the *Z* axis is ∞. It means that the plane is parallel to the *Z* axis.

Example 13.13 The lattice constant of a cubic lattice is 4.12 Å. Find the lattice spacings between (111), (112) and (123) lattice planes.

Given Data

The lattice constant $a = 4.12$ Å

Solution The distance between any two successive parallel planes is given by

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

(i) For (111) plane,

$$d_{111} = \frac{4.12 \times 10^{-10}}{\sqrt{1^2 + 1^2 + 1^2}} = 2.3787 \times 10^{-10} \text{ m}$$

(ii) For (112) plane,

$$d_{112} = \frac{4.12 \times 10^{-10}}{\sqrt{1^2 + 1^2 + 2^2}} = 1.68198 \times 10^{-10} \text{ m}$$

(iii) For (123) plane,

$$d_{123} = \frac{4.12 \times 10^{-10}}{\sqrt{1^2 + 2^2 + 3^2}} = 1.1011 \times 10^{-10} \text{ m}$$

The lattice spacing for the planes (111) $= 2.3787 \times 10^{-10}$ m
The lattice spacing for the planes (112) $= 1.6820 \times 10^{-10}$ m
The lattice spacing for the planes (123) $= 1.1011 \times 10^{-10}$ m

Example 13.14 Show that in a simple cubic lattice, the separation between the successive lattice plane (100), (110) and (111) are in the ratio of 1 : 0.71 : 0.58.

Solution The interplanar distance is given by

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

$$\text{For (100) plane, } d_{100} = \frac{a}{\sqrt{1^2+0^2+0^2}} = a$$

$$\text{For (110) plane, } d_{110} = \frac{a}{\sqrt{1^2+1^2+0^2}} = \frac{a}{\sqrt{2}}$$

$$\text{For (111) plane, } d_{111} = \frac{a}{\sqrt{1^2+1^2+1^2}} = \frac{a}{\sqrt{3}}$$

The ratio of $d_{100} : d_{110} : d_{111} = a : \frac{a}{\sqrt{2}} : \frac{a}{\sqrt{3}}$

i.e. $d_{100} : d_{110} : d_{111} = a : 0.707a : 0.577a$

The ratio of $d_{100} : d_{110} : d_{111} = 1 : 0.71 : 0.58$

or $d_{100} : d_{110} : d_{111} = \sqrt{6} : \sqrt{3} : \sqrt{2}$

Example 13.15 The lattice constant for a FCC structure is 4.938 A°. Calculate the interplaner spacing of the (220) plane.

Given Data

The lattice constant $a = 4.938 \times 10^{-10}$ m

Solution

The distance between any two successive parallel planes is given by

$$d = \frac{a}{\sqrt{h^2+k^2+l^2}}$$

Substituting the values for the 220 plane, we get,

$$d_{220} = \frac{4.938 \times 10^{-10}}{\sqrt{2^2+2^2+0^2}}$$

$$= 1.7458 \times 10^{-10}$$

The lattice spacing for the plane 110 is 2.968×10^{-10} m.

Example 13.16 The unit cell of aluminum is FCC with lattice constant $a = 0.405$ nm. How many unit cells are there in a 0.054 cm thick and 25 cm^2 square aluminum foil?

Given Data

The lattice constant $a = 0.405 \times 10^{-9}$ m.
The thickness of Al foil $t = 0.005 \times 10^{-2}$ m
The side length of the Al foil $A = 25 \times 10^{-4}$ m

Solution

We know that the number of unit cells in the Al foil is

$$= \frac{\text{Volume of Al foil}}{\text{Volume of unit cell}}$$

i.e., $$= \frac{t \times A}{a^3}$$

Substituting the values, we get,

$$= \frac{0.005 \times (25 \times 10^{-2})}{(0.405 \times 10^{-10})^3}$$

$$= \frac{1.25 \times 10^{-5}}{6.643 \times 10^{-29}} = 1.88 \times 10^{-26}$$

Therefore, the number of atoms in the Al foil is 1.88×10^{-26}.

Example 13.17 The lattice constant of a metal with cubic lattice is 2.88 A°. The density of the metal is 7200 kg m^{-3}. Calculate the number of unit cell present in the 1 kg of the metal.

Given Data

The lattice constant	$a = 2.88 \times 10^{-10}$ m.
Density of metal	$\rho = 7200$ kg m^{-3}

Solution

We know that the number of unit cells in 1 kg of the metal is

$$= \frac{1}{\text{mass of the unit cell}}$$

i.e., $$= \frac{1}{\text{volume of the unit cell} \times \text{density}}$$

Substituting the values, we get,

$$= \frac{1}{(0.288 \times 10^{-10})^3 \times 7200} = 5.8142 \times 10^{24}$$

Therefore, the number of unit cells present in 1 kg metal is 5.8142×10^{24}.

Example 13.18 Metallic iron changes from BCC to FCC form at 910°C. At this temperature, the atomic radii of the iron atoms in the structures are 0.1258 nm and 0.1292 nm respectively. Calculate the volume changes in percentage during this structural change.

Given Data

The atomic radius at BCC state $= 0.1258 \times 10^{-9}$ m

The atomic radius at FCC state $= 0.1292 \times 10^{-9}$ m

Solution

We know that, the lattice constant for BCC $a = \frac{4r}{\sqrt{3}}$

Substituting the values, we get,

$$= \frac{4 \times 0.1258 \times 10^{-10}}{\sqrt{3}} = 2.9052 \times 10^{-10} \text{ m}$$

Therefore, the volume occupied by one atom in BCC is $= \frac{a^3}{2}$

Substituting the values, we get,

$$= \frac{(2.9052 \times 10^{-10})^3}{2} = 1.226 \times 10^{-29} \text{ m}^3$$

The volume occupied by one atom in BCC state is $= 1.226 \times 10^{-29}$ m^3.

We know that, the lattice constant for FCC $a = \frac{4r}{\sqrt{2}}$

Substituting the values, we get,

$$= \frac{4 \times 0.1292 \times 10^{-10}}{\sqrt{2}} = 3.6543 \times 10^{-10} \text{ m}$$

Therefore, the volume occupied by one atom in FCC is

$$= \frac{a^3}{4}$$

Substituting the values, we get,

$$= \frac{(3.6543 \times 10^{-10})^3}{4} = 1.22 \times 10^{-29} \text{ m}^3.$$

The volume occupied by one atom in FCC state is $= 1.22 \times 10^{-29}$ m^3.
Therefore, change in volume in percentage is

$$= \frac{(1.226 \times 10^{-29} - 1.22 \times 10^{-29})}{1.226 \times 10^{-29}} \times 100 = 0.4894$$

The volume change in percentage during the structural changes is 0.4894.

Example 13.19 Copper has FCC structure and its atomic radius is 1.273 Å. Find the lattice parameter and the density of copper.

Atomic weight of copper = 63.5 g
Avogadros number = 6.023×10^{-26} atoms k mol^{-1}

Given Data

The atomic radius for FCC system $r = 1.273$ Å

Atomic weight of copper $A = 63.5$ g

$= 63.5 \times 10^{-3}$ kg

Number of atoms per unit cell for FCC $n = 4$

Solution

We know that Lattice parameter for FCC $a = \frac{4r}{\sqrt{2}}$

Substituting the values, we get,

$$= \frac{4 \times 1.273 \times 10^{-10}}{1.414}$$

$$\underline{\alpha} = 3.6 \times 10^{-10} \text{ m}$$

Density of the material is $\rho = \frac{nA}{Na^3}$

Substituting the given values, we get,

Density $$\rho = \frac{4 \times 63.5 \times 10^{-3}}{6.023 \times 10^{26} \times (3.6 \times 10^{-10})^3}$$

$$= 9.0388$$

Therefore, the density of copper is 9.0388 gm m^{-3}

Example 13.20 α-iron of 55.85 atomic weight solidifies into BCC structure and has a density of 7860 kg m^{-3}. Calculate the radius of an atom.

Given Data

The density of α-iron $\rho = 7860$ kg $m^{-3.}$

The atomic weight of α-iron $M = 55.85$

Number of atoms per unit cell for BCC $n = 2$

Avogadro's number $N = 6.023 \times 10^{26}$ atoms k mol^{-1}.

Solution

We know that the volume of the cube is

$$a^3 = \frac{\text{Number of atoms per unit cell} \times \text{molecdular weight}}{\text{Avogadro's number} \times \text{density}}$$

i.e., $$a^3 = \frac{n\,M}{N \times \rho}$$

Substituting the values, we get,

$$= \frac{2 \times 55.85 \times 10^{-3}}{6.626 \times 10^{26} \times 7860} = 2.3595 \times 10^{-32} \text{ m}^3.$$

The McGraw·Hill Companies

i.e., $a = 3\sqrt{2.3596 \times 10^{-32}}$

or $a = 0.2868 \times 10^{-10}$ m.

The lattice constant of α-iron is 0.2868 A°.

We know that the atomic radius for BCC is

$$r = \frac{a\sqrt{3}}{4}$$

Substituting the values, we get,

$$r = \frac{1.732 \times 0.2868 \times 10^{-10}}{4} = 0.12418 \times 10^{-10} \text{ m}$$

The atomic radius of the α-iron is $0.12418 \times ^{-10}$ m.

Example 13.21 Copper crystalline in the FCC structure. The density and atomic weight of copper are 8960 kg m^{-3} and 63.54 respectively. Calculate its lattice constant.

Given Data

The density of copper	$\rho = 8960$ kg m^{-3}.
The atomic weight of copper	$M = 63.54$
Number atoms per unit cell for FCC	$n = 4$
Avogadro's number	$N = 6.023 \times 10^{26}$ atoms k mol^{-1}.

Solution

We know that, the volume of the cube is

$$a^3 = \frac{\text{Number of atoms per unit cell} \times \text{molecdular weight}}{\text{Avogadro's number} \times \text{density}}$$

i.e., $a^3 = \dfrac{nM}{N \times \rho}$

Substituting the values, we get,

$$= \frac{4 \times 63.54 \times 10^{-3}}{6.023 \times 10^{26} \times 8960} = 4.7096 \times 10^{-32} \text{ m}^3.$$

i.e., $a = 3\sqrt{4.7096 \times 10^{-32}}$

or $a = 0.3611 \times 10^{-10}$ m.

The lattice constant of copper is 0.3611 Å.

The McGraw·Hill Companies

Objectives Questions

13.1 Anisotropic structures have _____ properties.
13.2 Examples for metallic crystals are _____, _____ and _____.
13.3 The structures which are not having directional properties are known as _____ substances.
13.4 Complete the equation, Crystal structure = Lattice + _____.
13.5 _____ is the fundamental building block of a crystal structure.
13.6 Body centered cubic unit cell is not a primitive cell. (True/False)
13.7 In a cubic crystal structure, the values of angles $\alpha = \beta = \gamma =$ _____.
13.8 The crystal plane is represented by a set of three numbers within _____.
13.9 Crystal directions are represented by a set of three numbers within _____.
13.10 The inter-atomic distance is equal to _____

a) $d = \dfrac{a^2}{\sqrt{h^2 + k^2 + l^2}}$ b) $d = \dfrac{\sqrt{\ } a^2}{\sqrt{h^2 + k^2 + l^2}}$

c) $d = \dfrac{a}{\sqrt{h^2 + k^2 + l^2}}$ b) $d = \dfrac{a^2}{h^2 + k^2 + l^2}$

13.11 The co-ordination number for a simple cube is _____.
13.12 The packing density for a simple cube is equal to _____.
13.13 _____ is the co-ordination number for body centered cubic structure.
13.14 _____ and _____ are respectively the co-ordination number and packing density for fcc structure.
13.15 The co-ordination number and packing density for hcp structure are respectively _____ and_____.
13.16 In the hcp structure, the relation between c and a is _____.

a) $\dfrac{c}{a} = \dfrac{8}{3}$ b) $\dfrac{c}{a} = \dfrac{\sqrt{8}}{3}$

c) $\dfrac{c}{a} = \sqrt{\dfrac{8}{3}}$ d) $\dfrac{c}{a} = \dfrac{8}{\sqrt{3}}$

13.17 Give three examples for cubic crystal planes _____, _____ and _____.
13.18 In BCC structure, the packing density of crystal is equal to

a) $\dfrac{\sqrt{3\pi}}{8}$ b) $\dfrac{\sqrt{3}\ \pi}{8}$

c) $\dfrac{\pi}{8}\sqrt{3}$ d) $\dfrac{3\sqrt{3\pi}}{8}$

KEY

13.1 Directional
13.2 Copper, silver, aluminium
13.3 Isotropic
13.4 Basis

The McGraw·Hill Companies

13.5 Unit cell
13.6 True
13.7 90º
13.8 Paranthesis
13.9 Paranthesis
13.10 c)
13.11 6
13.12 0.52
13.13 8
13.14 12, 0.74
13.15 12, 0.74
13.16 c)
13.17 (110), (010), (101)
13.18 b)

Exercises

13.1 Calculate the density of the NaCl crystal from the following data: Molecular weight of NaCl = 58.45; lattice constant a = 5.64 Å; Avogadro constant = 6.023×10^{26} per kg mole.

13.2 Calculate the lattice constant of potassium bromide. Density of potassium bromide is 2.7×10^3 kg m^{-3} and belongs to fcc lattice. The molecular weight of KBr is 119 and the Avogadro constant is 6.022×10^{26} per kg mole.

13.3 Calculate the next neighbour's distance in a simple cubic lattice and also determine the ratio between the nearest neighbour's distance and the next neighbour's distance for a simple cubic lattice.

13.4 Find the Miller indices of a plane, which makes intercepts at 3a, 2b and 2c.

13.5 A crystal plane makes intercepts at 2a, 3b and 6c along the three crystallographic axes. Here a, b and c are the primitive vectors of the unit cell. Determine the Miller indices of the given crystal plane.

13.6 The lattice parameters of a unit cell is 1.6 Å, 2.22 Å and 1.84 Å. A plane having Miller indices value of (231) makes an intercept at 1.84 Å along the Z axis. Find the lengths of the intercepts made by the plane along X and Y axes.

13.7 Calculate the interplanar spacing for a (321) plane in a simple cubic lattice whose lattice constant is 4.12×10^{-10} m.

Short Questions

13.1 Define space lattice.
13.2 What is meant by basis?
13.3 Define the term crystal structure.
13.4 What are interfacial angles?
13.5 Define the term unit cell.
13.6 What is meant by lattice Paranthesis?
13.7 What is Primitive cell?
13.8 Disting with the Primitive cell and unit cell.

The McGraw·Hill Companies

13.9 Name the seven types of crystal systems.
13.10 What is Bravais lattice?
13.11 Name the crystal structure of the following (a) Gold, (b) Geranium, (c) Bartu, and (d) Zinc.
13.12 Bismith has a = b = c = 4.74 AU and angles $\alpha = \beta = \gamma = 60°$ what it's crystal structure?
13.13 What are Miller indices?
13.14 Explain fow the miller indrices are determined.
13.15 Describe the structure of hcp.
13.16 Define the following terms:
(i) Atomic radius
(ii) Coordination number, and
(iii) Packing density
13.17 State the values of Co-ordination number for hcp structure.
13.18 Calculate the packing factor for a simple cubic structure.
13.19 Calculate the packing factor for a body contracte cubic.
13.20 Calculate the packing factor for a face centred cubic.
13.21 Define the Co-orination number and the calculate the same for fcc lattice
13.22 Show that for a bcc crystal structure, the lattice constant is given by $a = 4r/\sqrt{3}$, where r is the atomic radius.
13.23 Establish the relation between the radius and the interatomic distance for a face centred cubic crystal.
13.24 Draw the diagram of the unit cell of the close packed hexagonal lattice.
13.25 Fero cubic lattice drow the followig plane showing the value of intercapte with the Co-ordination axis (a) [110], (b) [111] and (c) [021].
13.26 Draw the schematic diagrams of a simple cubic, body centred cubic and face centred cubic unit cells.

Descriptive Questions

13.1 What are Miller indices? How will you determine the Miller indices of a given plane? What are the distinct features of Miller indices?
13.3 Define the terms coordination number, atomic radius, and packing density. Calculate the above factors for a simple cubic, body centred cubic and face centred cubic crystals.
13.4 Determine the coordination number and packing density for a hexagonally closed packed structure. Show that a hcp structure demands an axial ratio of 1.633.
13.5 What is packing factor? Prove that the packing faction of hcp is 0.74

Chapter

14

Waves And Particles

OBJECTIVES

- To explain the particle and wave nature (matter waves) of the quantum particles
- To derive the de-Broglie equation related with momentum and wavelength of the particle
- To discuss the properties of matter waves and explain the experimental evidence of matter waves
- To derive the Schrödinger wave equation for the motion of the quantum particles
- To discuss the importance of the uncertainty principle
- To derive application of Schrödinger equation for a particle in one-dimensional box and its importance

14.1 INTRODUCTION

Classical mechanics which is based on Newton's laws, explains the large-scale experiments. On the other hand, Newtonian mechanics fails to explain the different phenomena which take place at microscopic or atomic scales. Quantum theory was able to explain the nature of light waves in terms of packets or bundles known as photons or quanta. Some of the phenomena like interference, diffraction, polarization, etc., are explained based on the wave nature of light. However, the phenomena like photoelectric effect, Compton effect, Zeeman effect, etc., are explained based on the particle nature of light. Therefore, radiation, or light, behaves like a wave as well as a particle. Thus, the wave–particle property of light radiation is known as the dual nature of light. The concepts which prompted de Broglie to propose the idea of matter waves are namely i) the universe is made up of radiation and matter, and ii) nature is symmetrical everywhere.

In the present chapter, an attempt has been made to explain the dual nature of light, i.e., de Broglie waves, their experimental verification, Schrödinger wave equations and their applications in quantum mechanics.

The McGraw·Hill Companies

14.2 DE BROGLIE WAVE

According to de Broglie, light waves behave sometimes as particles, and conversely particles, can have wavelike characteristic properties. Therefore, the particles like electrons, protons, neutrons, atoms or molecules will have associated waves with them known as *matter waves* or *pilot waves* or *de Broglie waves*. A variable quantity known as *wave function* is used to characterise the de Broglie waves. The wave function is denoted by the symbol ψ (psi).

14.3 DE BROGLIE WAVELENGTH

Let us consider that a photon with frequency ν is traveling with the velocity of light c.

Then, the velocity of light $c = \nu\lambda$ (14.1)

where λ is the wavelength of light.

According to Einstein, the energy of each photon

$$E = h\nu \tag{14.2}$$

where E is the energy of the photon.

Considering the momentum of the photon subjected to an external force, Eq.(14.1) and Eq. (14.2) can be combined as,

$$p = \frac{E}{c} \tag{14.3}$$

Substituting the value of $c = \upsilon\lambda$ and $E = h\nu$, we get,

$$p = \frac{h\,\nu}{\nu\,\lambda}$$

or

$$p = \frac{h}{\lambda} \tag{14.4}$$

From the above equation, it is clear that the light photon with wavelength λ and momentum $p = h/\lambda$ will have an associated wavelength λ and also a wavelike character. In 1924, de Broglie suggested that the matter particles will have associated waves known as de Broglie waves (or) matter waves. The value of the wave function associated with a moving particle at any point (x,y,z) and time t determines the probability of finding the particle at that corresponding point and time.

Consider a particle of mass m moving with a velocity v. Then, the wavelength associated with the particle is

$$\lambda = \frac{h}{mv} \tag{14.5}$$

where $p = mv$ is the momentum of the particle.

Equation (14.5) is known as the de Broglie wavelength of the matter wave.

14.3.1 de Broglie Wavelength in terms of KE

Consider a particle of mass m moving with a velocity v. Then, the kinetic energy (KE) of the particle is,

$$E = (1/2)\,m v^2$$

$$= \frac{1}{2m}\, m^2 v^2 = \frac{p^2}{2m}$$

We know that $p = mv$,

$\therefore$ The momentum $p = \sqrt{2mE}$ (14.6)

Substituting the value of p in Eq. (14.5), we get,

The wavelength $\lambda = \dfrac{h}{\sqrt{2mE}}$ (14.7)

Equation (14.7) represents the de Broglie wavelength interms of kinetic energy.

14.3.2 de Broglie Wavelength in terms of v

Consider an electron of mass m and charge e, accelerated through a potential difference of V volts. Then, the KE of the electron is equal to the loss in potential energy.

i.e., $\dfrac{1}{2} m v^2 = eV$ (14.8)

or

$$m^2 v^2 = 2\ \mathrm{m}\ eV$$

$\therefore$ The momentum of the electron

$$p = m v = \sqrt{2\,m\,e\,V} \quad (14.9)$$

Substituting the value of mv in Eq. (14.5), we get,

de Broglie wavelength of electron $\lambda = \dfrac{h}{\sqrt{2\,m\,e\,V}}$ (14.10)

Substituting the value of h, e and m in Eq. (14.10), we get,

De Broglie wavelength of electron $\lambda = \dfrac{6.626\times10^{-34}}{\sqrt{2\times9.1\times10^{-3}\times1.6\times10^{-19}\times V}}$

$$= \frac{12.28\times10^{-10}}{\sqrt{V}}\ \mathrm{m}$$

$$\lambda = \frac{12.28}{\sqrt{V}}\ \text{Å} \quad (14.11)$$

Equation (14.11) gives the de Broglie wavelength of an electron.

The McGraw·Hill Companies

14.4 PROPERTIES OF MATTER WAVES

The following are the dominant properties of matter waves:

(i) Matter waves are not physical phenomena, but a wave probability.

(ii) The wavelength associated with matter waves decreases with increase in the mass of the particles,

i.e., $$\lambda = \frac{h}{\sqrt{2\,m\,e\,V}} \tag{14.12}$$

(iii) Matter waves are a new kind of waves and are not electromagnetic in nature.

(iv) Matter waves are mainly used as a wave to pilot or guide the matter particles and hence, are known as *pilot waves* or *guide waves*. The schematic representation of pilot or guide waves are shown in Fig. 14.1.

(v) The velocity of matter waves depends on the velocity of the material particle.

(vi) The phase velocity of a matter wave is inversely proportional to its wavelength.

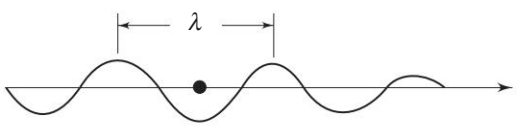


Fig. 14.1 *Pilot or guide wave*

(vii) Matter waves are not constant as electromagnetic waves and move faster than the velocity of light,

i.e., $$u = c^2/v \tag{14.13}$$

14.5 MATTER WAVES—EXPERIMENTAL VERIFICATION

The de Broglie hypothesis clearly states that particles like electrons will have associated wavelike characteristics and obey the properties of ordinary light like refraction, reflection, diffraction, etc. One can prove the wave nature of the particles through simple experiments, namely, Davission–Germer and Thomson's experiments. In both the experiments, electrons have been selected to prove the wave natures mainly due to their well known characteristic properties and for the easy generation and detection.

14.5.1 Davission–Germer Experiment

In 1927, Davission and Germer made an attempt to prove the wave nature of atomic particles. The schematic representation of the experimental set-up used to demonstrate the dual nature of matter is shown in Fig. 14.2.

The McGraw·Hill Companies

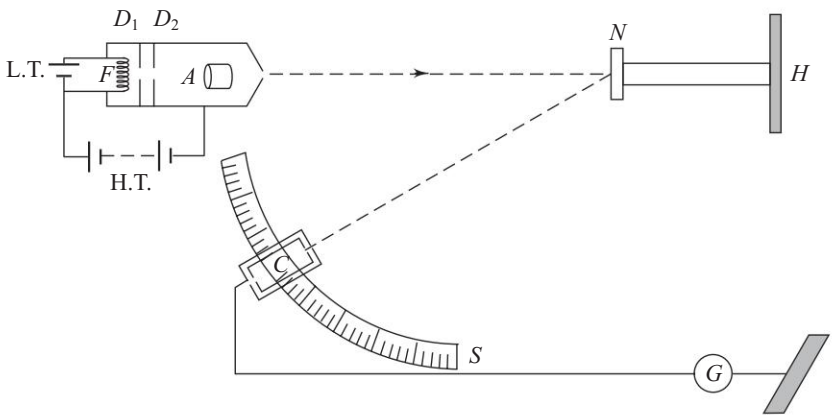


Fig.14.2 *Davission–Germer experiment—dual nature verification*

The electrons are produced by the hot filament F by applying the necessary potential through LT. The electrons are allowed to pass through the two thin aluminum diaphragms D_1 and D_2 in order to get a fine parallel beam of electrons. By applying the necessary potential V to the aluminium cylinder A, the electrons are accelerated, which are then allowed to strike the single crystal of nickel N which is fixed with the handle H. In order to get the diffraction at different points on the surface of the nickel N, one can rotate the nickel crystal through the handle H at any desired axis of the incident electron beam.

When the fast-moving electron beam strikes the nickel crystal, scattering of electrons takes place according to the Braggs diffraction formula for X-rays,

$$2d \sin \theta = n\lambda \tag{14.14}$$

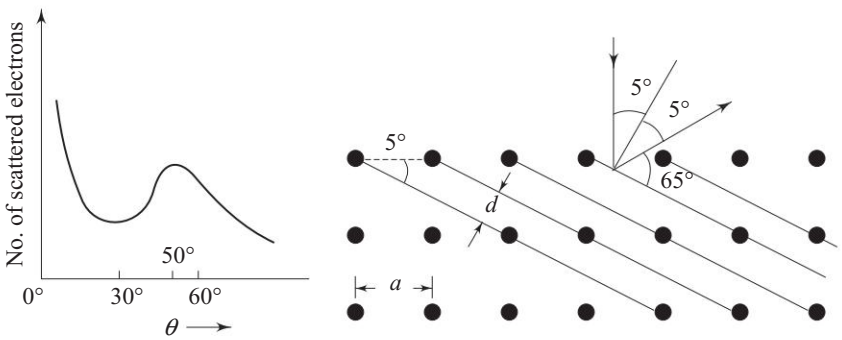
where d is the interplanar distance; θ, the scattering angle; n, the order of the spectrum and λ, the wavelength of the incident beam.

The scattered electron will be collected with the help of the detector known as the Faraday cylinder C. The detector moves on a graduated circular scale S. The Faraday cylinder collects the scattered electrons in all the directions by moving the same on the graduated scale. The amount of current produced by the collected electrons by C is measured through the galvanometer G. The position of the cylinder on the scale measures the angle of scattering of the electrons. Then, by measuring the peak position of the deflection in the galvanometer and the position of the cylinder on the scale, one can obtain respectively the maximum current for the scattered electrons and the maximum scattering angle. One can measure the number of electrons received on the Faraday cylinder and the scattering angle for a given accelerating voltage V. The experiment can be repeated for different accelerating voltages, V. The above observation can be plotted by taking the angle along the x-axis and the number of electrons scattered along the y-axis, as shown in Fig.14.3a. A peak in the number of scattered electrons occurs at an angle of 50° for an applied voltage of 54 V. This means that the diffraction effect which exhibits from the surface layer of the crystal acting as a plane grating produces the first-order spectrum at an angle of 50°.

The wave nature of the electron can be verified by determining the wavelength of the reflected waves by considering both its particle and wave nature. The wavelength of the electron, by considering the particle nature, can be determined from Bragg's diffraction formula. The arrangements of atoms in the nickel crystal are shown in Fig.14.3b. The interatomic distance for the nickel crystal is 2.12×10^{-10} m. From the top surface of the crystal, the reflecting planes in the crystal are inclined at an angle of 25^0.

Therefore, the interplanar distance

$$d = 2.15 \times 10^{-10} \sin 25^\circ$$
$$= 0.909 \times 10^{-10} \text{ m} \qquad (14.15)$$



a) Arrangement of atoms *b) Scattering angle versus number of electrons collected*

Fig.14.3 *Scattering angle and arrangement of atoms in nickel crystal*

According to Bragg's formula,

$$2 \times 0.909 \times 10^{-10} \sin (90^\circ - 25^\circ) = 1\lambda$$

Therefore, the de Broglie wavelength $\lambda = 1.65$ Å (14.16)

According to de Broglie, the electrons, which are accelerated by a potential difference of 54 V,

de Broglie wavelength $\lambda = \dfrac{12.28}{\sqrt{54}} = 1.67$ Å (14.17)

Equations (14.16) and (14.17) confirm the close agreement between the two wavelengths of the electron obtained by considering the wave and particle nature of electrons.

14.5.2 Thomson's Experiment

Another experiment which is used to study the wave nature of the electron was developed by GP Thomson. The Thomson experiment is similar to Von Laue's X-ray diffraction experiment. The schematic representation of the experimental set-up is shown in Fig.14.4.

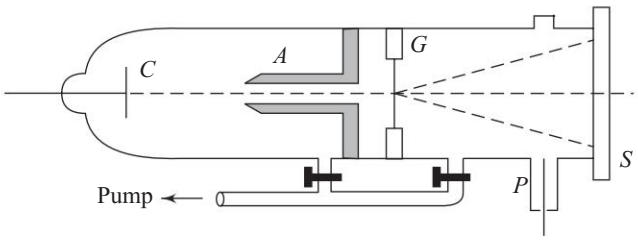


Fig.14.4 *Thomson experiment—dual nature verification*

The accelerated fast-moving narrow beam of electrons is produced by the cathode C by applying a potential of 50 kV. In order to get a fine narrow beam of electrons, it is further passed into a fine hole *H* as shown in Fig. 14.5. These electrons are made incident on the thin gold foil *F*. Thus, the diffraction of the electrons takes place at *F*. The diffracted electrons are made to fall either on a florescent screen *S* or a photographic plate *P,* as shown in Fig.14.4. One can use either a photographic plate (or) screen to obtain the pattern. When an electron is incident on a thin foil, both diffraction and generation of secondary X-rays takes place. The wave nature of the electrons can be verified by considering the reason for the existence of the diffracted pattern, i.e., the pattern is due to the existence of electrons or secondary X-rays.

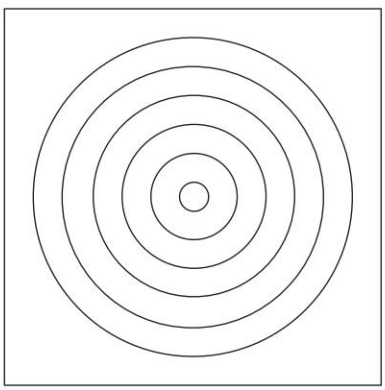


Fig.14.5 *Diffraction pattern*

The different patterns produced by the electrons which are diffracted at the foil *F* is shown in Fig.14.5. The pattern consists of concentric rings with increasing diameter. This is possible due to the periodic arrangement of atoms in the gold foil and it obeys the Braggs diffraction condition to give proper reflection. The above results confirm that the pattern is only due to the existence of electrons and not due to the secondary X-rays coming out of the thin foil *F*. The same can be explained in another way. When the foil is removed, the entire pattern disappears which confirms that the foil is essential to obtain the diffraction pattern. On the other hand, if the electron behaves as a particle when it is incident on the foil, it gets scattered over a wide angle. The diffracted electrons are the ones that produce the diffraction pattern.

From the above experiment, it is clear that electrons behave like a wave since the obtained diffraction pattern is possible only by waves. The size of the crystal unit

The McGraw·Hill Companies

cell can be obtained by knowing the ring diameter and the de Broglie wavelength of electrons. The size of the unit cell obtained by this method and X-rays are in close agreement.

14.6 SCHRÖDINGER WAVE EQUATION

In 1926, Schrödinger developed mathematical equations to explain both microscopic and macroscopic particles by predicting the wave functions at any particular point. The Schrödinger equation is one of the basic equations in quantum mechanics similar to Newton's law of motion. Schrödinger derived two forms of wave equations, namely, time independent wave equation, and time-dependent wave equation, to explain the dual nature of matter waves.

de Broglie made two assumptions to derive the equation for the wave. In his first assumption, the de Broglie wavelength concept was applied to any matter waves under any field of force. When such an external force is applied, the particles will have both potential and kinetic energy, i.e.,

$$E = \text{PE} + \text{KE}$$

$$E = V + (1/2)\, mv^2$$

The total energy
$$E = V + \frac{p^2}{2m} \tag{14.18}$$

Rearranging Eq.(14.18), we get,

$$p = [2m\,(E - V)]^{1/2} \tag{14.19}$$

According to de Broglie wavelength, we know that,

or
$$\lambda = \frac{h}{p}$$

$\therefore$ the de Brogile wavelength
$$\lambda = \frac{h}{[2m(E - V)]^{1/2}} \tag{14.20}$$

According to de Broglie's second assumption, the one-dimensional wavefunction associated with the particle is

$$\psi = \psi_0 e^{-i\omega t} \tag{14.21}$$

where ψ_0 is the amplitude of the wave function at the point (x,y,z) at any time which is independent of time and $\omega = 2\pi\nu$, where ν is the frequency of radiation.

14.6.1 Time Independent Wave Equation

Consider a system of stationary waves associated with a moving particle. ψ is the wave function of the particle along x, y and z co-ordinate axes at any time t. Based on the Cartesian co-ordinates, one can write the differential wave equation of a progressive wave with wave velocity u as,

$$\frac{d^2\psi}{dx^2} + \frac{d^2\psi}{dy^2} + \frac{d^2\psi}{dz^2} = \frac{1}{u^2} \times \frac{d^2\psi}{dt^2} \tag{14.22}$$

The solution for Eq. (14.22) is Eq. (14.21).

The value of $\frac{d^2\psi}{dt^2}$ can be obtained by differentiating Eq. (14.21) w.r.t. t twice.

$$\frac{d\psi}{dt} = -i\omega\psi_0 e^{-i\omega t} = -i\omega\psi$$

or

$$\frac{d^2\psi}{dt^2} = -\omega^2\psi \tag{14.23}$$

Substituting the value of $d^2\psi/dt^2$ in Eq. (14.22), we get,

$$\frac{d^2\psi}{dx^2} + \frac{d^2\psi}{dy^2} + \frac{d^2\psi}{dz^2} = -\frac{\omega^2}{u^2}\psi \tag{14.24}$$

We know that $\omega = 2\pi\nu$.

Therefore, Eq. (14.24) can be written as, $= -\frac{(2\pi\nu)^2}{u^2}\psi$

$$\frac{d^2\psi}{dx^2} + \frac{d^2\psi}{dy^2} + \frac{d^2\psi}{dz^2} = -\frac{4\pi^2\nu^2}{u^2}\psi \tag{14.25}$$

Substituting the value of u $(= \nu\lambda)$ in Eq. (14.25), we get,

$$\frac{d^2\psi}{dx^2} + \frac{d^2\psi}{dy^2} + \frac{d^2\psi}{dz^2} = -\frac{4\pi^2\nu^2}{\lambda^2\nu^2}\psi \tag{14.26}$$

Substituting the value of the wavelength from Eq. (14.20) in Eq. (14.26), we get,

$$\frac{d^2\psi}{dx^2} + \frac{d^2\psi}{dy^2} + \frac{d^2\psi}{dz^2} = -\frac{4\pi^2}{h^2} \times 2m(E-V)\psi \tag{14.27}$$

We know that ∇^2 is the Laplacian operator and is equal to

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{14.28}$$

The Laplacian operator operates on the wave function and reproduces the corresponding wave function and its energy.
Equation (14.27) can be written as,

$$\nabla^2\psi = -\frac{8\pi^2}{h^2} m(E - V)\psi \tag{14.29}$$

$$= -\frac{2m}{\hbar^2}(E - V)\psi$$

where $\hbar = \frac{h}{2\pi}$

Therefore, $$\nabla^2 \psi + \frac{2m}{\hbar^2}(E - V)\,\psi = 0 \tag{14.30}$$

The Schrödinger time independent equation can be taken either from Eq. (14.29) or (14.30) based on the requirements.

14.6.2 Time Dependent Wave Equation

On the other hand, for a free particle one cannot use Eq. (14.30), since for a free particle, the potential energy $V = 0$. Hence, a more general time-dependent equation is required. Therefore, Eq. (14.30) can be written by taking $V = 0$.

$$\nabla^2 \psi + \frac{2m}{\hbar^2} E\psi = 0 \tag{14.31}$$

According to Schrödinger's second assumption, Eq. (14.20) can be differentiated w.r.t. t, and we get,

$$\frac{d\psi}{dt} = -i\omega\psi = -i\,2\,\pi\,\nu\,\psi \tag{14.32}$$

Where $E = h\nu$. We get,

$$\frac{d\psi}{dt} = \frac{-i\,2\pi E}{h}\psi = -i\,\frac{E}{\hbar}\psi$$

$$E\psi = -\frac{\hbar}{i}\frac{d\psi}{dt}$$

or

$$E\,\psi = i\,\hbar\,\frac{d\,\psi}{dt} \tag{14.33}$$

According to Schrödinger's time-independent equation,

$$\nabla^2 \psi + \frac{2m}{\hbar^2} E\psi - \frac{2m}{\hbar^2} V\,\psi = 0$$

Substituting the value of $E\psi$ from Eq. (14.33), we get,

$$\nabla^2 \psi + \frac{2m}{\hbar^2}\,i\,\hbar \times \frac{d\psi}{dt} - \frac{2m}{\hbar^2} V\,\psi = 0 \tag{14.34}$$

Multiplying Eq. (14.34) by $\hbar^2 / 2m$, we get,

$$\frac{\hbar^2}{2m}\nabla^2\psi + i\hbar\frac{d\psi}{dt} - V\psi = 0$$

or

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = i\hbar\frac{d\psi}{dt} \tag{14.35}$$

H is a Hamiltonian operator and is equal to

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V \tag{14.36}$$

Substituting the value of H in Eq. (14.36), we get,

$$H\psi = E\psi \tag{14.37}$$

where $E\psi = i\hbar \times \frac{d\psi}{dt}$ is an energy operator.

The Schrödinger time-dependent wave equation can be either in the form of Eq. (14.35) or (14.37) depending on the applications.

14.6.3 Physical Significance of Wave Functions

The following are the significance of a wave function:

(i) It gives a statistical relationship between the particle and wave nature.

(ii) It is a complex quantity and hence, one cannot measure it.

(iii) It is a function of position and time co-ordinate. Hence, it cannot be used to locate the position of particle.

(iv) A new term known as position *probability density* $P(r,t)$ is introduced to explain the wave function. It is equal to the product of the wave function ψ and its complex conjugate as,

$$P(r, t) = |\psi(r, t)|^2 \tag{14.38}$$

(v) The probability of finding a particle within a volume dv is

$$P = \int |\psi|^2 dv \tag{14.39}$$

where $dv = dx\, dy\, dz.$

(vi) When a particle definitely exists in a volume dv then its probability density is one,

i.e,
$$P = \int_{-\infty}^{+\infty} |\psi|^2 dv = 1 \tag{14.40}$$

(vii) Wave function does not have any physical meaning while the probability density has physical meaning.

14.7 APPLICATION OF SCHRÖDINGER'S EQUATION TO A PARTICLE IN A BOX

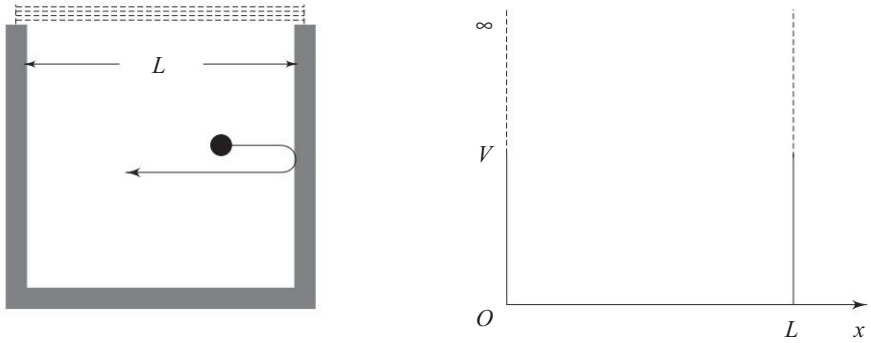


Fig. 14.6 *Schrödinger equation application Particle in a box*

Consider a one-dimensional box with width $x = 0$ to $x = L$ and infinite height as shown in Fig.14.6. Let E be the total energy of the particle with a mass m moving freely in the x-direction inside the potential box. The particle is moving inside the box with limitations $x = 0$ and $x = L$, and hence, its motion is restricted by the walls. Therefore, the particle will be inside the box and hence, its potential energy is zero. Therefore, one can use the Schrödinger time-independent equation to obtain the wave function and energy of the particle inside the box.

The particle is inside the box and hence, the potential energy is zero

i.e., for $0 < x < L$ $\qquad V = 0$

Therefore, the wave function

i.e., for $0 > x > L$ $\qquad |\psi|^2 = 0$

$$\psi = 0$$

The particle cannot be outside the box and hence, the potential energy outside the box is infinite,

i.e., for $0 \geq x \geq L$ $\qquad V = \infty$

Therefore, the wavelength $|\psi|^2 \neq 0$ for $0 < x < L$

According to Schrödinger, the one-dimensional time independent wave equation is

$$\frac{d^2\psi}{dx} + \frac{2m}{\hbar^2}(E - V)\psi = 0 \tag{14.41}$$

We know that for a freely moving particle, $V = 0$ in Eq. (14.41), and we get,

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}E\psi = 0 \tag{14.42}$$

or

$$\frac{d^2\psi}{dx^2} + k^2\psi = 0 \tag{14.43}$$

The McGraw·Hill Companies

where $$k^2 = \frac{2m}{\hbar^2} E \quad (14.44)$$

Equation (14.43) is similar to the equation of harmonic motion, and the solution of the above equation gives the wave function of the particle inside the box,

$$\psi = A \sin kx + B \cos kx \quad (14.45)$$

where A and Bare the amplitude of the electron's wave function.

Applying the boundary conditions, we get,

$$x = 0 \text{ and } \psi = 0$$

Therefore at this boundary condition the move function of electron is Zero.

Substituting the boundary conditions no. Eq. (14.45),we get, $B = 0$

Similarly, when $x = L$ and $\psi = 0$, we get $0 = A \sin kL$

We know that the electron is present inside the box and hence, $A \neq 0$

Therefore, we can write, $\sin kL = 0$

But, $\sin kL = 0$, only when $kL = n\pi$

Therefore, $$k = \frac{n\pi}{L} \quad (14.46)$$

Squaring Eq. (14.46) and comparing with the value of k^2 in Eq. (14.44), we get,

$$\frac{2m}{\hbar^2} E = \frac{n^2 \pi^2}{L^2}$$

Rearranging the above equation, we get,

$$E_n = \frac{n^2 \pi^2}{L^2} \times \frac{\hbar^2}{2m}$$

Substituting the value of $\hbar (= h / 2\pi)$, we get,

$$E_n = \frac{n^2 \pi^2}{L^2} \times \frac{h^2}{8\pi^2 m}$$

or

$$E_n = \frac{n^2 h^2}{8mL^2} \quad (14.47)$$

Eq. (14.47) given the permited energy levels of an electron in one dimentional at potential box.

Therefore, the wave function of the particle can be written by applying the boundary conditions to value of k from Eq. (14.46) in Eq. (14.45), and we get,

$$\psi_n = A \sin \frac{n\pi x}{L} \quad (14.48)$$

One can determine the value of A by considering the definite existence of an electron inside the box,

i.e., $$\int_{-\infty}^{\infty} | \psi |^2 \, dx = 1 \quad (14.19)$$

$$\int_0^L A^2 \sin^2\left(\frac{n\pi x}{L}\right) dx = 1 \tag{14.50}$$

Rearranging the above equation, we get,

$$\int_0^L \sin^2\left(\frac{n\pi x}{L}\right) dx = \frac{1}{A^2}$$

or

$$\frac{1}{A^2} = \int_0^L \left[\frac{1-\cos(2\pi n x / L)}{2}\right] dx$$

Integrating the above equation, we get,

$$\frac{1}{A^2} = \frac{L}{2}$$

Therefore,

$$A = \sqrt{\frac{2}{L}} \tag{14.51}$$

Substituting the value of A in Eq. (14.48), we get,
The wave function of the electron,

$$\psi_n = \sqrt{\frac{2}{L}} \sin\frac{n\pi x}{L} \tag{14.52}$$

Equations (14.52) and (14.47) give respectively the wave function and the corresponding energy levels. For each value of n, the above equations give the corresponding wave functions and energy levels. The allowed value of the wave function is known as *eigen function,* and the energy value is known as *eigen value*. We know that the particle is inside the box and hence, it will have a discrete value of wave function and energy. These wave functions and energies are normalised and are represented respectively by ψ_1, ψ_2, ψ_3, ψ_4 etc., and E_1, E_2, E_3, E_4, etc. The normalised wave function and energy of the particle inside the box are shown respectively in Fig.14.7.

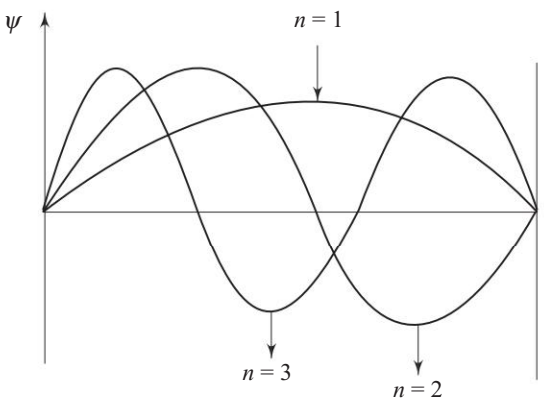


Fig.14.7 *One-dimensional potential box—energy levels*

The McGraw·Hill Companies

14.8 HEISENBERG UNCERTAINTY PRINCIPLE

Based on the classical theory, it is clear that each and every particle will have a fixed position and momentum in a given region of space. One can determine these quantities exactly at any given point of time. During measurement, there will be an inherent uncertainty in the determination of the position and momentum. The uncertainty in the simultaneous determination of position and momentum can be explained based on the Heisenberg uncertainty principle.

The Heisenberg principle states that the simultaneous determination of a pair of physical quantities like position and momentum of a particle cannot be determined with the required accuracy. Let Δx be the error in determining its position and Δp, the error in determining its momentum during the simultaneous determination. According to the Heisenberg principle, the above error quantities are related as

$$\Delta x \, \Delta p \simeq \frac{1}{2} \frac{h}{2\pi} \text{ or } \frac{h}{4\pi} \tag{14.53}$$

Equation (14.53) states that the product of two errors is approximately of the order of Planck's constant.

Let ΔE and Δt be respectively the uncertainty in the measurement of energy and time. Then according to the Heisenberg principle,

$$\Delta t \, \Delta E = \frac{1}{2} \frac{h}{2\pi} \text{ or } \frac{h}{4\pi} \tag{14.54}$$

Equation (14.53) and (14.54) clearly state that it is impossible to design an experiment to prove the wave and particle nature of matter at any given instant of time. If one measures the position or momentum accurately then there will be an uncertainty in the other. Thus, the Heisenberg uncertainty principle gives the probability of determining the particle at any given instant of time instead of a certainty.

14.8.1 Illustration

Consider that a beam of electrons is incident on a slit S of width Δx, which is perpendicular to the direction of the beam as shown in Fig. 14.8. The initial momentum of the electrons is $p = mv$. The electrons acquire momentum along OA after getting diffracted at the slit S. The angular deflection θ of the diffracted electrons depend on the component of the momentum parallel to the slit,

i.e., $$\Delta p = p \sin \theta$$

For a small angular deflection,

$$\Delta p = p\theta \tag{14.55}$$

The deflection angle θ_0 for the first minimum of the diffracted pattern is

$$\theta_0 = \frac{x}{\Delta x}$$

or

$$\Delta x = \frac{\lambda}{\theta_0} \tag{14.56}$$

The product of uncertainty in position and momentum

$$\Delta x \, \Delta p = p\theta \frac{\lambda}{\theta_0}$$

or

$$\Delta x \, \Delta p = p\lambda \frac{\theta}{\theta_0} \tag{14.57}$$

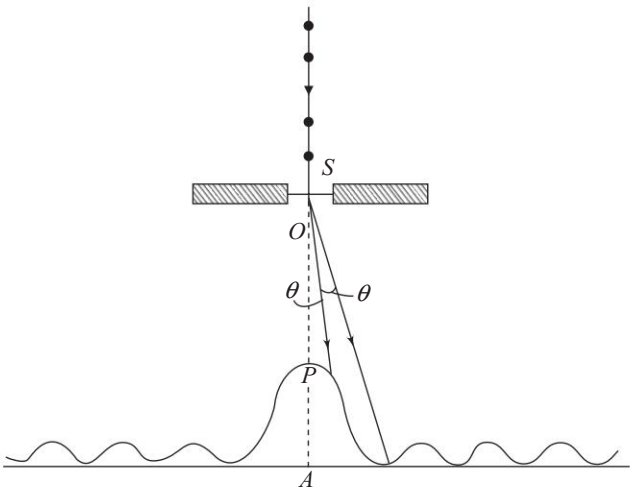


Fig. 14.8 *Heisenberg uncertainty principle—illustration*

For a small value of θ_0, it is nearly equal to θ (i.e., $\theta \approx \theta_0$ approximately)

Therefore, $$\Delta x \Delta p \sim h \tag{14.58}$$

where $$p = h/\lambda$$

The probable deflection θ of the electron is less than θ_0.

Therefore,

$$\Delta x \Delta p \geq \frac{1}{2} \frac{h}{2\pi} \text{ or } \frac{h}{4\pi} \tag{14.59}$$

Equation (14.59) states the uncertainty relation for the position and momentum according to the Heisenberg principle.

Keypoints to Remember

- All matter particles will have an associated wave known as matter wave or de Broglie wave.

- The de Broglie wavelength $\lambda = \frac{h}{mv} = \frac{h}{p}$
- The de Broglie wavelength of an electron $\lambda = \frac{h}{\sqrt{2meV}}$
- Matter waves are known as pilot waves (or) guide waves.
- The Davission–Germer and GP Thomson experiments are used to verify the wave nature of the particles experimentally.
- Schrödinger's time-independent and time-dependent wave equations are used to describe the dual nature of matter waves.
- Schrödinger's time-independent equation is $\nabla^2 \psi + \frac{2m}{\hbar^2}(E - V)\psi = 0$
- Schrödinger's time-dependent equation is $-\frac{\hbar^2}{2m}\nabla^2 \psi + V\psi = i\hbar \frac{d\psi}{dt}$
- Schrödingers time–dependent equation in terms of the Hamiltonian operator $H\psi = E\psi$
- According to the Heisenberg uncertainty principle, it is not possible to determine simultaneously a pair of physical quantities like position and momentum to a very high precision.

Solved Problems

Example 14.1 Determine the de Broglie wavelength of an electron accelerated by a potential difference of (i) 150 V, (ii) 5000 V, (iii) 400 V.

Given Data

(i) Potential difference $V = 150$ V
(ii) Potential difference $V = 5000$ V
(iii) Potential difference $V = 400$ V

Solution

Case I

The de Broglie wavelength of an electron, $\lambda = \frac{h}{\sqrt{2meV}}$ (i)

where h is the Planck's constant; m, the mass of the electron; and V, the potential difference applied.

Substituting the constant value in Eq. (i), we get,

The de Broglie wavelength of an electron $\lambda = \dfrac{6.626\times10^{-34}}{\sqrt{2x\times9.1\times10^{-31}\times1.6\times10^{-19}\times150}}$

$= 1.0018 \times 10^{-10}$ m

The de Broglie wavelength of the electron is 1.0018×10^{-10} m

Case II

The de Broglie wavelength of an electron, $\lambda = \dfrac{h}{\sqrt{2meV}}$ (ii)

where h is the Planck's constant; m, the mass of the electron; and V, the potential difference applied.

Substituting the constant value in Eq. (ii), we get,

The de Broglie wavelength of an electron $\lambda = \dfrac{6.626\times10^{-34}}{\sqrt{2\times9.1\times10^{-31}\times1.6\times10^{-19}\times5000}}$

$= 0.17353 \times 10^{-10}$ m

The de Broglie wavelength of the electron is 0.17353×10^{-10} m

Case III

The de Broglie wavelength of an electron, $\lambda = \dfrac{h}{\sqrt{2meV}}$ (iii)

where h is the Planck's constant; m, the mass of the electron; and V, the potential difference applied.

Substituting the constant value in Eq. (iii), we get,

The de Broglie wavelength of an electron $\lambda = \dfrac{6.626\times10^{-34}}{\sqrt{2\times9.1\times10^{-31}\times1.6\times10^{-19}\times400}}$

$= 0.6136 \times 10^{-10}$ m

The de Broglie wavelength of the electron is 61.36×10^{-10} m

The de Broglie wavelength of the electron

(i) accelerated for a potential of 150 V is 1.0018×10^{-10} m
(ii) accelerated for a potential of 5000 V is 0.17353×10^{-10} m
(iii) accelerated for a potential of 400 V is 0.6136×10^{-10} m

Example 14.2 Determine the wavelength associated with an electron having an energy of 100 eV

Given Data

Energy of the electron, $E = 100$ eV

Solution

Energy of the electron, $E = 100 \times 1.6 \times 10^{-19} = 1.6 \times 10^{-17}$ J

The de Broglie wavelength of an electron $\lambda = \dfrac{h}{\sqrt{2mE}}$

Substituting the values of *h, m* and v, we get $= \dfrac{6.626\times10^{-34}}{\sqrt{2\times9.11\times10^{-31}\times1.6\times10^{-17}}}$

$$\lambda = 1.227\times10^{-10}\text{ m}$$

The de Broglie wavelength of the electron is 1.227×10^{-10} m.

Example 14.3 A neuton of 1.675×10^{-27} kg mass is moving with a kinetic energy of 10 KeV. Calculate the de Broglie wavelength associated with it.

Given Data

Kinetic energy of the neutron $E = 10$ keV

$E = 10\times10^{3}\times1.6\times10^{-19}$

$= 1.6\times10^{-15}$ J

Solution

The de Broglie wavelength of the neutrons $\lambda = \dfrac{h}{\sqrt{2mE}}$

$$= \frac{6.626\times10^{-34}}{\sqrt{2\times1.675\times10^{-27}\times1.6\times10^{-15}}}$$

$$= 2.861\times10^{-13}\text{ m}$$

The de Broglie wavelength is 2.861×10^{-13} m.

Example 14.4 Determine (i) the velocity of an electron, (ii) phase velocity of the electron wave, (iii) the de Broglie wavelength, (iv) momentum, and (v) the wave number of an electron wave. The electrons which are at rest are accelerated through a potential difference of 125 V.

Given Data

Potential difference applied, $V = 125$ V

Solution

(i) The velocity of the electron, $v = \sqrt{\dfrac{2eV}{m}} = \sqrt{\dfrac{2\times1.6\times10^{-19}\times125}{9.1\times10^{-31}}}$

$$= 6.6299\times10^{6}\text{ m s}^{-1}$$

(ii) The phase velocity of the electron, $u = \dfrac{c^2}{v} = \dfrac{(3\times10^{8})}{6.6299\times10^{6}}$

$$= 1.3574\times10^{10}\text{ m s}^{-1}$$

(iii)The de Broglie wavelength, $\lambda = \dfrac{h}{mv} = \dfrac{6.626\times10^{-34}}{9.1\times10^{-31}\times6.6299\times10^{6}}$

$$= 1.0982 \times 10^{-10} \text{ m}$$

$$\lambda = 1.0982 \text{Å}$$

(iv) The momentum of the electron, $p = mv$

$$= 9.1 \times 10^{-31} \times 6.6299 \times 10^{6}$$

$$= 6.033 \times 10^{-24} \text{ kg m s}^{-1}$$

(v) The wave number of the electron wave, $\nu = \frac{1}{\lambda}$

$$= \frac{1}{1.0982 \times 10^{-10}}$$

$$= 9.1058 \times 10^{9} m^{-1}$$

The velocity of the electron is 6.6299×10^{6} m s^{-1}.
The phase velocity of the electron is 1.3574×10^{10} m s^{-1}.
The de Broglie wavelength is 1.0982 Å.
The momentum of the electron is 6.033×10^{-24} kg m s^{-1}.
The wave number of the electron wave is 9.1058×10^{9} m^{-1}

Example 14.5 Calculate the energy of the electron in the energy level immediately after the lowest energy level, confined in a cubical box of 0.1 nm side. Also, find the temperature at which the energy of the molecules of a perfect gas would be equal to the energy of the electrons in the above-said level.

Given data

The length of the potential well $L = 0.1 \times 10^{-9}$m

Solution

For the lowest energy level $n_x = n_y = 1$
Let n_z be the higher level $n_z = 2$
and hence, the lowest level will be $n_x = n_y = 1$
The energy of the electron in the (1 1 2) state is

$$E_{112} = \frac{h^2}{8mL^2}(1^2 + 1^2 + 2^2)$$

$$= \frac{6h^2}{8mL^2}$$

Substituting the values of *h, m* and *L* in the above equation, we get,

$$E_{112} = \frac{6 \times (6.62 \times 10 - 34)^2}{8 \times 9.1 \times 10^{-31} \times (0.1 \times 10^{-9}) \times 2}$$

$$= 36.12 \times 10^{-18} J$$

Therefore, the energy of the electron is 36.12×10^{-18} J

The energy E_{112} is equal to the average energy of the molecule.

Therefore, $$E_{112} = \frac{3}{2}kT$$

$\therefore$ The temperature $$T = E_{112}\,\frac{2}{3k}$$

Substituting the values of E_{112} and k in the above equation, we get

$$T = \frac{2\times 36.12\times 10^{-18}}{3\times 1.38\times 10^{-23}}$$

$$= 17.45\times 10^{5}\ \text{K}$$

The energy of the electron is 36.12×10^{-18} J
The temperature of the system is 1.745×10^{6} K

Example 14.6 An electron is trapped in a one-dimensional box of 0.1nm length. Calculates the energy required to excite it from its ground state to the fifth excited state.

Given data

The length of the one-dimensional box $= 0.1$ nm
$= 0.1\times 10^{-9}$ m

Solution
We know that the energy of the electron in a one-dimensional well is

$$E_n = \frac{n^2 h^2}{8mL^2}$$

For the ground state, $n = 1$
Substituting the values of h, m, and L in the above equation

$$E_1 = \frac{1^2\times(6.626\times 10^{-34})^2}{8\times 9.11\times 10^{-31}\times(0.1\times 10^{-9})^2}$$

$$= 6.0233\ \text{J}$$

$$= \frac{6.0233}{1.609\times 10^{-19}} = 37.64\ \text{eV}$$

For the fifth excited state, $n = 6$

$\therefore$ $$E_6 = \frac{6^2\times(6.626\times 10^{-34})^2}{8\times 9.11\times 10^{-31}\times(0.1\times 10^{-9})^2}$$

$$= 2.16805\ \text{J}$$

$$= \frac{2.16805}{1.609\times 10^{-19}} = 1355.03\ \text{eV}$$

$\therefore$ The energy required to excite the electron from the ground state to the fifth excited state is given by

The McGraw·Hill Companies

$$E = E_6 - E_1$$
$$= 1355.03 - 37.64$$
$$= 1317.93 \text{ eV}$$

$\therefore$ The energy required to excite the electron from the ground state to the fifth excited state is 1317.39 eV.

Example 14.7 Find the energy of the electron moving in one dimension in an infinitely by high potential box of 0.1 nm width.

Given data

The length of the one-dimensional potential box is 0.1 nm
$$= 0.1 \times 10^{-9} \text{ m}$$

Solution

For the lowest energy level, $n = 1$

The energy of the electron in a ground state is

$$E_1 = \frac{h^2}{8mL^2}(1)^2$$
$$= \frac{h^2}{8mL^2}$$

Substituting the value of *h, m* and *L* in the above equation, we get,

$$E_1 = \frac{(6.626\times10^{-34})^2}{8\times9.11\times10^{-31}\times(0.1\times10^{-9})^2}$$

$$= 6.0233 \times 10^{-18} \text{ J} = \frac{6.0233\times10^{-18}}{1.609\times10^{-19}} \text{ J}$$

$$= 37.64 \text{ eV}$$

The energy of the electron in the ground level is 37.64 eV

Example 14.8

For the one-dimensional well, show that limit $\frac{E_{n+1} - E_n}{E_n} = 0$
$n \to \infty$

Given data

The limit of the quantum number $n \to \infty$

Solution

We know that the energy of the electron in a one-dimensional well is

$$E_n = \frac{n^2h^2}{8mL^2}$$

Similarly, for the $(n + 1)^{th}$ state, the energy

$$E_{n+1} = \frac{(n+1)^2h^2}{8mL^2}$$

∴ The difference in the energy between n^{th} and $(n + 1)^{th}$ state,

$$E_{n+1} - E_n = \frac{h^2}{8mL^2}\left[(n+1)^2 - n^2\right]$$

Rearranging the above equation, we get,

$$\frac{E_{(n+1)} - E_n}{E_n} = \frac{(n+1)^2 - n^2}{n^2}$$

Given that $n \to \infty$
∴ $(n + 1)^2 - n^2 \to 0$
Therefore,

$$\text{limit } \frac{E_{n+1} - E_n}{E_n} = 0$$
$$N \to \infty$$

Example 14.9 If the uncertainty in position of an electron is 4×10^{-10} m, calculate the uncertainty in its momentum.

Given data

The electron in the position $\Delta x = 4 \times 10^{-10}$ m

Solution

We know from the Heisenberg uncertainty principle,

$$\Delta x \, \Delta p \approx h$$

Therefore,

$$\Delta p = \frac{h}{\Delta x}$$

Substituting the value of h and Δx in the above equation, we get,

The error in the momentum $\Delta p = \dfrac{6.6 \times 10^{-34}}{4 \times 10^{-10}} = 1.65 \times 10^{-24}$ kg m s^{-1}.

The uncertainty in momentum Δp is 1.65×10^{-24} kg m s^{-1}.

Example 14.10 Calculate the minimum energy that an electron can posses in an infinitively deep potential well of 4 nm width.

Given data

The width of the potential well $L = 4$ nm
$= 4 \times 10^{-9}$ m

Solution

Energy of the electron in a one-dimensional potential well $= \dfrac{n^2 h^2}{8mL^2}$

$$E = \frac{1^2 \times (6.626 \times 10^{-34})^2}{8 \times 9.11 \times 10^{-34} \times (4 \times 10^{-9})^2} = 3.765 \times 10^{-21} \text{ Joules}$$

$$E = \frac{3.765\times10^{-21}\,J}{1.6\times10^{-19}} = 0.235 \text{ eV}$$

Example 14.11 An electron is confined to a box of 10^{-9} m length. Calculate the minimum uncertainty in its velocity.

Given data

The electron in the position $\Delta x = 10^{-9}$ m
Mass of the electron $m = 9.1 \times 10^{-31}$ kg
Planck's constant $h = 6.6 \times 10^{-34}$ J s

Solution

We know that $\Delta x \Delta p = h$

When Δx is maximum, the error in Δp, i.e., the uncertainty in momentum is minimum.

i.e., $(\Delta x)_{max} (\Delta p)_{min} = h$

or $$(\Delta p)_{min} = \frac{h}{(\Delta x)_{max}}$$

Substituting the values of Δx and h in the above equation, we get,

$$(\Delta p)_{min} = \frac{6.6\times10^{-34}}{10^{-9}}$$

$$= 6.6 \times 10^{-25} \text{ kg m s}^{-1}$$

We know that minimum error in momentum $(\Delta p)_{min} = m\ (\Delta v)_{min}$

Therefore, $$(\Delta v)_{min} = \frac{(\Delta p)_{min}}{m}$$

Substituting the values of $(\Delta p)_{min}$ and m in the above equation, we get,

The minimum error in velocity

$$(\Delta v_x)_{min} = \frac{(\Delta p_x)_{min}}{m} = \frac{6.6\times10^{-25}}{9.1\times10^{-31}} = 7.3\times10^{5} \text{ m s}^{-1}$$

The minimum uncertainty in velocity is 7.3×10^{5} m s^{-1}.

Objectives Questions

14.1 Electrons, protons and neutrons are the examples for _____.

The McGraw·Hill Companies

14.2 The wave function is denoted by _____.
14.3 Matter waves are not _____ waves.
14.4 Matter waves are used as _____ to guide the matter particles.
14.5 In Davission and Germer experiment, electron is selected as a particle due to _____ and _____.
14.6 Bragg's diffraction formula $2d \sin\theta = n$ _____.
14.7 The de Broglie wavelength of electron accelerated through a potential of 54 V is _____ A°.
14.8 Two forms of Schrödinger's equation are _____ and _____.
14.9 When the particle is inside the box, it's potential energy is equal to _____.
14.10 When the electron is inside the well, the wave function associated with the electron is _____.
14.11 The wave function corresponding to each E_n is known as _____.
14.12 Heisenberg uncertainty principle $\Delta x\ \Delta p$ = _____.

KEY

14.1 Matter particles
14.2 ψ
14.3 Electromagnetic
14.4 Pilot
14.5 Easy generation, detection
14.6 λ
14.7 1.67
14.8 Time independent, time dependent
14.9 Zero
14.10 Not equal to zero
14.11 Eigen functions
14.12 $\frac{h}{2\pi}$

Exercises

14.1 Determine the de Broglie wavelength of an electron accelerated through a potential difference of 200 eV.
14.2 Compute the de Broglie wavelength of a proton whose kinetic energy is equal to the rest energy of an electron. Mass of a proton is 1836 times than that of the electron.
14.3 Compare the de Broglie wavelength, speed and momenta of a proton and α-particle having the same kinetic energy.
14.4 Determine the energy levels and wave function for an infinitely deep two-dimensional potential well.
Given that $V = 0$ for $0 < \times < d$ and $0 < y < d$ and $v = \alpha$ for outside of this region.
14.5 Determine the energy levels for a particle with mass m in a one-dimensional potential box shown in Fig. 14.10.
Given that $V = V_1$ for $x \leq 0$, $V = 0$ for $0 < x <$ and $V = V_2$ for $x \geq d$

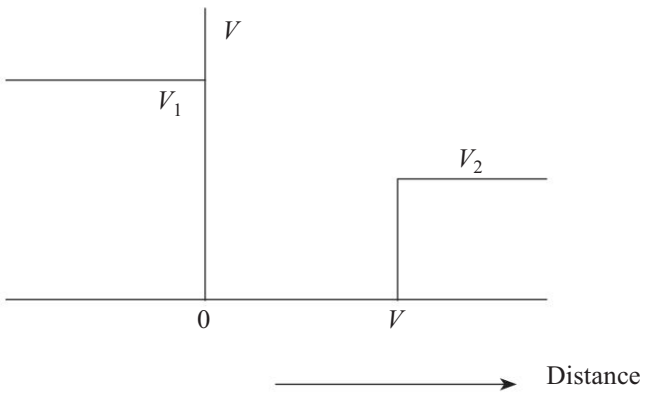


Fig. 14.10 *Energy levels of a particle*

14.6 Applying the uncertainty principle, show the non-existence of an electron inside the nucleus.
14.7 A particle is different at a certain angle and the uncertainty in the measurement of the angle is one second of the angle. Estimate the error in the measurement of its angular momentum.
14.8 An electron has a speed of 600 m s^{-1} with an accuracy of 0.005%. Determine the certainty with which one can locate the position of the electron.
14.9 Determine the uncertainty in velocity if the uncertainty in the location of a particle is equal to its de Broglie wavelength.

Short Questions

14.1 What is meant by dual nature?
14.2 Explain particle wave duality.
14.3 What is meant by matter waves?
14.4 Define de Broglie wave length.
14.5 Write down the de-Broglie matter wave equation? .
14.6 What are the ideas prompted de Broglie, to propose his concept about matter waves?
14.7 Define wave function.
14.8 What is meant by Schrödinger equation?
14.9 Write the Schrödinger time independent equation and explain the various terms.
14.10 Write the Schrödinger time dependent equation and explain the various terms.
14.11 Drive an expression for the de Broglie wavelength of an electron.
14.12. What is the physical significance of wave function?
14.13 What is meant by eigen function?

The McGraw·Hill Companies

14.14 Define Eigen value.
14.15 State and explain Heisenberg uncertainty principle.

Descriptive Questions

14.1 Explain the particle wave duality. Drive the equation for the de Broglie equation for an electron accelerated through a potential difference V.
14.2 What are the matter waves? Explain with neat sketch the existence of matter waves using Thomson's experimental method.
14.3 Explain with neat sketch the experimental verification of matter waves using Davisson-Germer experiment.
14.4 Derive the Schrödinger time independent and time dependent wave *equation*.
14.5 Explain the physical significance of the wave function ψ? Explain the *application* of Schrödinger wave equation to a particle in a box.
14.6 Explain Heisenberg uncertainty principle with suitable illustration.
14.7 Write note on the following
- Matter waves
- Particle in a box
- Heisenberg uncertainty principle

Chapter

Semiconductors

15

OBJECTIVES

- To explain the property of a solid material on the basis of band theory
- To study the classification of semiconductors such as intrinsic and extrinsic semiconductors
- To explain the arrangement of atoms in intrinsic semiconductors
- To explain the structural changes of extrinsic semiconductors when added with tri/ pentavalent element
- To study the principle behind LED and different types of LED's
- To explore the possible LED materials and their properties
- To understand the principle behind the passive display device, namely, LCD and the classification of LCD devices
- To study the characteristic properties of LCD
- To compare the relative merits of LED and LCD
- To review the applications of display devices

15.1 INTRODUCTION

Most electronic devices are made of semiconductors such as silicon and germanium. The applications of semiconductors and their devices are ever increasing in all fields of science and technology.Hence, knowledge on the classification, structures and properties of semiconductors has become very essential to identify the right materials for their applications. In the next chapters, electrical properties of materials will be studied in detail. Therefore, in this chapter, the classification, bonding and structure of solids are discussed in the first part; while in the second part,certain applications such as solar cells and display devices are discussed in detail.

15.2 CLASSIFICATION OF SOLIDS ON THE BASIS OF BAND THEORY

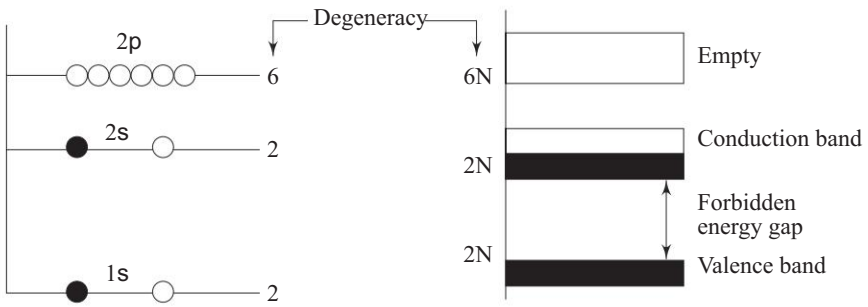
Based on the energy band structure, the arrangement of electrons and forbidden bands, solid materials are classified into the following three categories:

(i) Conductors
(ii) Insulators, and
(iii) Semiconductors

Let us discuss these materials briefly and move on to semiconducting materials.

15.2.1 Conductors

Materials which conduct electric current when a potential difference is applied across them, are known as conductors. In case of a conductor, the valence band is completely filled, while the conduction band is half filled, as shown in Fig. 15.1.
Therefore, when a small potential difference is applied to a solid material, it provides sufficient energy to the electron in the valence band to shift to the conduction band. Thus, the shifting of electrons from the valence band to the unfilled conduction band results in the flow of current in the material. Examples for some good conductors are copper, lithium, etc.



(a) Energy level diagram (b) Energy bands

Fig. 15.1 *Conductors–lithium atom*

15.2.2 Insulators

Solid materials which do not conduct electric current under normal conditions are known as insulators. In insulators, the valence band is completely filled and it has no electron in the conduction band. Further, the forbidden energy gap will be very high when compared with a conductor. The energy band diagram of an insulator (for example, ebonite) is shown in Fig. 15.2. Therefore, the energy required to shift an electron from the valence band to the conduction band to make electrical conduction is very high. Hence, it is not possible to provide enough energy by an ordinary electric field. However, one can achieve electrical conduction in an insulator with very high voltage, known as *breakdown voltage.*

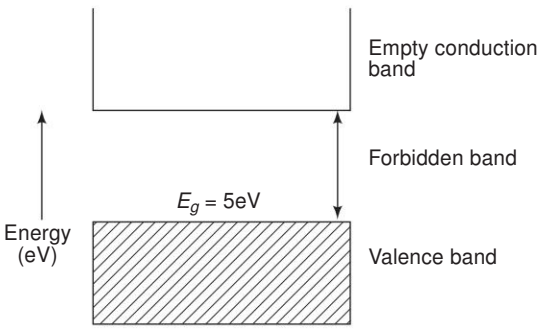


Fig. 15.2 *Energy bands in insulators*

15.2.3 Semiconductors

Semiconductors (for example, silicon or germanium) are materials whose electrical conductivity lies between that of conductors and insulators. The conductivity of semiconductors is of the order of 10^4 to 10^{-4} mho m^{-1}. The magnitude of the forbidden energy gap of a semiconductor lies in between the forbidden energy gap of insulators and conductors, as shown in Fig. 15.3.

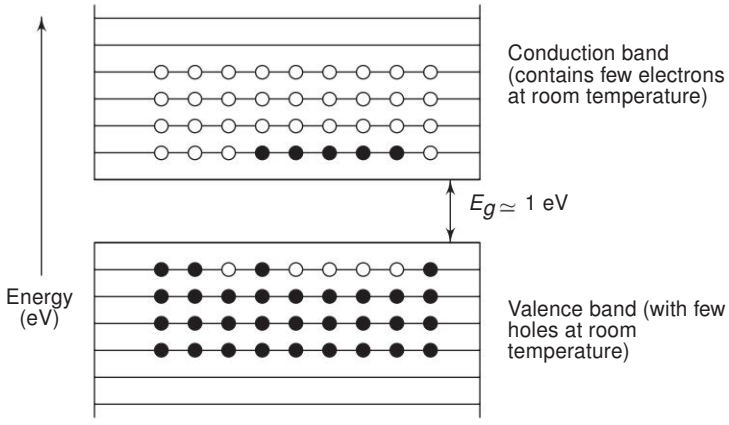


Fig. 15.3 *Energy bands in semiconductors*

Semiconducting materials, whether elemental, compound or oxide, all are crystalline solids in nature. Well-known semiconductors, such as silicon and germanium, are elemental semiconductors, while gallium arsenide (GaAs), cadmium sulphide (CdS), etc., are known as compound semiconductors. Some of the oxide semiconductors are Bi_2O_3, Te_2O_3, ZnO_3, Cu_2O, etc.

Let us discuss the different types of semiconductors and their applications in detail.

15.3 CLASSIFICATION OF SEMICONDUCTORS

Semiconductors are of two types and are classified on the basis of the concentration of electrons and holes in the materials

(i) Pure or intrinsic semiconductors,
(ii) Doped or extrinsic semiconductors

Let us discuss the above two types of semiconductors in detail.

15.3.1 Pure or Intrinsic Semiconductors

Highly pure semiconductors are called *intrinsic semiconductors*, which means that the concentration of electrons must be equal to the concentration of holes.

In order to understand the electrical conducting property in a semiconducting material, let us first consider the arrangement of atoms in a material, say, silicon (or germanium). Both germanium and silicon have four valence electrons as shown in Fig. 15.4.

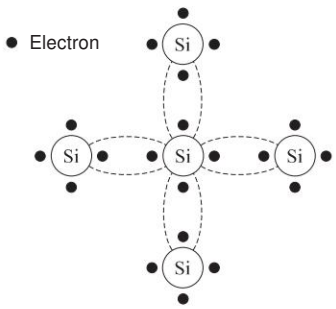


Fig. 15.4 *Two dimensional arrangement of atoms in silicon at T= 0 K*

Let us consider the crystal arrangement of silicon (or germanium). Silicon atom is represented by a circle, while the valence electron is marked by black dots. The central silicon atom is surrounded by four valence electrons constituting a covalent bond between the two atoms, due to the sharing of valence electrons.

If we consider the crystal structure at 0 K, all the valence electrons are engaged in forming covalent bonds with the neighbours. Therefore, no free electron is available. At this stage, the material does not conduct current due to the lack of mobile charges, i.e., at 0 K, the material behaves as an insulator. Electrical conduction can be more lucidly explained by considering the energy level diagram of silicon at 0 K, as shown in Fig. 15.5. At 0 K, the valence band is completely filled and there is no empty space in the valence band. Therefore, electrons can shift from the valence band to the conduction band through the forbidden gap. The shifting of electrons is not possible even for a large applied field strength at 0 K.

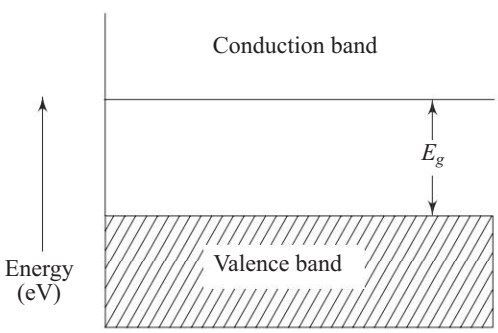


Fig. 15.5 *Energy-band-diagram of silicon at T = 0 K*

However, the thermal energy is sufficient to liberate an electron from the valence band at room temperature. When the temperature of the intrinsic semiconductor is

The McGraw·Hill Companies

raised, the atoms in silicon vibrate about their mean positions. This provides sufficient energy to the electrons due to which the breaking of the covalent bonds takes place as shown in Fig. 15.6. The broken electron is now said to be a free electron.

When an electrical field is applied, the free electron acquires sufficient energy and shifts from the valence band to the conduction band. This results in the creation of a hole in the valence band, as shown in Fig. 15.6. The free electrons move in the conduction band, while the holes move in the valence band. Thus, electrons and holes are moving opposite to each other, as shown in Fig. 15.7.

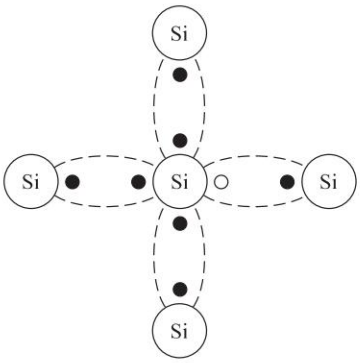


Fig. 15.6 *Two-dimensional arrangement of atoms in silicon at T= 0 K*

When a potential difference is applied across a silicon/germanium crystal, the electric force experienced by the electrons and holes are opposite. Thus, due to the opposite charges, the movement of electrons and holes give rise to an electric current in the same direction. The conductivity of germanium is higher than silicon.

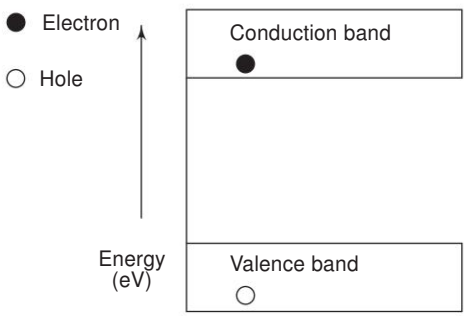


Fig. 15.7 *Energy band diagram of silicon at T > 0 K*

15.3.2 Extrinsic semiconductors

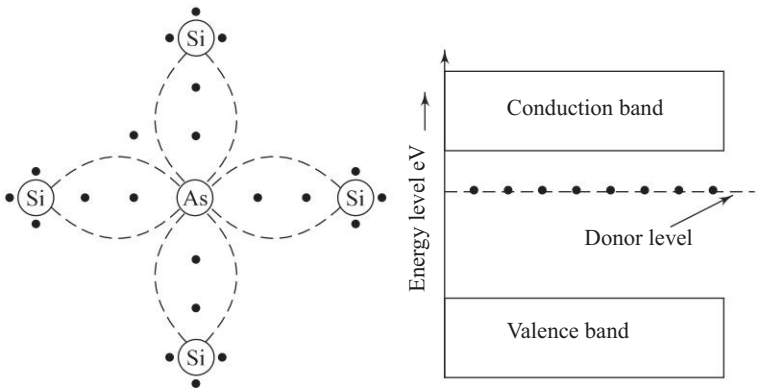
The application of intrinsic semiconductors is restricted due to their low conductivity. In electronic devices, high conducting semiconductors are more essential. The concentration of either electrons or holes in a semiconductor is increased depending upon the requirements in the electronic devices.

This can be carried out simply by adding impurities (one atom in 10^7 host atoms) to the intrinsic semiconductors. The process of adding an impurity to the intrinsic semiconductors is known as *doping*. The doped semiconductor is called *extrinsic semiconductor*. The concentration of electrons and holes are not equal in an extrinsic semiconductor.

Extrinsic semiconductors are classified into two categories based on the concentration of the charge carriers as given below:

(i) *n*-type semiconductors
(ii) *p*-type semiconductors

(i) n-type semiconductors when a pentavalent atom such as arsenic (antimony, bismuth, phosphorus) is added as a dopant to the tetravalent silicon atom, the arsenic atom will occupy one site of the silicon atom. Thus, out of five free electrons in arsenic, four electrons make covalent bonds with the four neighbouring silicon atoms and the fifth one is loosely bound to the silicon atom, as shown in Fig. 15.8.

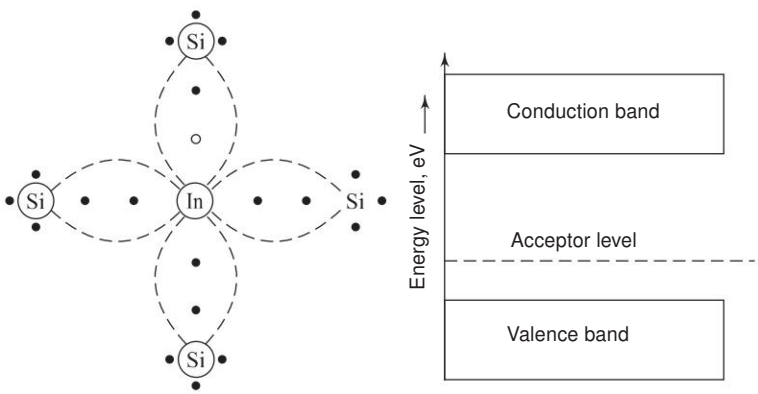


(a) With free electrons (b) Energy at donor level

Fig. 15.8 *Doping in n-type semiconductors*

The energy required to ionise the fifth electron is very less and hence, the thermal energy of the material shifts the free electron to the conduction band. Each arsenic atom contributes one free electron to the crystal and hence, it is called a *donor impurity*. In this type of semiconductor, the concentration of charge carriers (i.e., electrons) is more than that of holes. Therefore, these semiconductors are called *n-type semiconductors*. In an *n*-type semiconductor, electrons are the majority carriers while holes are the minority carriers.

(ii) p-type Semiconductors Instead of a pentavalent atom, the addition of a trivalent atom indium (In) to the tetravalent silicon atom, occupies the crystal site of the silicon atom as shown in Fig. 15.15.



(a) With vacancy (b) Energy at acceptor level

Fig. 15.9 *Doping in p-type semiconductors*

The McGraw·Hill Companies

The three valence electrons in indium make covalent bonds with the three neighbouring silicon atoms, whereas the fourth bond has an empty space known as hole due to the deficiency of one electron. Therefore, when a trivalent atom is added to silicon, it creates a hole in the valence band. The dopant (indium) accept an electron from the neighbouring silicon atom to form a covalent bond and hence, it is called an acceptor. The hole in the valence band moves freely and hence, the current flows through the material. This type of electrical conduction will take place only when the dopant valency is less than that of the parent atom. Such semiconductors are called *p-type semiconductors*. In a *p*-type semiconductor, holes are the majority current carriers and electrons are the minority current carriers.

15.4. SOLAR CELLS

Let us discuss in this section the general description, characteristics and materials of some important solar cells and their applications.

A solar cell is nothing but a *p-n* junction device based on the principle of photoelectric effect. It directly converts light into electricity and, hence, is known as a *photovoltaic cell*.

15.4.1 Description of a Solar Cell

The first solar cell made of silicon was developed in 1954. Even today, silicon cells are the only cells which have commercial status. Conventional silicon cells are thin wafers about 300 μm thick and 3 to 6 cm in diameter, sliced from a single crystal of *n*-type or *p*-type doped silicon (see Fig. 15.10).

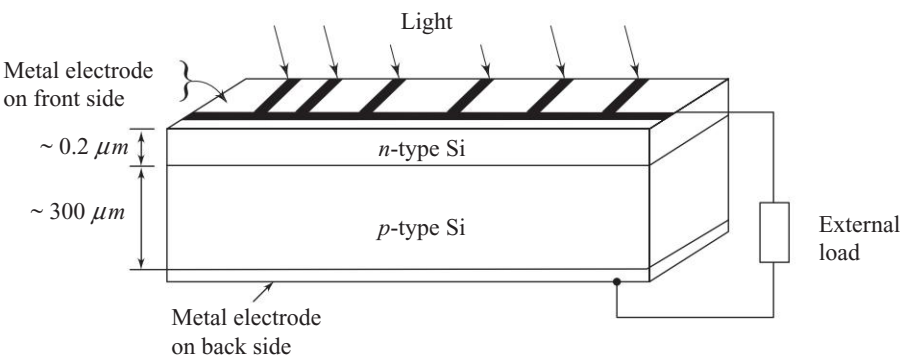


Fig.15.10 *Diagrammatic representation of a solar cell*

A shallow junction is formed at one end by the diffusion of the other type of impurity. Metal electrodes are made of Ti-Ag solder and are attached to the front and back sides The electrodes are in the form of a metal grid with fingers which permit sunlight to pass through it. On the back side, the electrode completely covers the surface. An antireflection coating of SiO having a thickness of about 0.1 μm, is also put on the top surface.

15.4.2 Working of a Solar Cell

When radiation falls on a solar cell, it is absorbed and pairs of positive and negative charges, called *electron-hole pairs*, are created. The positive and negative charges are separated because of the *p-n* junction. The direct current thereby produced is collected by the metal electrodes and flows through the external load.

(i) V-I Characteristics of a Solar Cell The volt-ampere characteristics of a solar cell are determined by connecting a decade resistance box and a voltmeter as shown in Fig15.11

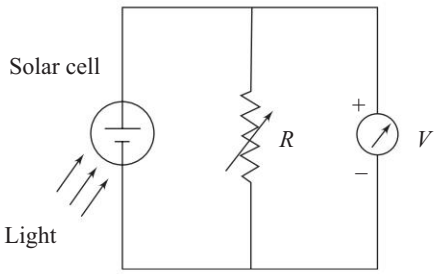


Fig15.11 *Volt-ampere characteristics of a solar cell*

A known intensity of light from a halogen lamp is made to fail on it. The resistance values are varied step by step and the corresponding voltages across the resistance box is measured using a voltmeter. From the known values of V and R, the value of I is determined using the relation, $I = V/R$. A graph is drawn between voltage and current as shown in Fig. 15.12.

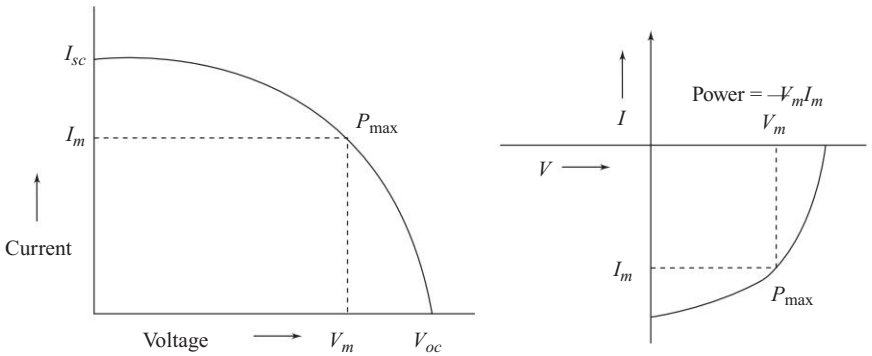


Fig. (15.12) *Characteristic curve*

The current I_{sc} shown in Fig. 15.12 is obtained by short-circuit in the two terminals of the solar cell and this current is known as *short-circuit current*. Similarly, voltage V_{oc} is known as *open-circuit voltage*. The product of these two quantities gives the ideal power of the cell. The maximum useful power is the area of the largest rectangle that can be formed in the *V-I* curve. The corresponding voltage and current are represented by V_m and I_m. Therefore, the maximum useful power is $V_m I_m$. The ratio of the maximum useful power to the ideal power is called the *fill factor*. A typical value of the V_{oc}, I_x and fill factor for a silicon cell are. respectively, 450 to 600 mV, 30 to 50 mA and 0.65 to 0.8.

The McGraw·Hill Companies

(ii) Efficiency of a Solar Cell Efficiency of a solar cell is defined as the ratio of the total power converted by the solar cell to the total solar power available for energy conversion.

$$\eta = \frac{\text{Maximum output electrical power}}{\text{Input optical power}}$$

$$\eta = \frac{P_{max}}{\text{Light intensity} \times \text{area of the solar cell}} \times 100 \qquad (15.2)$$

Equation (15.2) is used to calculate the efficiency of the solar cell and the same found to vary from 10 to 15 per cent.

(iii) Fill Factor The fill factor of a solar cell is given by the relation.

$$f = \frac{\text{Maximum output power}}{V_{oc} \times I_{sc}} \qquad (15.2)$$

The fill factor of the solar cell is found to be 0.65 to 0.8.

15.4.3 Solar Cell Materials

There are many materials which can be used for solar cells. The first parameter affecting the choice of the semiconductor for the photovoltaic application is the band-gap energy. The band-gap energy of a semiconductor should be comparable with the energy of the photons present in the solar spectrum, so as to absorb more number of photons. The energy gap of the materials which are used as solar cell is given in Table 15.1

Table 15.1 *Energy gap of solar cell materials*

S. No.	*Material*	*Energy gap (eV)*
1.	Si	1.11
2.	CdS	2.42
3.	GaAs	1.40
4.	$CuInSe_2$	1.01
5.	CdSe	1.74
6.	CdTe	1.44
7.	$CuInS_2$	1.50
S.	$CuInTe_2$	1.90
9.	ZnSe	2.60
10.	ZnTe	2.20

15.4.4 Classification of Solar Cells

Solar cells are classified into different categories, namely,

(i) *p-n* homojunction
(ii) *p-n* hetrojunction
(iii) Schottky barrier

(iv) Homojunction hetrostructure
(v) MIS (metal-insulator-semiconductor cell), and
(vi) SIS (semiconductor-insulator-semiconductor cell)

Let us discuss the gallium arsenide solar cells in detail

Gallium Arsenide Solar Cells GaAs solar cells perform well in monocrystalline form, compared to polycfystailine thin film solar cells. They have high absorption co-efficieni for visible light. They absorb all light in a surface layer of 1 μm thickness. The schematic diagram of a GaAs solar cell is shown in Fig. 15.13

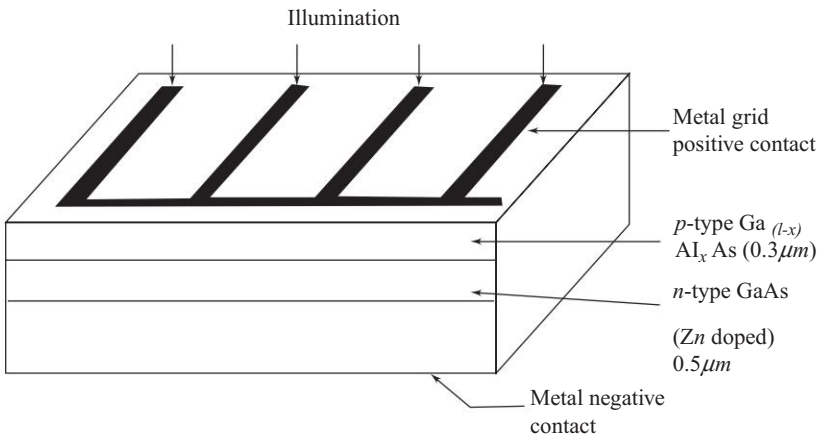


Fig. 15.13 *Schematic diagram of a GaAs solar cell*

In the case of GaAs solar cells, the surface recombination is very high and at the same time the carriers have short diffusion lengths. Therefore, the efficiency of the GaAs solar cells, is low. Low efficiency in this solar cells is also due to the lack of understanding the material fabrication parameters which are used to determine the minority carrier life time. Therefore, a new hetrojunction solar cell is made using GaAs/ $Ga_{(1-x)}$ Al_x As material with an efficiency of 19 percent. The junction at the interface of GaAs and $Ga_{(1-x)}$ Al_x. As does not affect the crystal structure very much and, thus, does not introduce recombination centres near the junction.

15.4.5 Application of Solar Cells

There are three broad categories for terrestrial photovoltaic applications: industrial, social and consumer applications.

(i) Industrial Applications Telecommunication PV systems are used effectively for the operation of telecommunication equipment in remote and difficult to access areas like mountain tops, islands and deserts.

Cathodic Protection To prevent corrosion of pipeJines, bridges, etc., a small direct current is impressed on the structures at regular intervals. This is called cathodic protection. For this, small PV panels may be used to provide this current efficiently.

Alarm Systems PV systems are also used to power railway signals, alarm systems, fog. fire and flood hazard warnings, traffic lights and highway telephones.

The McGraw·Hill Companies

Defense Equipment Defense equipment like mobile telephone, remote instrumentation, remote radar, water purifier, etc., are powered by PV systems.

Remote Aircraft beacons Remote radio and light beacons are powered by PV systems.

Automatic Meteorological Stations For recording the whether report at a frequent interval of time, meteorological stations are powered by PV systems.

(ii) Social Applications

(i) PV devices are used to provide electric power to remote villages
(ii) PV powered water pumps are being installed to provide potable waterf flow or plant
(iii) PV is used to power refrigerators at community health centres for keeping vaccines at low temperatures.

(iii) Consumer Applications Some consumer products, such as pocket calculators, watches, torches, garden lights, portable fans for cars and houses, radios, toys, electric fences, etc., are powered by PV solar systems

15.5 DISPLAY DEVICES

Recent advancements in the field of science and technology require information on the classification, basic principles, working mechanism, performance and applications of different display devices. Display devices are an output unit which give visual representation of data, i.e., display device is a device which produces the information on the screen. Generally, the display devices are classified into two different categories, namely, active and passive devices. The device which emits radiation on its own is known as an *active device*. On the other hand, the device which modulates incident radiations to provide necessary display information is known as a *passive device*. Display devices play an important role in many areas such as electronic instruments, advertisements, healthcare, home products, etc. The principle, operation and applications of both active and passive display devices are given in detail in the following sections.

15.6 ACTIVE DISPLAY DEVICES

An active display device is a device which emits its own radiation when a required power is given to the device. The examples of active display devices are

(i) Cathode ray tubes
(ii) Light emitting diodes

The principle, working and application of light emitting diodes is given briefly in the following heading:

15.6.1 Light Emitting Diode

A Light Emitting Diode (LED) is a junction diode which emits light when it is forward biased. The principle behind LED is electroluminescence. The process of

injecting electrons and holes into the *n*-type and *p*-type materials is known as *injection electroluminescence*.

Principle

Electrons are charge carriers in a semiconductor and it absorbs energy when an electric energy is applied. The total energy of the charge carrier electrons increases and hence,they are excited to the higher energy state (E_2). The excited electrons in the higher energy states stay there only for a few seconds, and after the mean lifetime eventually return back to the ground state energy level (E_1). During this process, a spontaneous emission of the radiation of light takes place. The energy of the emitted photon ($h\nu$) is equal to the energy of band gap (E_g) of materials, i.e.,

$$E_g = h\nu \tag{15.4}$$

where h is the Planck's constant and ν, the frequency of the emitted radiation.

Substituting the value of $\nu = \frac{c}{\lambda}$ in Eq. (15.4), we get,

$$E_g = \frac{hc}{\lambda} \tag{15.5}$$

where c is the velocity of light and λ, the wavelength of light.
Equation. (15.5) can be written as,

$$\lambda \sqrt{\frac{hc}{E_g}} \tag{15.6}$$

Equation. (15.6) gives wavelength of emitted photon.
Substituting the values of $h = 6.62 \times 10^{-34}$ J s and $c = 2.998 \times 10^8$ m s^{-1} in Eq. (15.6), we get,

The wavelength of emitted photon $\lambda = \frac{6.62 \times 10^{-34} \times 2.998 \times 10^8}{E_g}$

$$\lambda = \frac{12400}{E_g (ev)} \text{Å} \tag{15.7}$$

Equation. (15.7) indicates that the wavelength of emitted photons depends on the energy gap in the semiconductor. Thus, the energy gap of a semiconductor plays a major role in selecting a suitable material for LED applications. The type of material is also an important factor to be considered for the materials selection.

Operation

LED is a forward biased *p-n* junction (Fig. 15.14). When it is forward biased suitably, it emits visible light. During the forward biasing, the charge carriers,

The McGraw·Hill Companies

namely, electrons and holes, are injected respectively into the anode and cathode regions. The recombination of the charge carriers, namely, the electrons from the *n*-side and the holes from the *p*-side, takes place at the junction. During the recombination, the difference in the energy is given up in the form of heat and light radiation, i.e., photons. The energy of light radiation depends on the strength of recombination. Thus, the diode current controls electroluminous efficiency of the LED. The emitted light is very small in intensity and is of the order of microampere range. The emitted light colour depends on the types of materials used. For example, materials like GaAs, GaP and GaAsP are used to produce infrared, red or green and red or yellow colours.

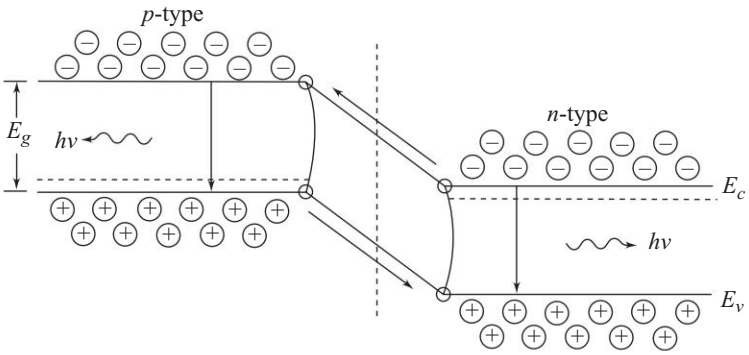


Fig 15.14 *p – n Junction spontaneous emission*

15.6.2 LED in Electrical Circuits

The symbol used for LED is shown in Fig 15.15a. Similarly, a typical forward biased LED circuit and its symbols is shown in Fig. 15.15 b and c.

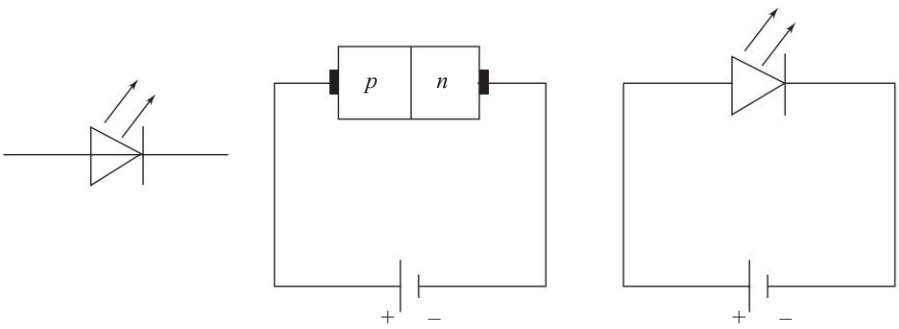


Fig. 15.15 *Light emitting diode*

The cross - sectional view of gallium arsenide phosphide (GaAsP) semiconductor and its symbol is shown in Fig.15.16. Gallium arsenide is used as substrate to grow

The McGraw·Hill Companies

a thin layer of *n* type gallium arsenide phosphate through the epitaxy method. Above the *n* type layer, a very thin layer of p-region is diffused. Similarly, very a thin layer of *p*-region is diffused into the epitaxial layer. The anode is laid down on the *p*-region in the shape of a comb or other similar configurations. In order to allow more central surface area for the light to escape, the provisions for metal oxide connections are given at the outer edge of the *p*-layer.

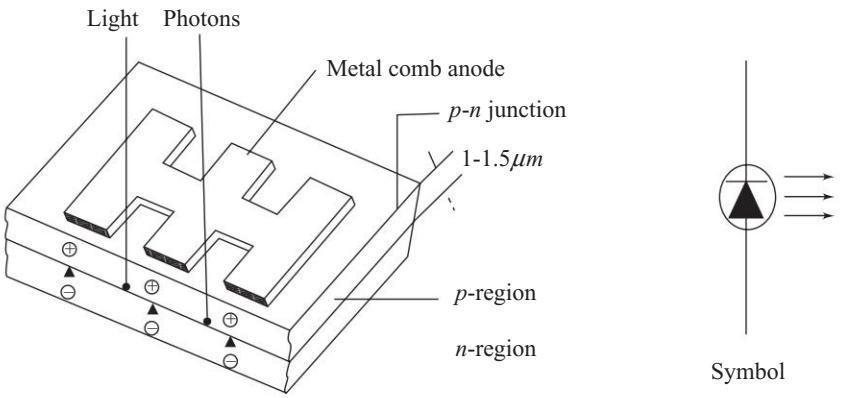


Fig. 15.16 *Gallium arsenide phosphate semiconductor—sectional view*

15.6.3 Types of LEDs

LEDs are available as discrete components, packaged for specific applications, or as high intensity light source products depending on the applications.

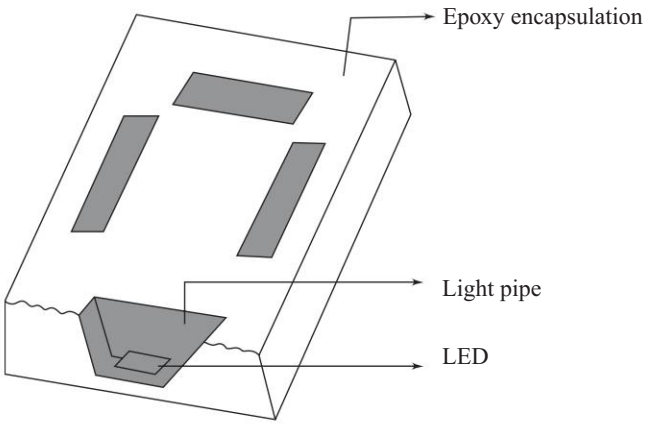
(i) Discrete LED The popular standard type LED is in the form of a small round dome epoxy encapsulation. LEDs are also available in a variety of other shapes and sizes. In particular, rectangular, square and triangular LEDs are available for applications like panel indicators. LED components like axial leaded LEDs, bar graph displays, tri-coloured RGB LEDs, and surface mount technology components are also available in the market. The salient features of LEDs include low current, high brightness, high voltage, flashing and multicolour variants.

(ii) Alphanumeric Displays In 1967, LED displays are introduced by comprising 7 or more individual LED's. In most of the electrical appliances and other applications, these types of displays are used. The displays are arranged either by using a multi-segment LED display or a dot-matrix LED display. The alphanumeric character is displayed by means of using array LED's and seven-segment LEDs. The schematic representation of 7×5 array and seven segment LED arrangements for the alphanumeric display are shown respectively in Fig.15.17a and b. In 7×5 array LED display, the illuminated points are shown as dark points while non-illuminated points are shown by white points. A typical representation of decimal 5 read-out is shown in Fig. 15.17a. The seven- segment display is energised by applying a potential ranges from 1.5 to 3.3 V with a constant current of 50 to 100 mA. The above display can produce different colours such as red, blue,

The McGraw·Hill Companies

green and orange. In most of the display applications, seven-segment displays are commonly used.

(a) 7×5 array of LED's



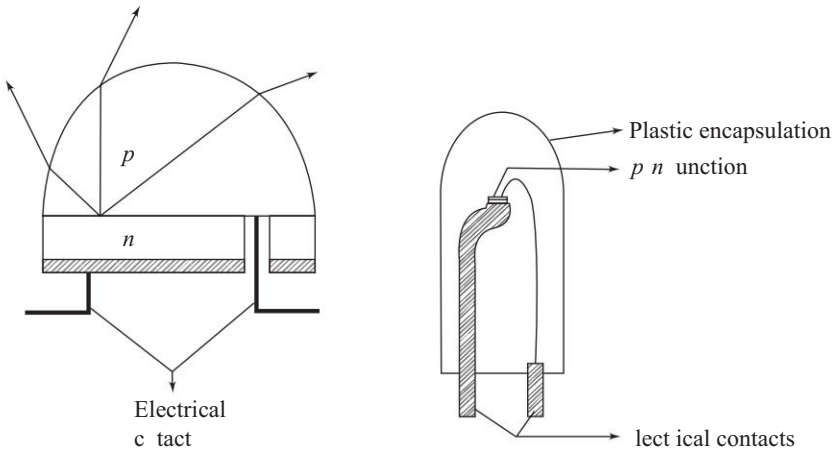
(b) Seven-segment LED

Fig. 15.17 *Alphanumeric display*

(iii) LED Clusters and Lights In order to produce brighter and high power lights, groups or clusters of LEDs are used. The total brightness of the group or cluster of LED may failure even if there is a breakdown of one or two individual LEDs with in the group. However, the application of LED technology is increasing in the market due to the high density light yield.

15.6.4 LEDs—Reduction in Reflection Loss

Generally, when an LED emits light, there is a loss of light energy due to the reflection losses which arise in LEDs. There are two methods by which one can reduces the loss due to reflections. One way of reducing loss of reflection is by making the LED with a hemispherical dome as shown in Fig. 15.18a. In this method, the *p*-type material is used to obtain a hemispherical domelike structure. Thus, the dome forms a semiconductor/air interface. When the light rays strike the semiconductor/air interface at an angle less than the critical angle, they emerge out without any loss. The energy loss is further reduced by replacing the dome by means of plastic encapsulation as shown in Fig.15.18b. The replacement of semiconductor/air interface in a hemispherical dome by a plastic/air interface further reduces the loss of energy due to reflection. In most of the commercial LEDs, the plastic encapsulation techniques are employed to improve brightness of the light energy.



(a) Hemispherical dome (b) Plastic encapsulation

Fig. 15.18 *LEDs— reduction in reflection loss*

15.6.5 LED Materials

The material selection for the fabrication of LED is essential due to the emission of different colours (wavelength). The emission of light with different colours depends on the semiconductor materials. The essential factors such as energy gap, both *n*- and *p*-type with efficient pathways are to be considered for the fabrication of LEDs. The wavelength of the light emitted depends on the energy gap of the semiconducting materials. For example, infrared and far infrared light are emitted by lower band–gap materials, while visible light is emitted by larger band gap (≈ 2 eV) materials. Due to the high melting point, high resistivity and low structural stability, the high band gap materials have advantages in LED applications than low band gap materials. The commercial by available materials for LED applications along with their properties are briefed in the following headings:

(i) Gallium Arsenide Gallium arsenide (GaAs) is an III–V compound semiconductor consisting of the element gallium from the column III and the element arsenic from V of the periodic table. It is a direct band gap semiconductor. At room temperature,the band gap energy is 1.44 eV. It has high luminescent efficiencies due to the readily formed *p-n* junction. The mass radiation is due to the band -to- band transitions. Hence, the emitted radiations are reabsorbed. The efficient diodes are obtained by doping silicon, either as a donor or acceptor. The incorporation of silicon is used to replace either Ga or As depending on the doping. Thus, the main radioactive transition is based on the recombination between conduction band and complete acceptor level. In such a case, the band-to-band transition is not effective due to the insufficient energy of radiation and hence, re-absorption is very low. Generally, the wavelength of light emitted from GaAs LED lies between 910 and 1020 nm.

(ii) Gallium Phosphate Gallium phosphate (GaP) is an indirect band–gap semiconductor with $E_g = 2.26$ eV. These semiconductors are converted into an efficient light emitter by doping impurity in the band gap. Generally, it emits red and green

The McGraw·Hill Companies

light. There is no band-to-band radioactive transition. Elements such as N and Bi from group V, have the same valence as that of P and are doped. It is useful to get the radioactive recombination rather in forming donor or acceptor states. When the doping concentration of N is greater than 10^{14} cm^{-3}, it results in a shift in the emitted light radiation from 550 nm (green) to 590 nm (yellow). This observed shift in the light radiation is due to the defect level in the band gap. A simultaneous doping of Zn and O in GaP results in light radiation at 640 nm (red).

(iii) Gallium Arsenide Phosphide Gallium arsenide phosphide ($GaAs_{1-x}P_x$) is an important semiconductor for LED fabrication due to the change in band gap energy. The band gap energy depends on the composition of x. When the composition is low ($x < 0.45$), it gives a direct band gap semiconductor, while when it reaches $x = 0.45$, a change in the direct baned gap to indirect band gap takes place. Thus, the energy gap at $x = 0.45$ is 2.1 eV. One can produce light emission for different colours such as red, orange and yellow by changing the composition of x and the doping content N.

(iv) Gallium Aluminium Arsenide Gallium aluminum arsenide ($Ga_xAl_{1-x}As$) is a new semiconductor used for the fabrication of LEDs. The light emitted by these materials lies in between red and near infrared region. The intensities of the light are very high and hence, it is called *super bright LEDs*. The standard type GaAsP LEDs are replaced by GaAlAs LEDs. The high-intensity radioactive emission is due to the formation of n-type $Ga_{0.3}Al_{0.7}As$ and p-type $Ga_{0.6}Al_{0.4}As$ semiconductor by doping Zn. The emission is achieved at 650 nm with high efficiency. The doping of Si with a change in Al composition is possible due to increase the band gap and hence, emission rate from 870 to 890 nm.

(v) Indium Gallium Arsenide Indium gallium arsenide ($In_{0.53}Ga_{0.47}As$) semiconductor is used for fiber optical communications. The light emitted by the LEDs is in the range of 1.1 to 1.6 μm. This wavelength range is more suitable for optical communications since the spectral region of optical fibers lies in the same range. The general characteristic properties of common LED materials are given in Table 15.2.

Table 15.2 *Properties of LED materials*

Materials	*Band gap Type*	*Band gap (eV)*	*Peak wave length (nm)*	*Colour*
GaP:ZnO	Indirect	1.78	700	Red
GaAsP	Direct	1.99	650	Red
GaAsP: N	Indirect	1.95	630	Red
AlGaAs	Direct	1.91	650	Red
AlGaP	Direct	2.00	620	Red
AlInGaP	Direct	2.08	595	Amber
GaAsP:N	Indirect	2.10	585	Yellow

AlInGaP	Direct	2.18	570	Yellow–Green
GaP:N	Indirect	2.20	565	Yellow–Green
GaP	Indirect	2.26	555	Green
SiC	Indirect	2.9 – 3.05	480	Blue
GaN	Direct	3.5	450	Blue

15.6.6 Advantages

The following are the advantages of LEDs**:**

- LEDs are used to produce more light per watt than incandescent bulbs.
- It is easy to operate than through battery power.
- LEDs are highly protected as they are built inside solid cases.
- The lifespan of LEDs is more than 100,000 hours.
- LEDs light up very quickly.
- A typical red indicator LED will achieve full brightness within few microseconds.
- The response time of LEDs used in communication devices is much faster than normal sources.

15.7 PASSIVE DISPLAY DEVICES

Non-emissive display devices are known as passive display device,s i.e., they modify the parameters of light, generated by an external source. The modification may take place in optical parameters such as optical path, path length, absorption, reflection, scattering or a combination of all these parameters. Each modulation will have its own effects. For example, when wavelength is modulated, then the display is colored. Materials such as liquid crystals, electrophoresis and electrochromium are used to achieve the electro optic effects. Based on the properties of electro optic materials, one can select the right material for passive display devices. The Liquid Crystal Display (LCD) is widely used due to the variety of modulation effects. The properties of passive display materials are given in Table 15.3. In the following heading, the principle, types, operation and applications are given in detail.

Table 15.3 *Properties of passive display materials*

Effect	*Liquid crystals*		*Electrophoresis*	*Electrochromism*
	Hydro-dynamic	*Field effect*		
Operation in transmission	T	T	-	-
Operation in reflection	R	R	R	R
Viewing angle	Narrow	Medium	No restriction	No restriction

Operating voltage (V)	15–40	5–10	50–100	1–2
Power consumption continuous operation mW cm^{-2}	0.1	0.002	0.4	-
Switching speed	0.05	0.2	1	0.001

T - Transmission R-Reflection

15.8.1 Liquid Crystal Display

Liquid crystal is a kind of substance which lies in an intermediate state between a conventional liquid and a solid state. In a restricted temperature range, the organic materials exhibit a state of matter known as *liquid crystal state*.A liquid crystal becomes crystalline solid at low temperature, and at high temperatures, it turns into a clear isotropic liquid. Thus, the properties of liquid crystals are intermediate between liquids and solids. LEDs generate lights, while liquid crystal displays do not generate light, rather they reflects light. LCD reflects only a part of the surrounding light, while the other part of the light is absorbed by the display. In view of the above reasons, LCDs are used in lighted areas. LCD displays are used in watches, display boards, TVS, computers and packet calculators.

15.8.2 Principle of LCD

Liquid crystal display is based on two principles, namely, polarisation and twisted nematic property of the liquid crystal.

Polarisation

The process of filtering certain unwanted light wavelengths employing glass with parallel thin impenetrable black lines is known as polarisation. When the wavelength of the incident light is parallel to the parallel lines, it will pass through the glass, while the other wavelengths are stopped. Consider two polarised glass sheets are placed one in front of the another with perpendicular polarisation. When light is incident on the glass sheet, it will not pass through it.

Twisted Nematic

Consider that the liquid crystal is sandwiched between two sheets of polarised glass. The molecules in the liquid crystal arrange themselves in a twisted form. Under this condition, when a light is entering into the first sheet of glass, it comes out of the glass as a polarised light. Thus, the polarised light enters into the twisted form of molecules. The entered light waves are twisted along with the orientation of liquid crystal molecules. The light waves which are entering the molecules are in line with the polarisation of the second glass sheet. Therefore, the region in which the light passes through the glass sheet is displayed. On the other hand, a liquid crystal does not pass the light when it is untwisted. Therefore, the light is not passed through the second glass sheet, and hence, it creates a dark region on the display. The position of the liquid crystal is controlled by applying a necessary charge.

15.8.3 Classification of Liquid Crystals

Liquid crystals consist of three different types of states, namely, nematic, cholesteric and smectic states. In most of the electro optic devices, the first two types of states are used for practical display devices.

(i) Nematic Liquid Crystals In nematic ordering, the molecules known as *directors* are arranged parallel to each other. Rod or disclike molecules oriented with respect to each other as well as the liquid crystal surface are known as *directors*. In other words, a director is a unit vector pointing along the orientation of molecules in a small volume. In nematic ordering, molecules are free to move along each other due to the liquid properties. Generally, two benzene rings are linked with a central group to form the nematic phase. In nematic ordering, the parallel alignment of molecules (or directors) is shown in Fig.15.19. Thus, it acquires liquid properties. In nematic crystals, two benzene rings are linked with a central group. The chemical formula for the nematic crystal molecule 4–methoxybenzylidene–4–butylanaline (MBBA), is given below,

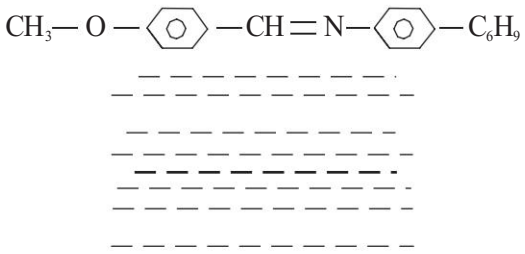


Fig. 15.19 *Liquid crystalline phases – Nematic ordering*

The schematic arrangements of nematic liquid crystal molecules are shown in Fig.15.19. The MBBA shows liquid crystal behaviour in the temperature range of 293 to 320 K.

(ii) Cholesteric Liquid Crystals A small modification in the nematic crystal results in a cholesteric liquid crystal. The structure of a cholesteric liquid crystal is derived from cholesterol molecules. The molecules in the crystal are optically active, and hence, have the ability to rotate the plane polarisation of light passing through it. The structure of the cholesteric liquid crystal is shown in Fig.15.20. Rod or disclike molecules are used in liquid crystal phase.

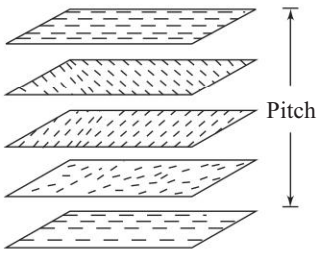


Fig. 15.20 *Liquid crystalline phases – Cholesteric ordering*

The McGraw·Hill Companies

A cholesteric liquid crystal consists of a large number of vertical sheets stacked one above the other as shown in Fig 15.20. The molecules in the respective sheets are aligned in the nematic state, keeping the orientation of the molecules in each sheet twisted when compared with the molecules in other sheets. Thus, the director directions of the sheets represent a helical twist in the molecules. The distance between any two successive planes is known as *pitch.*

When a light beam is propagated parallel to the helical axis, it undergoes phenomena like reflection, rotation and change of polarisation based on the wavelength of the incident light to pitch ratio. Similar to the direction, the elliptically polarised light is rotated above the helix axis when wavelength to pitch ratio is very small. The cholesteric liquid crystals show colour effects. When a white light is incident on it, it shows as a strongly coloured light. The pitch of the crystal and colour of reflected light depends on temperature. Under this condition, it obeys the Bragg's reflections, i.e., $p = n\lambda$ (where n is equal to 1, 2, 3, etc.). The cholesteric liquid crystals consisting of mixed materials are used for visual display devices like a surface temperature measurement thermometer. For example, a mixture of material with 20% cholesterol carbonate and 80% cholesterol nonanoate scatters bright yellow and deep violet lights respectively at wavelengths of 0.55 μm at 298 K and 0.39 μm at 318 K.

(iii) Smectic Liquid Crystals A smectic liquid crystal is also called *lyotropic liquid crystal* (lyo means liquid). An example of smectic liquid crystals is soap water. When soap is melted, it forms liquid crystals. The molecules in liquid crystals are arranged parallel to each other with irregular spacing as shown in Fig. 15.21. A layered structure is known to exist only in smectic structure. Thus, it gives rise to a strong lateral force between molecules within layer and weak interactions between layers. Therefore, a smectic liquid crystal is slippery.

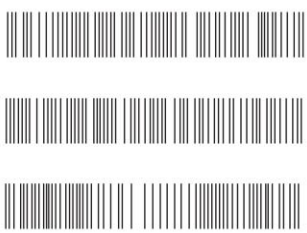


Fig. 15.21 *Smectic liquid crystal—molecular arrangements*

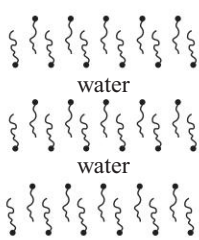


Fig. 15.22 *Lyotropic smectic liquid crystal–molecular arrangements*

Similarly, molecules which are arranged in a lyotropic liquid crystal also exhibit similar property. The arrangement of molecules in a lyotrpic liquid crystal is shown in Fig 15.22. The molecular layers are separated by a thin film of water. The layers can stack one above other and hence, give properties of soap. An example of lyotropic liquid crystals are body fluids like mucus and saliva.

15.9 DIFFERENT MODES OF LCD

We know that LCD depends on electro optic effects. Based on the effects, LCD operations are categorised as,

(i) Dynamic scattering
(ii) Twisted nematic
(iii) Phase change
(iv) Memory effects

A brief explanation about the different modes of LCD is given in the following headings:

15.9.1 Dynamic Scattering

In dynamic scattering, positive nematic materials with homogeneous alignment of molecules having low impendence are used. The material is purely transparent when there is no external field. The molecular ordering depends on the external field. The orientation of positive homogeneous molecules in LCD on application of electric field is shown in Fig. 15.23. For example, when the field is lower than E_c, the molecular ordering is not affected. However, when the field reaches E_c, the molecules tend to align parallel to the field. When the field is extremely higher, most of the molecules align along the field direction. At very high field, the existing current carries the movement of the liquid. As a result, the incident light gets a strong scattering. Thus, it gives a frosted appearance to the material. The dynamic scattering is possible only at low frequency, while at higher frequency, it is not possible due to the electrolysis of the liquid.

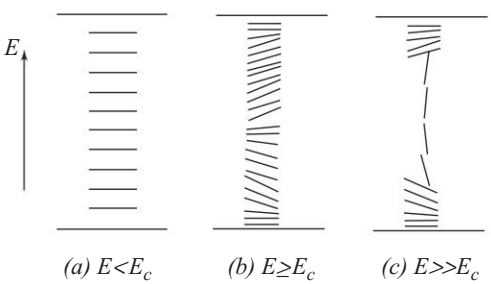


Fig. 15.23 *Orientation of positive nematic material with homogeneous ordered– applications of electric field*

LCD is used as an effective scatterer of white light under a suitable external field. However, in absence of a field, it is purely a transparent liquid crystal. The major

The McGraw·Hill Companies

disadvantages of LCD are poor contrast due to reflection, short lifetime and huge power consumption when compared to twisted nematic or field effect display.

15.9.2 Twisted Nematic Mode

In this method, dielectric properties of the molecules play a major role on the transport property of light. Consider that the molecules are arranged in a manner as shown in Fig.15.24a, during the absence of the field. When an external field is applied ($E >> E_c$), the molecules aligne themselves in the field direction as shown in Fig. 15.24b. The twisting of the molecules is eliminated under the influence of the field. When a light is incident on the LCD, it does not have effect on the incident polarised light. The portion in which the twisted nematic LCDs action takes place appears as a block in the ON condition. On the other hand, a shiny surface is obtained on the other faces. These kinds of LCDs help to produce frosty white characters on a dark background.

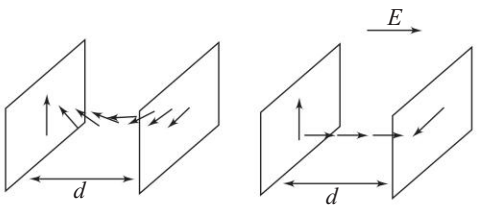


Fig. 15.24 *Arrangement of molecules in liquid crystal operating in twisted nematic mode*

15.9.3 Phase Change Effect

In this mode, cholesteric material with positive dielectric anisotropy and high resistivity is used. In this method, with a proper biasing, the cholesteric planes are maintained perpendicular to the cell walls. When a small field is applied, the widely scattered liquids cause the helix to unwind, leading to the arrangement of molecules in the direction of the field. Thus, the LCD is more nematic and highly transparent.

15.9.4 Memory Effect

Cholesteric material with negative anisotropy and low resistance are required to achieve this effect. The functions are very similar to dynamic scattering. The transition from the scattering to the transparent state is very slow.

15.10 LIQUID CRYSTAL DISPLAY SYSTEM

It consists of two glass plates, each attached with a crossed polaroid as shown in Fig. 15.25. On both sides of the glass plate, silicon monoxide is coated. The distance of separation between the coating surfaces is typically 10 μm. The liquid crystal is sealed in between the two glass surfaces. Thus, the liquid molecule has a transition in between the two glass plates. LCD devices are usually constructed in the form

of sandwich cells. The sandwich cells consist of two conducting glass plates made up of SnO_2 by coating on both sides of the plate by In_2O_3, etc. A thin layer of liquid crystal of the order of 3–50 μm is placed between the two conducting plates. The cell is perfectly sealed in order to avoid chemical reactions by interacting with the environment.

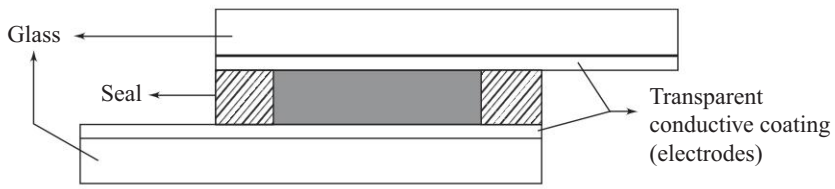


Fig. 15.25 *Liquid crystal display*

The lifetime of LCD is very high when the device is operated under ac conditions. The liquid crystal is operated under two modes, namely, reflective and tranmissive mode. In the reflective mode of operation, the incident light gets reflected back from the surface of the reflective aluminum film which is placed near the glass plate. In the transmission mode, the light is transmitted through the cell. There are two types of LCD displays, namely, dynamic scattering and the field effect type. Letters on a dark background are produced in *dynamic scattering* while in *field effect*, black letters on a silver background are produced.

15.10.1 Twisted Nematic Liquid Crystal Display

A positive nematic material with parallel alignment of molecules to the field directions is used in Twisted Nematic Liquid Crystal Display (TN-LCD) or field effect display. The liquid crystal cell shown in Fig. 15.25 is placed in between the two polarising filters known as polaroid (I and II). The polarizing directions of the filters are made very near to the molecular ordering direction. The arrangements of the filters and the LCD are shown in Fig. 15.26.

In the absence of an electric field, the molecules in LCD undergo a gradual twist. When an unpolarised light is incident on the first polaroid, it gives a polarised light. When the polarised light is incident on the cell, the twisted molecules rotate the plane polarised light through 90°. Thus, it gives horizontal polarised light and it passes through the second polaroid. In the absence of a field, it gives a clear bright cell. The action of an LCD is shown in Fig. 15.26a.

On the other hand, when a field is applied, the molecules are aligned in the field directions. Under this condition, when a light is incident on the first polariser, it is passed into the liquid medium. Therefore, the polarised light is simply absorbed by the liquid molecules. Therefore, no light is reflected back from its original path and hence, the liquid medium appears dark. The corresponding display action of an LCD is shown in Fig. 15.26b. A reflector is placed behind the second polaroid in case of reflective mode of operation. It rotates the incident polarised light through 90° similar to the light when it passes through the cell. The polarised light passes

The McGraw·Hill Companies

through the second polariser and then, it is reflected back along its path. Therefore, it appears bright.

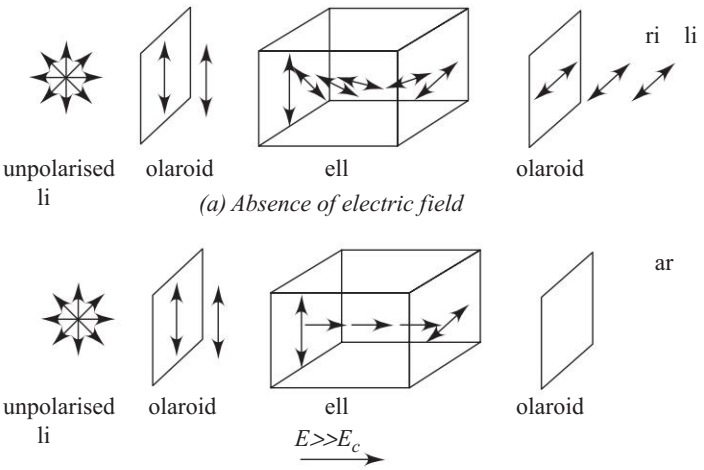


Fig. 15.26 *Twisted nematic liquid crystal display function*

It is clear from the above study that during the absence of a field, LCD reflects the incident light and appears as bright. During the presence of a field, light is not passed through the second polariser due to the absence of 90° twist. Hence, there is no reflected light and the, device appears as dark. The amount of light reflected from LCD depends on the applied voltage. In LCD, the amount of light reflected as a function of applied voltage is shown in Fig.15.27. The reflected light remains constant up to the critical voltage V_c . The critical voltage V_c is equal to $E_c d$, where d is the cell thickness. When the applied voltage is increased, an exponential decay is obtained up to the saturation voltage V_{sat}.

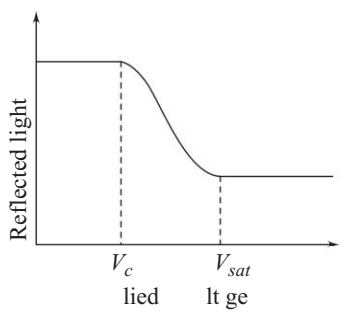


Fig. 15.27 *Reflected light versus applied voltage*

The turn on time T_{on} and turn-off time T_{off} are important factors which are considered for the operation of an LCD device. In twisted nematic LCD, the T_{on} is directly

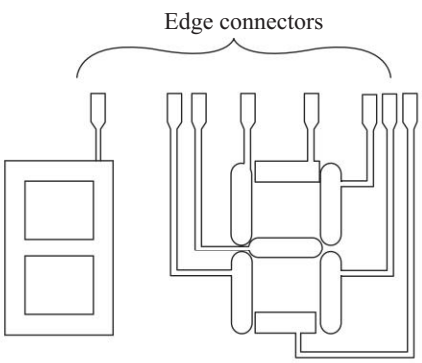
proportional to average viscosity η and inversely proportional to $\Delta\in E^2$, where ∈ is dielectric anisotropy and E, the electric field. The electric field is has more influence on T_{on} than Δ∈ and η. Similarly, the T_{off} is directly proportional to λd^2, where λ is the twist viscosity coefficient. The nature of mode of display and transformation temperature of few a LCDs is given in Table 15.4.

Table 15.3 *Liquid crystals operating mode and transition temperature*

Liquid crystals	*Mode*	*Transition temperature K*
n-p-Methoxy Benzylidene *p-n* Butyl-Aniline (MBBA)	Nematic	293.4
4 Methoxy- 4-*n*-Butyl-Azox-Benzene (MBAB)	Nematic	315.8
n-p-Ethoxy Benzyliden *p-n*- Butyl-Anidene (EBBA)	Nematic	308.4
Cholesteryl-Erucate (CE)	Cholesteric	283.6
Cholesteryl Nonanoate (CN)	Cholesteric	351.9

15.10.2 Numeric Display

The commonly used simple format for numeric display is the seven segment display. It is based on the dynamic scattering method. It consists of two glass plates coated with a transparent conductive material like SnO_2. These two glass plates are separated through a small distance of nearly 10 μm. The liquid crystal material is filled in between the glass plates. The conducting coating available at the lower electrode is used as common electrodes. The upper electrodes are energised through latch and driver unit separately. The schematic representation of a seven–segment display is shown in Fig. 15.28.



a) Lower electrode b) Upper electrode

Fig. 15.28 *Seven segment display—dynamic scattering display*

When the electric field is absent, the liquid appears as transparent. On the other hand, when an electric field is applied, the interaction of the doped ions with molecules takes place. Thus, the molecules are oriented in the field direction. Therefore, the incident light on the liquid gets scattered and hence, the liquid appears white in colour. The required electrical field is applied to each segment employing external devices. The power required to generate the electric field is very low. Thus, the required display either in dots or segments is made with the application of electric field by designing the glass plate with a conductive layer on one side.

The LCD is not suitable for use in matrix displays mainly due to the dependence of the threshold voltage, the flow response on the operating temperature and the flow response time.

15.10.3 Advantages and Disadvantages of LCDs

The following are the advantages of LCDs:

- LCD displays are used in conventional televisions and computer screens.
- LCDs are much lighter in size and consume less power.
- It provides an opportunity for the development of many portable devices.
- LCDs improve the quality of the image and meet the increasing demand for quality display.
- LCDs are replacing CRT screens in many household appliances owing to its lower price.
- It can be viewed over a wide range of lighting conditions.
- The device is very small in thickness.

The following are the disadvantages of LCDs:

- The decay time of LCD is very small. Therefore, the segments slowly fade when the power is switched off.
- It requires an external illumination and is invisible in darkness.
- The angle of view is limited.
- The intensity of the light produced is very small.
- Its functions are limited in the temperature range of 293 to 333 K.
- It is highly sensitive to environmental conditions like temperature, interference due to electromagnetic field, etc.

15.11 COMPARISON BETWEEN LEDs AND LCDs

A comparison between the LEDs and LCDs are given in Table 15.5 for easy understanding.

Table 15.3 *Comparison of LED and LCD*

Light emitting diode	*Liquid crystal display*
(i) An electronic device that lights up when an electricity is passed through it	(i) Digital display that uses liquid crystal cells that change reflectivity in a applied electric field.
(ii) LED or OLED displays does not have any light loss	(ii) Backlit by a light source, hence, there is a little light loss passing the light through the LCD panel
(iii) Requires 10–250 mW power per digit	(iii) Requires 10–200 μW power per digit
(iv) Generates visible light	(iv) Reflects a part of the surrounding light
(v) Gives high intensity light	(v) Gives high intensity light
(vi) Operate over wide range of temperatures (303 – 358 K)	(vi) Limited range of temperatures (293–333 K)
(vii) Higher lifetime	(vii) Limited lifetime due to chemical degradation
(viii) Emits different colours	(viii) Extra illumination required
(ix) Operating voltage ranges from 1.6 to 5 V dc	(ix) Operating voltage ranges from 3 to 20 V dc
(x) Response time is 50–500 ns	(x) Response time is 50–500 ns

15.12 APPLICATIONS

The following are the applications of display devices, namely, CRT, LED and LCD in various fields.

CRT

- CRT is an important versatile display device used for the analysis of electrical signals and optical display.
- CRTs are used in radar systems, televisions and computer monitors.

LED

- Traffic lights and signals
- Exit signs
- Bicycle lights
- Railroad crossing signals
- Flashlights
- Light bars on emergency vehicles.
- Thin and lightweight message displays at airports, railway stations, buses, trains, ferries etc.
- Remote controls, of TVs, VCRs, DVDs
- In optical fiber and Free Space Optical (FSO) communications

LCD

In recent years, the applications of LCDs have been endless. Most of the watches and pocket calculators use LCD. A simple example of a liquid crystal display in a digital watch is shown in Fig. 15.29. The display is created in a sandwich liquid crystal which is placed in between two glass sheets G2 and G3. G2 and G3 glass sheets are not polarised, however the one in front of G2 has grooves etched into it, and electrodes inside those grooves are used to determine charge required for the liquid crystal. The glass sheet G4 is coated with a conducting layer and acts as an electrode. Glass-film sheets G1 and G5 are with perpendicular polarisation and are placed on either side of these two glass sheets. As a result, the direction of the grooves line up with the direction of their nearest polarising films. The reflective surface glass plate G6 is placed behind the system. This will helps to view the light excited by liquid crystal pixel of LCD.

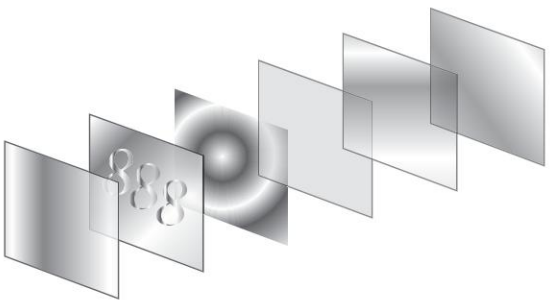


Fig. 15.29 *Digital water — LCD display*

- Liquid crystal display devices are now preferred to cathode ray tube (CRT) devices as they can offer superior performance in terms of maximum luminance and intensity.
- It can also offer some features like automatic calibration and remote quality assurance.

Keypoints to Remember

- Solids are classified based on energy band structure, arrangements of electrons and forbidden bands as conductors, insulators and semiconductors.
- In conductors, the valence band is completely filled while the conduction band is half filled.
- Solid materials which do not conduct electric current under normal conditions are known as insulators.
- In insulators, the valence band is completely filled and these are no electrons in conduction band.
- Three different types of semiconductors, namely, elemental, compound and oxide semiconductors.

- Semiconductors are classified into two types, namely, pure or intrinsic and doped or extrinsic semiconductors.
- In intrinsic semiconductors, the carrier concentration of electrons and holes are equal. These semiconductors are highly pure semiconductors.
- Doping is the process of adding impurity to an intrinsic semiconductor.
- Doped semiconductors are called extrinsic semiconductors.
- Extrinsic semiconductors are of two types, namely, *n*-type and *p*-type.
- In *n*-type semiconductors, electrons are majority carrier while in *p*-type semiconductors, the holes are majority carriers.
- When a dopant with valency higher than the parent atom is doped then the resultant semiconductor is known as *n*-type semiconductors.
- When a dopant with valency lower than the parent atom is doped, then the resultant semiconductor is known as *p*-type semiconductors.
- A solar cell is nothing but a *p-n* junction diode based on the principle of photoelectric effect and it converts light into electrical energy.
- Characteristics of solar cells are analysised from the *V-I* curve.
- Efficiency of a solar cell is defined as the ratio of the total power converted by the solar cell to the total power available for energy conversion.
- The requirements of a solar cell are that band gap energy of semiconducting materials is comparable with the energy of the photon present in the solar spectrum.
- Different classification of solar cells is *p-n* homojuction, *p-n* heterojunction, Schottky barrier, homojuction heterostrcuture, metal-insulator semiconductor cell and semiconductor insulator semiconductor cell.
- Active display device is the one which emits radiation on its own.
- Passive display device is the device which modulates incident radiation to provide necessary display information.
- The energy of the emitted photon is $E = hv$.
- LED is a forward-biased *p-n* junction.
- In LED, the emitted light colour depends on the materials used.
- Alphanumeric characters are displayed employing array LED's and seven segment LEDs.
- In LEDs, the energy loss during radiation is reduced employing plastic encapsulation.
- The colour of light emitted from LED depends on the semiconductor materials used.
- Large band gap ($\approx$ 2 eV) materials are used for LED applications.
- Lifespan of an LED is more than 1,00,000 hours.
- Two different types of principles, namely, polarisation and twisted nematic property of liquid crystal are employed for LCD.
- The process of filtering certain unwanted wavelengths employing glass with parallel thin impenetrable black lines is known as polarisation.

The McGraw·Hill Companies

- In twisted nematic display, the molecules in the liquid crystal are twisted and hence, the incident light entering the liquid crystal comes out as polarised light.
- Liquid crystals are classified as nematic, cholesteric and smectic liquid crystals.
- LCD has four different types of modes of display, namely, dynamic scattering, twisted nemactic scattering, phase change and memory effects.
- Dynamic scattering and twisted nematic scattering methods are mostly used for LCD devices.
- Letter on a dark background are produced in dynamic scattering methods.
- Black letters on a silver background are produced in field effect type.
- In twisted nematic LCD, the T_{on} is directly propositional to viscosity of the liquid medium.
- LED generates visible light while LCD reflects a part of the surrounding light.

Solved Problems

Example 15.1 Calculate the wavelength of light emitted by an LED with band gap of energy 1.8 eV.

Given Data

Bandgap of given LED is $E_g = 1.8$ eV

$= 1.8 \times 1.609 \times 10^{-19}$

$= 2.8962 \times 10^{-19}$ J

Solution

We know that wavelength emitted from given LED

$$\lambda = \frac{hc}{E_g}$$

Substituting the values, we get,

$$= \frac{6.626\times10^{-34}\times3\times10^{8}}{2.8962\times10^{-19}}$$

$= 6.863 \times 10^{-7}$ m

The wavelength of light emitted from given LED is 0.6863 μm.

Example 15.2 Calculate the band-gap energy of GaAsP semiconductor whose output wavelength is 6715A˚.

Given Data

The wavelength of green light from mercury lamp = 6715×10^{-10} m

Solution

We know that band gap is $E_g = \frac{hc}{\lambda}$

Substituting the values, we get,

$$= \frac{6.626\times10^{-34}\times3\times10^{8}}{6751\times10^{-10}} = 2.944 \times 10^{-19} \text{ J}$$

$$= 2.944 \times 10^{-19} \text{ J}$$

or $$= \frac{2.944\times10^{-19}}{1.6\times10^{-19}} = 1.8 \text{ eV}$$

The band gap of the given GaAsP is 1.8 eV.

Objectives Questions

15.1 In conduction, the valence band is _____ while the conduction band is _____ filled.
15.2 In conducting materials, the flow of current is due to shifting of _____ from valence band to conduction band.
15.3 In insulators, _____ band is completely filled and does not have any_____ is conduction band.
15.4 Examples for insulators are_____.
15.5 The forbidden gap energy of an insulator is _____ eV.
15.6 Examples for semiconductors_____, _____.
15.7 The conductivity of semiconductor is of the order of _____ $\text{ohm}^{-1}\ \text{m}^{-1}$.
15.8 Give an example for compound semiconductors _____, _____.
15.9 Give two examples for oxide semiconductors_____, _____.
15.10 In intrinsic semiconductors, the concentration of electrons and holes are not equal. True / False.
15.11 Give any two examples for intrinsic semiconductors_____, _____.
15.12 The valence electron for silicon is _____.
15.13 The conductivity of semiconductors depends on energy gap. True / False.
15.14 The majority current carrier is *n*-type semiconductor is _____.
15.15 _____ is the majority current carrier is *p*-type semiconductors.
15.16 Give any two examples for pentavallent atoms _____, _____.
15.17 _____ and _____ are the example for trivalent atom.
15.18 The device which converts the light energy into electrical energy is known as _____.
15.19 _____ is used for the antireflection coating at the surface of solar cell.
15.20 Efficiency of a solar cell is equal to
a) (Maximum output electrical power / input optical power) × 100
b) (Maximum input optical power/ output electrical power) × 100
c) (Maximum output electrical power × input optical power) × 100
d) (Maximum output electrical power / input optical power) / 100

The McGraw·Hill Companies

15.21 Fill factor of a solar cell is equal to
a) Maximum output power x (V_{oc} I_{SC})
b) Maximum output power x (V_{oc} / I_{SC})
c) Maximum output power / (V_{oc} I_{SC})
d) (V_{oc} I_{SC}) / Maximum output power

15.22 _____ device display the information on the screen.

15.23 The device which emits radiation is known as _____.
a) Passive device
b) Active and passive device
c) Active device
d) None of the above

15.24 The device which modulates the incident light is known as_____.

15.25 _____ is the principle behind LED.

15.26 LED is a _____ biased *p-n* junction.

15.27 Energy of the emitted photon is _____.

15.28 The colour of the emitted LED light depends on _____.

15.29 _____ and _____ LED's are used to display alphanumeric character.

15.30 The loss of emitted light energy in LED is due to _____.

15.31 _____ band gap materials are used to emit visible light.

15.32 Lower band gap materials are used to emit _____ and _____ light.

15.33 The wavelength of light emitted by GaAs LED lies between _____ and _____ nm.

15.34 GaP is a _____ band-gap semiconductor with energy _____.

15.35 The band-gap energy for $GaAs_{1-x}P_x$ at X = 0.45 is _____ eV.

15.36 The non-emissive display device is known as _____ display device.

15.37 _____ device reflects lights.

15.38 MBBA shows liquid crystal behaviour in the temperature range from _____ to _____ K.

15.39 A mixture of 20% cholesterol carbonate and 80% cholesterol non-anoate scatter _____ and _____ colours.

15.40 Dynamic light scattering is possible at _____ frequency.

15.41 The twisted molecules in a liquid crystal rotate the plane polarized light through an angle _____ .

15.42 In dynamic scattering, letters on _____ background is produced.

15.43 In field effect scattering type, black letters on _____ background are produced.

15.44 Response time for an LED is from _____ to _____ ns.

KEY

15.1 completely, half
15.2 electrons
15.3 valence, electron
15.4 carbonite
15.5 5
15.6 silicon, germanium
15.7 10^4 to 10^{-4}
15.8 GaAs, GaP
15.9 Bi_2O_3, Te_2O_3
15.10 False
15.11 silicon, germanium
15.12 4
15.13 True
15.14 electrons
15.15 holes
15.16 arsenic, antimony

15.17 indium, boron
15.18 photovoltaic cell
15.19 SiO_2
15.20 a)
15.21 c)
15.22 Display
15.23 c)
15.24 Passive devices
15.25 Electroluminescence
15.26 Forward
15.27 hν
15.28 Materials
15.29 Array and seven segments
15.30 Reflection loss
15.31 Larger
15.32 Infrared and far infra red
15. 33 910 and 1020
15.34 Indirect and 2.26 eV
15.35 2.1
15.36 Passive
15.37 LCD
15.38 293 and 320
15.39 Bright yellow and deep violet
15.40 Low
15.41 90°
15.42 Dark
15.43 Silver
15.44 50 and 500

Short Answer Questions

15.1 What is a semiconductor?
15.2 Define a conductor.
15.3 What is meant by an insulator?
15.4 Draw the energy level diagram of a conductor.
15.5 Draw the energy level diagram of an insulator.
15.6 Draw the energy level diagram of a semiconductor.
15.7 State the properties of semiconductors.
15.8 What is meant by an intrinsic semiconductor?
15.9 What is an extrinsic semiconductor?
15.10 Distinguish between intrinsic and extrinsic semiconductor.
15.11 What are elemental, compound semiconductors? Give two examples.
15.12 Explain the crystal structure of germanium and silicon.
15.13 Mention any four advantages of semiconductors.
15.14 What is meant by doping?
15.15 Define *p*-type and *n*-type semiconductor.
15.16 Explain *n*-type semiconductor.
15.17 What are the advantages of semiconducting materials?
15.18 What is solar cell?
15.19 Explain photovoltaic cell.
15.20 Explain the working of solar cell.
15.21 Explain the *V-I* characteristics of a solar cell.
15.21 What is meant by an open circuit?
15.23 Define the efficiency of a solar cell.
15.24 What is fill factor?
15.25 Mention the different classifications of a solar cell.

The McGraw·Hill Companies

15.26 What are the materials used for the fabrication of a solar cell.
15.27 Explain GaAs solar cell.
15.28 Mention the industrial applications of a solar cell.
15.29 What are the basic requirements of a solar cell?
15.30 What is meant by display devices?
15.31 Explain the classification of display devices.
15.32 Why do we need a suitable material for the LED?
15.33 Explain how LED emits light radiations.
15.34 What is meant by electromagnetic radiation?
15.35 Why the shape of the LED is made hemispherical?
15.36 Explain plastic encapsulation.
15.37 Explain the different types of LED materials along with their radiant colour.
15.38 What is meant by passive display devices?
15.39 Mention some of the properties of passive display devices.
15.40 What is meant by liquid crystal display?
15.41 Mention few applications of LCD.
15.42 What is meant by liquid crystal state?
15.43 Explain how LCD illuminates light.
15.44 What is the range of wavelength emitted by LCD?
15.45 What is meant by a director?
15.46 Explain why extra light is required to illuminate LCD.
15.47 Mention the different types of LCD display.
15.48 Mention few applications of LCD's
15.49 What are the drawbacks of LCD?
15.50 List some of the applications of display devices in solid state electronics.
15.51 What is the difference between active and passive display devices?
15.52 Explain the active display devices.

Descriptive Questions

15.1 Explain with a neat sketch the intrinsic and extrinsic semiconducting materials.
15.2 What is an energy band diagram? Explain the energy diagram for and extrinsic semiconducting materials.
15.3 Write notes on the following:
a) Intrinsic semiconductors
b) Extrinsic semiconductors
15.4 Explain with a neat sketch the construction, working and *V-I* characteristics of solar cell.
15.5 Explain the working of GaAs solar cell with neat sketch. Mention the applications of solar cells.
15.6 Explain how a *p-n* junction diode acts as a lighting emitting diode.

15.7 Explain with a neat sketch the basic principle, working and the applications of LED.
15.8 Explain with a neat sketch for the working of LCD along with their salient features.
15.9 Compare the relative merits of LCD and LED.
15.10 Write notes on the following:
a) LED
b) LCD

The McGraw·Hill Companies

Chapter

16

Electron Theory of Solids

OBJECTIVES

- To understand the electrical conduction in materials
- To explain the conductivity of a material using free electron theory
- To derive the equation for electrical conductivity of a solid using free electron theory
- To derive the equation for thermal conductivity of a material and hence, derive the Wiedemann–Franz law
- To verify Ohm's law
- To discuss the limitations of classical free electron theory

16.1 INTRODUCTION

In solids, electrons in the outermost orbit of atoms determine its electrical properties. Electron theory is applicable to all solids, both metals and non-metals. In addition, it explains the electrical, thermal and magnetic properties of solids. The structure and properties of solids are explained employing their electronic structure by the electron theory of solids.

There are three main stages for development of this theory, namely, classical free electron theory, quantum free electron theory and zone theory. The first theory was developed by Drude and Lorentz in 1900. According to this theory, metal contains free electrons which are responsible for the electrical conductivity and metals obey the laws of classical mechanics. In 1928, Sommerfield developed the quantum free electron theory. According to Sommerfield, the free electrons move with a constant potential. This theory obeys the quantum laws. Bloch developed the Brillouin zone or Band theory in 1928. According to this theory, free electrons move in periodic potential provided by the lattice. This theory is known as band theory of solids.

Let us discuss in the following headings, electrical conduction and thermal conduction, their relationship and the advantages and drawbacks of classical free electron theory.

The McGraw·Hill Companies

16.2 ELECTRICAL CONDUCTION

When an electronic potential difference is applied across materials, it conducts electricity. Electrical conductivity is defined as the quantity of electricity flowing through a material per unit area per unit time maintained at unit potential gradient. The material with this property is called a *conducting material*. The most important property of a material is electrical resistance, and it is used to characterise the electrical properties of the material. Let R be the resistance of the material. A and l are respectively the area and length of the material.

Therefore, the resistivity of the material is

$$\text{Resistivity } (\rho) = \text{Resistance} \times \frac{\text{area}}{\text{length}}$$

or
$$\rho = R \frac{A}{l} \qquad \Omega\,\text{m} \tag{16.1}$$

The reciprocal of the electrical resistivity is known as *electrical conductivity* (σ),.

i.e.,
$$\sigma = \frac{1}{\rho}$$

$$= \frac{1}{R}\frac{l}{A} = \frac{l}{RA} r^{-1} m^{-1} \tag{16.2}$$

The electrical conductivity of a material depends only on the presence of free electrons or conduction electrons. These free electrons move freely in the metal and do not correspond to any atom. This group of electrons is also called *electron gas*.

16.3 CLASSIFICATION OF CONDUCTING MATERIALS

Based on the electrical conductivity, conducting materials are classified into three major categories, namely,

(i) Zero resistivity materials
(ii) Low resistivity materials, and
(iii) High resistivity materials

(i) Zero Resistivity Materials The material which conducts electricity at zero resistance is known as a zero resistivity material. Superconducting materials like alloys of aluminium, zinc, gallium, nichrome, nibium, etc., conduct electricity almost at zero resistance below the transition temperature. The above superconducting materials are perfect diamagnets. The applications of these materials are energy saving in power systems, superconducting magnets, memory storage, etc.

The McGraw·Hill Companies

(ii) Low Resistivity Materials The resistivity of high electrical conductivity materials is very low. The electrical conductivity of metals and alloys like silver, and aluminium is very high, of the order of $10^8\ \Omega^{-1}\ m^{-1}$. Low resistivity materials are used as resistors, conductors, electrical contacts, etc., in electrical devices. In addition, these materials are also used in electrical power transmission and distribution lines, winding wires in motors and transformers.

(iii) High Resistivity Materials Generally, materials with high resistivity will have low temperature coefficient of resistivity. For example, tungsten, platinum, nichrome, etc., have high resistivity and low temperature coefficient of resistance. High resistivity materials are used in the manufacturing of resistors, heating elements, resistance thermometers, etc.

16.4 CLASSICAL FREE ELECTRON OR DRUDE–LORENTZ THEORY OF METALS

Very high electrical and thermal conductivity are the outstanding properties of metals. The positive ion cores and the valence electrons are the charge carriers in metals and hence, good conductivity means the presence of a mobile charge carrier. The observed high conductivity with conduction by the drift ion cores is difficult to understand. If this happens, then conductivity must increase with increase in temperature. Hence, it may be concluded that mobile charges in metals are valence electrons. The classical free electron theory reveals that the free electrons are fully responsible for electrical conduction.

16.4.1 Postulates of Classical Free Electron Theory

The free electrons, or electron gas, available in a metal move freely here and there during the absence of an electronic field similar to the gas molecules moving in a vessel as shown in Fig. 16.1. These free electrons collide with other free electrons or positive ion cores or the walls of the container. Collisions of this type are known as *elastic collisions*. The total energy of an electron is assumed to be purely KE.

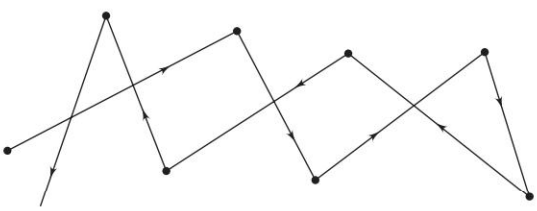


Fig 16.1 *Absence of the field—Free electron*

On the other hand, suppose an electric field is applied to the material through an external arrangement as shown in Fig 16.2. The free electrons available in the metal

gain some amount of energy and are directed to move towards a higher potential. These electrons acquire a constant velocity known as *drift velocity* <*v*>. The velocity of the electron obeys the Maxwell–Boltzmann studies.

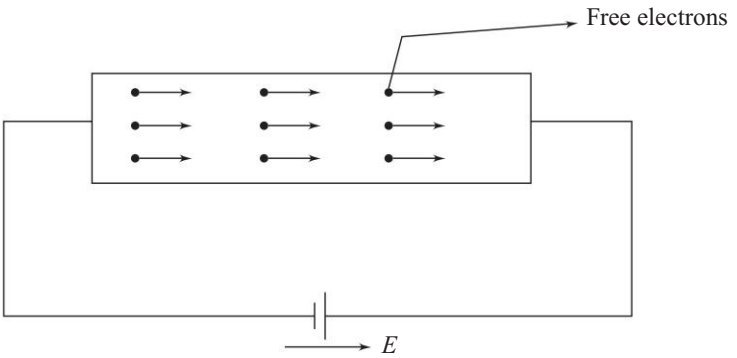


Fig. 16.2 *Presence of field—free electron*

Drift Velocity It is defined as the average velocity acquired by the free electron in a particular direction during the presence of an electric field.

Relaxation Time The relaxation time is defined as the time taken by a free electron to reach its equilibrium position from its disturbed position, during the presence of an applied field.

$$\tau = \frac{l}{< v >} \tag{16.3}$$

where l is the distance traveled by the electron.

16.5 EXPRESSION FOR ELECTRICAL CONDUCTIVITY AND DRIFT VELOCITY

In order to obtain the expression for electrical conductivity, the velocity of the electron in the conducting material is obtained by two methods. In the first method, the velocity of the electron is obtained without considering the collision of free electrons. In the second method, the collision of free electrons is taken into account for obtaining the velocity. Thus, the current density behaviour of a conducting material is studied under the above two approaches to obtain the expression for electrical conductivity.

Consider a conductor which is subjected to an electric field of strength E as shown in Fig. 16.3. Consider that n is the concentration of the free electrons; m, the mass and e, the electric charge of the electron in the conductor. According to Newton's second law of motion, the force F acquired by the electrons is equal to the force exerted by the field on the electrons.

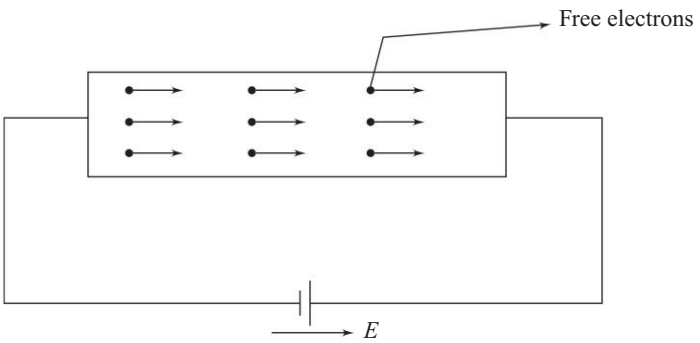


Fig.16.3 *Metal—conduction*

Therefore, the equation of motion $ma = -eE$ (16.4)

or Acceleration $a = \dfrac{-eE}{m}$

Integrating Eq. (16.4), we get,

The velocity of the electron $v = -\dfrac{eE}{m}t + C$ (16.5)

where C is an integration constant, and is obtained by applying the boundary conditions. During the absence of the electric field, the average velocity of the electron is zero,

i.e., when $t = 0, < v > = 0$

Substituting the above boundary condition in Eq. (16.5), we get, $C = 0$
Substituting the value of C in Eq. (16.5), we get,

Velocity of the electron $= -\dfrac{eE}{m}t$ (16.6)

According to Ohm's law, the current density in a conductor,

$$J = \frac{I}{A} \tag{16.7}$$

Therefore, the charge dq flowing through this area A in time dt is

$$dq = -e\,n\,A\,v \quad dt$$

$$\frac{dq}{dt} = I = -enAv \tag{16.8}$$

where I is the current flowing through the conductor with an area of cross-section A. Substituting the value of I from Eq. (16.8) in Eq. (16.7), we get,

$$J = -nev \tag{16.9}$$

Substituting the value of ν from Eq. (16.6) in Eq. (16.9), we get,

$$J = \frac{ne^2 E}{m} t \tag{16.10}$$

It is clear from Eq. (16.10) that the current density is directly proportional to the applied field E. When the field increases, the current density also increases, and it reaches infinity when the field is applied for a long period. However, in actual practice, J never becomes infinity. Rather, it remains constant beyond a certain field strength. This is due to the presence of collisions of free electrons which is not taken into account for obtaining the current density.

Due to the collision of free electrons, a frictional force acts on the electrons. The frictional force experienced by the electron

$$= m < v > / \tau_r \tag{16.11}$$

where <v> is the drift velocity of the electron and τ_r the relaxation time between any two successive collisions during the presence of an electron field.
Therefore, equation of motion,

$$m\frac{d < v >}{dt} + \frac{m < v >}{\tau_r} = - eE \tag{16.12}$$

Rearranging Eq. (16.12), we get,

$$\tau_r \frac{d < v >}{dt} = -\left[\frac{eE\tau_r}{m} + < v >\right]$$

or
$$\frac{d < v >}{\left[\frac{eE\tau_r}{m} + < v >\right]} = \frac{-dt}{\tau_r} \tag{16.13}$$

Integrating Eq. (16.13), we get,

$$\text{In}\left[\frac{e\,E\,\tau_r}{m} + < v >\right] = \frac{-t}{\tau_r} + \text{In}\, B \tag{16.14}$$

where In B is an integration constant.

Simplifying, Eq. (16.14), we get,

$$\frac{e\,E\,\tau_r}{m} + < v > = Be^{-t/\tau r} \tag{16.15}$$

In order to obtain the value of the integration constant, consider the boundary condition, when $t = 0$, $< \text{v} > = 0$,
Therefore Eq. (16.15) can be written as,

$$B = \frac{e\,E}{m}\,\tau_r \tag{16.16}$$

or $$< v > = \frac{e E \tau_r}{m} [e^{-t/\tau r} - 1] \quad (16.17)$$

Substituting the value of $< v >$ from Eq. (16.17) in Eq. (16.9), we get,

The current density $$J = \frac{ne^2 \tau_r E}{m} [1 - e^{-t/\tau r}] \quad (16.18)$$

$$J = \sigma E$$

The current density is proportional to the electric field intensity E and a constant σ, known as electrical conductivity.

Therefore, electrical conductivity $$\sigma = \frac{ne^2 \tau_r}{m} [1 - e^{-t/\tau r}] \quad (16.19)$$

Equation (16.19) shows that the conductivity of the material is directly proportional to E and it reaches a maximum value within 10^{-14} s (since τ_r is of the order of 10^{-14} s). Beyond the limit, it remains as constant.

Therefore, the steady state-value of the electrical conductivity and drift velocity of the electron are

Electrical conductivity $$\sigma = \frac{ne^2 \tau_r}{m} \quad (16.20)$$

and drift velocity $$< v > = \frac{e E \tau_r}{m} \quad (16.21)$$

Resistivity (ρ) We know that the resistivity of the material is the reciprocal of electrical conductivity.
Therefore, the resistivity

$$\rho = \frac{1}{\sigma}$$

$$= \frac{m}{ne^2 \tau_r} \quad (16.22)$$

Mobility (μ_e) We know that from Eq. (16.21), the average velocity of electrons

$$< v > = \frac{e \tau_r}{m} \text{ E} \quad (16.23)$$

or $$< v > \propto \text{E}$$

where $\frac{e \tau_r}{m}$ is the proportionality factor known as *mobility* μ_e .

i.e., $$\mu_e = \frac{e\tau_r}{m} \quad (16.24)$$

The mobility of a charge carrier is defined as the average velocity of the charge carrier per unit applied electric field intensity.

Mean Free Path (λ) The average distance travelled by an electron between any two successive collisions is known as the mean free path.

i.e, $$\lambda = v\tau_r \quad (16.25)$$

where v is the root mean square velocity of the electron.

16.6 THERMAL CONDUCTIVITY

Thermal conductivity (K) of *a* material is equal to the amount of heat energy (Q) conducted per unit area of cross-section per second to the temperature gradient (dT/dx).

i.e. $$Q \propto dT/dx$$

or $$Q = K\frac{dT}{dx}$$

Therefore, thermal conductivity $$K = -\frac{Q}{(dT/dx)} \quad (16.26)$$

In solids, the conduction takes place both by available free electrons and thermally excited lattice vibrations known as *phonons*.
Therefore, the total conductivity is

$$K_{\text{total}} = K_{\text{electrons}} + K_{\text{phonons}} \quad (16.27)$$

where phonons are the energy carriers for lattice vibrations.

Metals In case of metals, the concentration of free electrons is very high. Therefore, electrons, in addition to thermal vibrations, absorb and transport thermal energy. Hence, the thermal conductivity of a metal is due to the free electrons. ($K_{\text{electron}} >> K_{\text{phonon.}}$),

$$K_{\text{total}} = K_{\text{electrons}} \quad (16.28)$$

Insulators In insulators, the thermal conductivity is due to atomic or molecular vibration of the lattice,

i.e, $$K_{\text{total}} = K_{\text{phonons}} \quad (16.29)$$

Semiconductors The thermal conductivity is due to both electrons and phonons.

Therefore, $$K_{\text{total}} = K_{\text{electrons}} + K_{\text{phonons}} \quad (16.30)$$

The McGraw·Hill Companies

16.7 EXPRESSION FOR THERMAL CONDUCTIVITY

Consider a uniform rod *HC* with the temperature of the hot end *H* as T_1 and the temperature of the cold end *C* as T_2, as shown in Fig.16.4. Heat is flowing from the hot end *H* to the cold end *C*. Let *A* be the cross-sectional area which is at a distance equal to the mean free path λ of the electrons between the two ends *H* and *C*. The kinetic energy of the electrons at the hot end *H* is greater than that of the electrons at the colder end *C*.

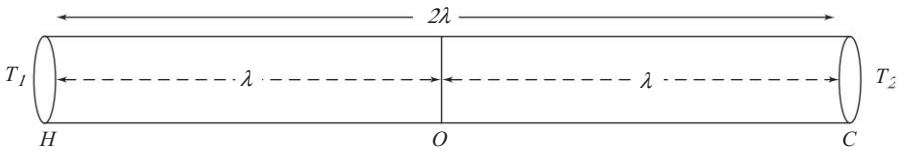


Fig. 16.4 *Thermal conductivity*

Let *Q* be the amount of heat flowing through the rod from the end *H* to *C* whose length is 2λ. Therefore, the heat conducted

$$Q \propto \frac{A(T_1 - T_2)}{2\lambda} t$$

or
$$Q = \frac{KA(T_1 - T_2)t}{2\lambda} \tag{16.31}$$

where *K* is the coefficient of thermal conductivity; *A*, the area of cross-section; *t*, the time for conduction and 2λ, the length of the rod.

Therefore, the thermal conductivity per unit area per unit time

$$K = \frac{Q\,2\lambda}{A(T_1 - T_2)t} \tag{16.32}$$

Let *n* be the number of available conduction electrons and *v* be the root mean square velocity of the electrons. Let us assume that the available free electrons in the metal have an equal probability to move in all the six possible directions. Therefore, an average of 1/6 nv electrons can travel in any one direction. We know that the free electrons are assumed to be free moving gas molecules.

The average kinetic energy of an electron at the hot end *H* of temperature T_1

$$= \frac{3}{2} k_B T_1 \tag{16.33}$$

where k_B is the Boltzman constant.

Similarly, the average kinetic energy of an electron at the cold end c of the temperature T_2

$$= \frac{3}{2} k_B T_2 \tag{16.34}$$

The number of electrons crossing the area at A per second,

$$= \frac{1}{6} n v \tag{16.35}$$

$\therefore$ the heat energy transferred per unit area per unit time from the hot end H to the cold end C is = number of electrons × average KE of electrons moving from H to C

$$= \frac{1}{6} n v \frac{3}{2} k_B T_1$$

$$= \frac{1}{4} n v k_B T_1 \tag{16.36}$$

Similarly, the heat energy transferred across A per second from C to H

$$= \frac{1}{6} n v \frac{3}{2} k_B T_2$$

$$= \frac{1}{4} n v k_B T_2 \tag{16.37}$$

$\therefore$ the resultant heat energy transferred from H to C across A per second,

$$Q = \frac{nv}{4} k_B (T_1 - T_2) \tag{16.38}$$

We know that, temperature gradient $\frac{dT}{dx} = \frac{T_1 - T_2}{2\lambda}$ (16.39)

Substituting the values of Q and temperature gradient from Eq.(16.38) and (16.39), in Eq. (16.32) we get,

or Thermal condition $K = \frac{n(v/6)(3/2)k_B (T_1 - T_2)}{(T_1 - T_2)/2\lambda}$

$$K = \frac{1}{4} nv k_B \frac{(T_1 - T_2)\, 2\lambda}{(T_1 - T_2)} \tag{16.40}$$

Simplifying the above equation, we get,

$$K = \frac{1}{2} nv k_B \lambda \tag{16.41}$$

We know from the classical free electron theory, the electronic specific heat capacity of the metal is

$$C_e = \frac{3}{2} R \tag{16.42}$$

where R is the gas constant and is equal to $n k_B$.

Substituting the R value in Eq. (16.41), we get,

Thermal conductivity $K = \frac{1}{3} C_e \, v \, \lambda$ (16.43)

Thermal conductivity of a metal is determined either using Eq. (16.41) or Eq (16.43) depending on the available data.

16.8 WIEDEMANN–FRANZ LAW

The ratio between the thermal conductivity and electrical conductivity of a metal is directly proportional to the absolute temperature of the metal when the temperature is not too low,

i.e., $$\frac{K}{\sigma} \propto T$$

$$\frac{K}{\sigma T} = L \tag{16.44}$$

where $K/\sigma\, T$ is a constant known as *Lorentz number*. The value of L is equal to 2.44×10^{-8} W Ω K^{-2} at 293 K.
We know that

thermal conduction $K = \frac{1}{2} nv k_B \lambda$ (16.45)

electrical conduction $\sigma = \frac{ne^2 \tau r}{m}$ (16.46)

$\therefore$ $$\frac{K}{\sigma} = \frac{1/2 nv k_B \lambda}{(ne^2 \tau_r / m)} \tag{16.47}$$

$$= \frac{mv k_B \lambda}{2e^2 \tau_r} \tag{16.48}$$

Substituting $\lambda = v \tau_r$ in the above equation, we get,

$$\frac{K}{\sigma} = \frac{mv^2 k_B \tau_r}{2e^2 \tau_r} \tag{16.49}$$

Simplifying the above equation, we get,

$$\frac{K}{\sigma} = \frac{3}{2} k_B T \frac{k_B}{e^2} \tag{16.50}$$

We know that kinetic energy of an election is $1/2\, mv^2 = 3/2\, k_B\, T$,

Therefore, $$\frac{K}{\sigma T} = \frac{3}{2} \frac{k_B^2}{e^2}$$

Lorentz number $L = \frac{3}{2} \times \frac{k_B^2}{e^2}$ (16.51)

Substituting the values of k_B and e in Eq. (16.51), we get,

Lorentz number $$L = \frac{3(1.38 \times 10^{-23})^2}{2(1.6 \times 10^{-19})^2}$$

$$= 1.11 \times 10^{-8} \text{ W } \Omega \text{ K}^{-2} \quad (16.52)$$

We observe that the theoretical value of the Lorentz number is only one half of the experimental value ($= 2.44 \times 10^{-8}$ W Ω K^{-2}). The observed discrepancy is due to the failure of the classical theory to give the correct value for the thermal conductivity of metals. In order to rectify the discrepancy, the concept of quantum theory has been used to obtain the expression for thermal conductivity. According to the quantum theory, the expression for thermal conductivity is modified by considering the electronic specific heat as,

$$K = \frac{n\pi^2 k_B^2 T \tau_r}{3m} \quad (16.53)$$

We know that Lorentz number, $L = \frac{K}{\sigma T}$

Substitutly the values of K and σ, we get,

or $$L = \frac{n\pi^2 k_B^2 T \tau_r}{3m} \times \frac{m}{ne^2 \tau_r T} \quad (16.54)$$

Lorentz number $$= \frac{\pi^2}{3}\left[\frac{k_B}{e}\right]^2$$

Substituting the values of k_B, e and π in Eq. (16.54), we get,

$$L = \frac{\pi^2 (1.38 \times 10^{-23})^2}{3(1.6 \times 10^{-19})^2}$$

Lorentz number $$= 2.44 \times 10^{-8} \text{ W } \Omega \text{ K}^{-2} \quad (16.55)$$

Equation (16.55) gives the correct value of the Lorentz number and is in good agreement with the experimental value. Thus, it confirms that the Wiedemann–Franz law is verified using the quantum theory. Further, it supports that Wiedemann–Franz law is not applicable for low temperature.

16.9 VERIFICATION OF OHM'S LAW

The classical free electron theory is used to verify Ohm's law. In order to verify the same, consider that the steady-state current density can be written from Eq. (16.19) as

$$J = \frac{ne^2 \tau_r E}{m} \quad (16.56)$$

The McGraw·Hill Companies

Similarly, the steady-state conductivity from Eq. (16.20), is

$$\sigma = \frac{ne^2 \tau_r}{m} \tag{16.57}$$

Compare Eqs. (16.56) and (16.57), one can write,

$$J = \sigma E \tag{16.58}$$

We know that the current density $J = I/A$, the conductivity $\sigma = 1/\rho$, and the electric field intensity $E = V/d$, Therefore, Eq.(16.58) is given by

$$\frac{I}{A} = \frac{1}{\rho} \frac{V}{d}$$

or
$$V = \left(\frac{\rho d}{A}\right) I \tag{16.59}$$

We know that, $\rho = \frac{RA}{d}$ and hence,

$$R = \frac{\rho d}{A} \tag{16.60}$$

Substituting the above value in Eq. (16.59), we get,

$$V = I R \tag{16.61}$$

Equation (16.61) states the Ohm's law. It is clear from the above equation that Ohm's law is verified employing classical free electron theory.

16.10 CLASSICAL FREE ELECTRON THEORY: ADVANTAGES AND RAWBACKS

Even though the classical free electron theory is the first theory developed to explain the electrical conduction and thermal conduction of metals, it has many practical applications. The advantages and disadvantages of the classical free electron theory are as follows:

Advantages

(i) It explains the electrical conductivity and thermal conductivity of metals.
(ii) It explains the Wiedemann–Franz law.
(iii) It verifies Ohm's law.
(iv) It is used to explain the optical properties of metals.

Drawbacks

(i) It fails to explain the electric specific heat and the specific heat capacity of metals.
(ii) It fails to explain superconducting properties of metals.

(iii) It fails to explain new phenomena like photoelectric effect, Compton effect, black-body radiation, etc.

(iv) It fails to explain electrical conductivity of semiconductors or insulators.

(v) The classical free electron model predicts the incorrect temperature dependence of σ. According to the classical free electron theory, $\sigma \propto T^{-1}$.

(vi) It fails to give a correct mathematical expression for thermal conductivity.

(vii) Ferromagnetism couldn't be explained by this theory.

(viii) Susceptibility has greater theoretical value than the experimental value.

Keypoints to Remember

- The three main theories developed for metals are the classical free electron theory, quantum free electron theory and zone or Band theory.
- The classical free electron theory is a macroscopic theory and it obeys classical laws.
- The quantum free electron theory is a microscopic theory and it obeys quantum laws.
- Zone theory is a macroscopic theory and is based on the energy bands of solids.
- Electrical conductivity is defined as the quantity of electricity flowing through the material per unit area per unit time maintained at unit potential gradient.
- The electrical conductivity $\sigma = \frac{ne^2\tau_r}{m}$ where n is the number of electrons per unit; e, the charge; m, the mass and τ_r, the relaxation time.
- The mobility of a charge carrier is defined as the average velocity of the charge per unit applied electric field intensity.

$$\mu_e = \frac{e\tau_r}{m}$$

- The mean free path is the average distance travelled by the electron between any two successive collisions $\lambda = v\tau_r$
- The thermal conductivity of a metal $K_{\text{total}} = K_{\text{electrons}} + K_{\text{phonons}}$
- Thermal conductivity of an insulator $K_{\text{total}} = K_{\text{phonons}}$
- The coefficient of thermal conductivity is defined as the ratio of the heat energy transferred across unit area of cross-section in one second to the temperature gradient.

$$K = \frac{Q}{(T_1 - T_2)} 2\lambda$$

- The thermal conductivity of a metal

$$K = \frac{1}{2} n v k_B \lambda$$

- The thermal conductivity of a metal in terms of electronic specific heat

$$K = \frac{1}{3} C_e v \lambda$$

- The Wiedemann–Franz law states that the ratio between thermal conductivity and electrical conductivity of a metal is directly proportional to absolute temperature.

$$\frac{K}{\sigma} \propto T$$

- Lorentz number $L = \dfrac{K}{\sigma T}$
- Ohm's law is verified using the classical free electron theory.

Solved Problems

Example 16.1 The following data are given for copper: i) Density = 8.92×10^3 kg m^{-3}, ii) Resistivity = 1.73×10^8 Ω m iii) Atomic weight = 63.5. Calculate the mobility and the average time collision of electrons in copper obeying classical laws.

Given Data

The resistivity of copper $\rho = 1.73 \times 10^8$ Ω m
The atomic weight of copper $z = 63.5$
The density of copper is $d = 8.92 \times 10^3$ kg m^{-3}

Solution

We know that the carrier concentration $= \dfrac{\text{Avogadro's} \times \text{density}}{\text{Atomic weight}}$

Substituting the values, we get

i.e, $n = 6.023 \times 10^{26} \times \dfrac{8.92 \times 10^3}{63.5}$

$n = 8.46 \text{ x } 10^{28}$ m^{-3}

The conductivity of copper is, $\sigma = \dfrac{1}{\rho}$

$= \dfrac{1}{1.73 \times 10^{-8}}$

$$= 5.78 \times 10^7 \ \Omega^{-1} \text{m}^{-1}$$

We know that conductivity is $\sigma = \dfrac{ne^2\tau}{m}$

Substituting the values, we get, $\tau = \dfrac{\sigma m}{ne^2}$

$$= \frac{5.78\times10^7 \times 9.11\times10^{-31}}{8.46\times10^{28}\times(1.6\times10^{-19})^2}$$

$$\tau = 2.43 \times 10^{-11} \text{ s.}$$

Mobility $\mu = \dfrac{\sigma}{ne} = \dfrac{5.78 \times 10^7}{8.46 \times 10^{28} \times 1.6 \times 10^{-19}}$

$$\mu = 4.27 \times 10^{-3} \text{m}^2 \text{ V}^{-1} \text{ s}^{-1}$$

The mobility of electrons in copper is $4.27 \times 10^{-3} \text{m}^2 \text{ V}^{-1} \text{ s}^{-1}$.

Example 16.2 Sodium metal with bcc structure has two atoms per unit cell. The radius of sodium atom is 1.85 $\overset{0}{\text{A}}$. Calculate its electrical resistivity at 0° C, if the classical value of mean free time at this temperature is 3×10^{-14} s.

Given Data

The radius of the sodium atom $= 1.85 \text{ Å}$
Mean free time of sodium at 0ºC $\tau_r = 3 \times 10^{-14}$ s

Solution

We know that resistivity $\rho = \dfrac{m}{n_2 e^2 \tau}$

For bcc structure, lattice constant is given by $a = \dfrac{4}{\sqrt{3}} r$

Therefore, $a = \dfrac{4\times1.85\times10^{-10}}{\sqrt{3}} = 4.27 \times 10^{-10}$ m

Density of electrons $n_e = \dfrac{\text{Number of atoms per unit cell}}{\text{Volume } (a^3)}$

where number of atoms per unit cell is 2 for bcc structure.

Therefore, $\therefore \quad n_e = \dfrac{2}{(4.27\times10^{-10})^3}$

$$n_e = 2.57 \times 10^{28} \text{m}^{-3}$$

$$\therefore \qquad \rho = \frac{9.11 \times 10^{-31}}{(2.57\times10^{28})\times(1.6\times10^{-19})^2 \times(3\times10^{-14})}$$

$$\rho = 4.616 \times 10^{-8}\ \Omega\text{m}$$

Therefore, the resistivity copper is $4.616 \times 10^{-8}\ \Omega\text{m}$.

Example 16.3 Calculate the electrical resistivity of sodium at 0°C. It has 25.33×10^{27} electrons per unit volume and has a mean free time of 3.1×10^{14}s.

Given Data

Number electrons per unit volume $n = 25.33 \times 10^{27}$
Mean free time of the electron $\tau_r = 3.1 \times 10^{14}$

Solution
We know that the electrical resistivity of sodium metal is

$$\rho = \frac{m}{ne^2 z}$$

where m and e are the mass and charge of electrons respectively.
Substituting the value of n, $\tau_{r,\,m}$ and e, we get,

$$= \frac{9.11 \times 10^{-31}}{25.33\times10^{27}\,(1.6\times10^{-19})^2 \times3.1\times10^{14}}$$

$$= 4.532 \times 10^{-36}\ \Omega\text{ m}$$

The electrical resistively of sodium at 0°C is $4.532 \times 10^{-36}\ \Omega$ m.

Example 16.4 The relaxation time of conduction electrons in a material is 3×10^{-14}s. If the density of electrons is 5.8×10^{28} per m^3, calculate the resistivity of the material and mobility of the electron.

Given Data

Number of force electrons per unit volume $n = 5.8 \times 10^{28}$ per m^3
The relaxation time of conduction electrons $\tau_r = 3 \times 10^{-14}$ s.

Solution

We know that the electrical resistivity of the metals is given by

$$\rho = \frac{m}{ne^2\,\tau_r}$$

Substituting the value of n, τ_r, mass of electron and charge of electron, we get,

$$\rho = \frac{9.11\times10^{-31}}{5.8\times10^{28}\times(1.6\times10^{-19})^2\times3.4\times10^{14}}$$

$$= 1.804 \times 10^{-36}\ \Omega\text{ m}$$

Therefore, the electrical conductivity of the material is $1.804 \times 10^{-36}\ \Omega^{-1}\ m^{-1}$.
We also know that,
the mobility of electron is

$$\mu_e = \frac{e\tau_r}{m}$$

Substituting the value of τ_r and m, we get,

$$= \frac{1.6\times10^{-19}\times3.4\times10^{14}}{9.11\times10^{-31}} = 5.9715$$

Therefore, the mobility of the electron in a metal is $5.9715\ m^2\ v^{-1}s^{-1}$.

Example 16.5 A uniform silver wire has a resistivity of 1.54×10^{-8} ohm/m at room temperature. For an electric field along the wire of 1 volt/cm, compute the average drift velocity of an electron assuming that there is 5.8×10^{28} conduction electrons/m^3. Also calculate the mobility.

Given Data

The resistivity of the silver = 1.54×10^{-8} ohm/m
Electric field along the wire = $100\ v\ m^{-1}$
The carrier concentration of electron = $5.8 \times 10^{28}\ m^{-3}$

Solution

We know that, the mobility of the electron $\mu = \frac{\sigma}{ne}$ (or) $\frac{1}{\rho\ ne}$

Substituting the values of n, e and ρ, we get,

$$\mu = \frac{1}{1.54\times10^{-8}\times5.8\times10^{28}\times1.6\times10^{-19}}$$

$$\mu = \frac{1}{142.912} = 6.9973\times10^{-3}$$

The mobility of the electrons in silver is $6.9973 \times 10^{-3}\ m^2\ v^{-1}s^{-1}$

The drift velocity $v_d = \mu\ E$

$$v_d = 6.9973\times10^{-3}\times100$$

Therefore, Drift velocity $(v_d) = 0.69973$
The drift velocity of the electron is silver is 0.69973 m s^{-1}.

Example 16.6 The density of silver is $10.5 \times 10^3\ kg\ m^{-3}$, assuming that each silver atom provides one conduction electron. The conductivity of silver at 20° C is $6.8 \times 10^7\ \Omega^{-1}\ m^{-1}$. Calculate the density and mobility of electron in silver with atomic weight of $107.9 \times 10^{-3}\ kg\ m^{-3}$.

The McGraw·Hill Companies

Given Data

The density of silver $= 10.5 \times 10^3$ kg m^{-3}
The conductivity of silver at 20^{0}C $= 6.8 \times 10^7\ \Omega^{-1}$ m^{-1}
The atomic weight of silver $= 107.9 \times 10^{-3}$ kg m^{-3}

Solution

We know that, the number of atoms present per m^3

$$= \text{Avogadro's number} \times \frac{\text{density}}{\text{atomic weight}}$$

$$= 6.023 \times 10^{26} \times \frac{10.5 \times 10^3}{107.9}$$

$$= 5.86 \times 10^{28} \text{ atoms/m}^3$$

As each silver atom contributes one electron, the density of electrons is 5.8×10^{28} atoms/m^3.

We know that the conductivity is given by $\sigma = ne\mu$

$$\mu = \frac{\sigma}{ne} = \frac{6.8 \times 10^7}{5.86 \times 10^{28} \times 1.609 \times 10^{-19}}$$

$$\therefore \qquad \mu = 0.72 \times 10^{-2} \text{ m}^2 \text{ V s}^{-1}$$

Therefore, the mobility of electron is 0.72×10^{-2} m^2 V s^{-1}.

Example 16.7 The thermal and electrical conductivities of Cu at 20^{o}C are 390 W m^{-1} K^{-1} and 5.87 x 10^7 Ω^{-1} m^{-1}, respectively. Calculate the Lorentz number.

Given Data

Electrical conductivity of copper $\sigma = 5.87 \times 10^7\ \Omega^{-1}$ m^{-1}
Thermal conductivity of copper $k = 390$ W m^{-1} k^{-1}
Temperature $T = 293$ K

Solution

We know that the Lorentz number is

$$L = \frac{K}{\sigma T}$$

Substituting the values of K, σ and T, we get,

$$= \frac{390}{5.87 \times 10^7 \times 293}$$

$$= 2.26756 \times 10^{-8}\ W\,\Omega\,K^{-2}$$

Therefore, the Lorentz number is 2.267×10^{-8} W Ω K^{-2}.

Example 16.8 Calculate the electrical and thermal conductivities for a metal with relaxation time of 10^{-14} second at 300 K. Also, calculate the Lorentz number using the above result. (Density of electrons = 6×10^{28} m^{-3}.)

Given Data

The relaxation time $\tau_r = 10^{-14}$ s
The temperature $T = 300$ K
The electron concentration $n = 6 \times 10^{28}$ m^{-3}

Solution

We know that the electrical conductivity is

$$\sigma = \frac{ne^2\tau_r}{m}$$

Substituting the values of n, e, τ_r and m, we get,

$$= \frac{6\times10^{28} \times (1.6\times10^{-19})^2 \times 10^{-14}}{9.1\times10^{-31}}$$

$$= 1.6879 \times 10^7$$

Therefore, the electrical conductivity is 1.6879 X 10^7 Ω^{-1} m^{-1} .
We know that the thermal conductivity

$$k = \frac{n\pi^2 k^2_B T\tau_r}{3m}$$

Substituting the values, we get,

$$= \frac{6\times10^{28} \times \pi^2 (1.38\times10^{-23})^2 \times 300 \times 10^{-14}}{3 \times 9.1\times10^{-31}}$$

$$= 123.927525 \text{ W m}^{-1} \text{ k}^{-1}$$

We know the thermal conductivity is 123.92 W m^{-1} k^{-1}.

Lorentz number is $L = \frac{k}{\sigma T}$

Substituting the corresponding value in the above equation, we get,

$$= \frac{123.927525}{1.6879\times10^7 \times 300} = 2.44737 \times 10^{-8}$$

The Lorentz number is 2.4474×10^{-8} W Ω K^{-2}.

Example 16.9 The density and atomic weight of Cu are 8900 kg m^{-3} and 63.5. The relaxation time of electrons in Cu at 300 K is 10-^{14} s. Calculate the electrical conductivity of copper.

Given Data

Density of copper = 8900 kg m^{-3}
Atomic weight of Cu = 63.5
Relaxation time = 10^{-14} s

The McGraw·Hill Companies

Solution

We know that the concentration of free electrons

$$n = \frac{\text{Avogadro's constant} \times \text{denstiy} \times \text{number of free electrons per atom}}{\text{atomic weight}}$$

Substituting the values, we get,

$$= \frac{6.022 \times 10^{23} \times 8900 \times 10^{3}}{63.5} = 8.44 \times 10^{28}$$

Since the number of free electrons per atom in copper is 1,

the electrical conductivity $\sigma = \frac{ne^2\tau_r}{m}$

$$= \frac{8.44 \times 10^{28} \times (1.6 \times 10^{-19})^2 \times 10^{-14}}{9.1 \times 10^{-31}}$$

The electrical conductivity is 2.374×10^7 Ohm^{-1} m^{-1}.

Example 16.10 The resistivity of a piece of silver at room temperature is 1.6×10^{-8} Ω m. The effective number of conduction electrons is 0.9 per atom and the Fermi energy is 5.5 eV. Estimate the mean free path of the conduction electrons. Calculate the electronic relaxation time and the electronic drift velocity in a field of 100 V m^{-1}. The density of silver is 1.05×10^4 kg/m^3 $\left(\frac{m}{m*} = 1\right)$.

Given Data

The resistivity of the silver piece at room temperature $\rho = 1.6 \times 10^{-8}$ Ω m
The effective number of conduction electrons for atom $= 0.9$
Fermi energy of the silver piece $= 5.5$ eV
$= 5.5 \times 1.60 \times 10^{-19}$
$= 8.8 \times 10^{-19}$ J

Solution

We know that the conductivity $\sigma = \frac{1}{\rho}$

$$= \frac{1}{1.6 \times 10^{-8}}$$

$$= 6.25 \times 10^7$$

The conductivity of the silver piece is 6.25×10^7 Ω^{-1} m^{-1}

$$n = 6.023 \times 10^{23} \times \frac{1.05 \times 10^4}{107.9 \times 10^{-3}} = 5.86 \times 10^{28} / m^3$$

We know that the relaxation time $\tau = \frac{\sigma m}{ne^2}$

Substituting the values we get,

$$= \frac{6.25\times10^{7}\times9.11\times10^{-31}}{5.86\times10^{28}\times(16.03\times10^{-19})^{2}}$$

$$\tau = 3.78\times10^{-14}$$

Therefore, the relaxation time is 3.78×10^{-14} s

The mean free path is given by $\lambda = c\tau$

$$= 3\times10^{8}\times3.78\times10^{-14}$$

$$\lambda = 1.134\times10^{-5} \text{ m}$$

We know that, the drift velocity is $V_d = \frac{\sigma E}{ne}$

Substituting the values, we get, $V_d = \frac{6.25\times10^{7}\times100}{5.86\times10^{28}\times1.6\times10^{-19}}$

Therefore, the drift velocity of electrons in the silver piece is 0.66 m s^{-1}

Example 16.11 A copper wire has a resistivity of 1.7×10^{-8} Ω m at room temperature of 300 K. If the copper is highly pure, find the resistivity at 700˚C.

Given Data

The resistivity of copper wire at room temperature 1.7×10^{-8} Ω m.

Solution

We know that $\rho \propto T^{1/2}$

i.e., $\frac{\rho 1}{\rho 2} = \frac{T_1^{1/2}}{T_2^{1/2}}$

$$\rho_2 = \frac{\rho_1 T_1^{1/2}}{T_2^{1/2}}$$

By rearranging the above equation, we get

$$\rho_2 = \rho_1\sqrt{\frac{T_1^{1/2}}{T_2^{1/2}}}$$

Substituting the value of ρ_1, T_1 and T_2, we get,

$$\rho_2 = 1.7\times10^{-8}\times\sqrt{\frac{(700+273)^{1/2}}{300^{1/2}}}$$

a resistivity of the copper wire is 3.0616×10^{-8} Ω m.

The McGraw·Hill Companies

Example 16.12 A conduction wire has a resistivelity of $1.54 \times 10^{-8}\ \Omega$ m at room temperature. The Fermi energy for such a conductor is 5.5 eV. There are 5.8×10^{28} conduction electrons per m^3. Calculate

(i) The relaxation time and the mobility of the electrons
(ii) The average drift velocity of the electrons when the electric field applied to the conductor is 1 V cm^{-1}
(iii) The velocity of an electron with Fermi energy
(iv) The mean free path of the electrons

Given Data

The resistivity $\rho = 1.54 \times 10^{-8}\ \Omega$ m
Fermi energy $E_F = 5.5$ eV
$= 5.5 \times 1.609 \times 10^{-19}$
$= 8.8495 \times 10^{-19}$ J
Concentration of electrons $n = 5.8 \times 10^{28}\ m^{-3}$

Solution

(i) We know that the relaxation time a

$$\tau_r = \frac{m}{\rho n e^2}$$

Substituting the value of ρ, E_F, n and e, we get,

$$\tau_r = \frac{9.1 \times 10^{-31}}{1.54 \times 10^{-8} \times 5.8^{28} \times (1.6 \times 10^{-19})^2}$$

$$= 3.97972 \times 10^{-14}\ \text{s}$$

Therefore the relaxation time is 3.98×10^{-14}
We know that the mobility of the electrons is

$$\mu = \frac{e\tau_r}{m}$$

$$\mu = \frac{1.6 \times 10^{-19} \times 3.97972 \times 10^{-14}}{9.1 \times 10^{-31}}$$

$$= 6.9973 \times 10^{-3}$$

The mobility of the electrons is $6.99 \times 10^{-3}\ m^2\ V^{-1}\ s^{-1}$.

(ii) We know that the drift velocity of the electrons.

$$v_d = \frac{e\tau_r E}{m}$$

Substituting the value of e, τ_r, E and m in the above equation, we get,

$$\mu = \frac{1.6 \times 10^{-19} \times 3.98 \times 10^{-14}}{9.11 \times 10^{-31}} \times 1 \times 10^2$$

$$= 0.69973$$

The drift velocity of the electrons is 0.69973 m s^{-1}.

(iii) We know that the Fermi velocity is

$$V_F = \sqrt{\frac{2E_F}{m}}$$

$$= \sqrt{\frac{2\times5.5\times1.6\times10^{-19}}{9.1\times10^{-31}}} = 1.3907 \times 10^6 \text{ m s}^{-1}$$

Therefore, the Fermi velocity is 1.39×10^6 m s^{-1}.
We also know that the mean free path $\lambda = V_F\tau_r$

$$= 1.3907 \times 10^6 \times 3.97972 \times 10^{-14}$$

$$= 5.5346 \times 10^{-8}$$

The mean free path is 5.535×10^{-8} m.

Objectives Questions

16.1 _____ is the macroscopic theory.
16.2 Quantum free electron theory is a macroscopic theory and it obeys _____ law.
16.3 Zone theory is based on _____ of solids.
16.4 The reciprocal of electrical resistivity is known as _____.
16.5 The unit for electrical resistivity is _____.
16.6 The free electrons in a metal are known as _____.
16.7 Electrical conductivity of low-resistivity materials is in the range of _____.
16.8 Examples for high-resistivity materials _____, _____ and _____.
16.9 According to Newton's second law of motion
(a) $F = ma^2$ (b) $F = ma$
(c) $F = m^2a^2$ (d) $F = \sqrt{ma}$
16.10. The thermal conductivity of a metal is

(a) $\sigma = \frac{ne^2\tau_r}{m^2}$ (b) $\sigma = \frac{ne^2\tau_r^2}{m^2}$

(c) $\sigma = \frac{ne^2\tau_r}{m}$ (d) $\sigma = \frac{ne\tau_r}{m^2}$

16.11. The resistivity of a material

(a) $\rho = \frac{m}{ne^2\tau_r}$ (b) $\rho = \frac{m}{ne\tau_r}$

(c) $\rho = \frac{m}{ne^2\tau_r^2}$ (d) $\rho = \frac{m^2}{ne^2\tau_r}$

The McGraw·Hill Companies

16.12 The mobility of the electron μ_e = _____

16.13 The thermal conduction in a metal is equal to K_{total} = _____

16.14 The thermal condition for an insulator is K_{total} = _____

16.15 The coefficient of thermal conductivity K is equal to

(a) $\frac{1}{2} n^2 v k_B \lambda$ (b) $\frac{1}{2} n^2 v^2 k_B \lambda$

(c) $\frac{1}{2} n v^2 k_B \lambda$ (d) $\frac{1}{2} n v k_B \lambda$

16.16 According to Wiedemann–Franz law

(a) $\frac{K}{\sigma} \propto T^2$ (b) $\frac{K}{\sigma} \propto T$

(c) $\frac{K}{\sigma} = T^2$ (d) $\frac{K}{\sigma} \propto T^2 L$

16.17 Ohm's law is verified using the classical free electron theory. True/False

KEY

16.1 Classical free electron
16.2 quantum
16.3 energy bands
16.4 Electrical conductivity
16.5 Ω^{-1} m^{-1}
16.6 electron gas
16.7 10^8
16.8 Tungsten, Platinum, nichrome
16.9 b)
16.10 c)
16.11 a)
16.12 $e \tau_r / m$
16.13 $K_{electron} + K_{phonon}$
16.14 K_{phonon}
16.15 d)
16.16 b)
16.17 True.

Short Questions

16.1 What is meant by electrical conductivity?
16.2 Define electrical conductivity.
16.3 What is meant by conductivity materials?
16.4 Define density of states and sketch the same for a metal?
16.5 What are the main theories developed for metals?
16.6 What is meant by free electrons?
16.7 What is classical free electron theory?
16.8 Mention any two important features of quantum free electron theory of metal?
16.9 Define electric collision.
16.10 What are the sources of resistance in metals?
16.11 Mention the postulates of free electron theory.

The McGraw·Hill Companies

16.12 Define relaxation time.
16.13 Distinguish between relaxation time and collision time.
16.14 Define mean free path.
16.15 Define the mobility of electrons?
16.16 Define drift velocity?
16.17 Distinguish between drift velocity and thermal velocity of an electron.
16.18 Define coefficient of thermal conductivity.
16.19 Define electrical conductivity.
16.20 Distinguish between thermal and electrical conductivity.
16.21 State Wiedemann–Franz law.
16.22 What is meant by Lorentz number?
16.23 What are the simultaneous between the electrical and thermal conductivity?
16.24 Verify Ohm's law using classical free electron theory.
16.25 Mention the application of classical free electron theory.
16.26 What are the successes of classical electron theory?
16.27 What are the drawbacks of classical free electron theory?

Descriptive Questions

16.1 What are the postulates of free electron theory? Drive an expression for electrical conductivity based on free electron theory?
16.2 Derive an expression for electrical conductivity in a material in terms of mobility of electrons and hence obtain Wiedemann–Franz law.
16.3 Define electrical and thermal conductivity and deduce the expression for the same.
16.4 Deduce Wiedemann Franz law classically and quantum mechanically.
16.5 Describe the classical free electron theory and deduce Lorentz number. Also, discuss the merits and demerits of classical free electron theory.

Chapter

17

Statistics and Band Theory of Solids

OBJECTIVES

- To derive the Fermi Dirac statistics for the energy states
- To study the probability of occupation of electrons
- To explain the density of states in metal or semiconductors
- To study the effect of temperature on Fermi energy function

17.1 INTRODUCTION

The classical free electron theory could not explain specific heat capacity, electronic specific heat etc. But they are well explained using quantum free electron theory. On the other hand, quantum free electron theory fails to explain the distinction among metals, semiconductors and insulators, the positive value of Hall Effect and some transport properties. All the above failures are overcome by the energy band theory of solid. The band theory of the solids is used to explain the band structure and the electrical properties of solids. The outcome of quantum mechanics, Pauli's exclusion principle and Bohr's atom model are results in the band theory of solids.

The statistics mainly used to derive the velocity and the energy distribution of particles is of three types namely, (i) Maxwell-Boltzmann statistics, (ii) Bore-Einstein statistics and (iii) Fermi-Dirac statistics. The Maxwell – Boltzmann statistics deals with particles having no spin (Eg. Gaseous particles). The Bose-Eintein statistics explains the particles like photon having integral spins, while the particle with half integral spin was dealt by the Fermi-Dirac statistics. Examples for Fermi-Dirac statistics are the particles like electrons which have half integral spin and hence, these particles are known as Fermi particles or Fermions. The Fermi-Dirac statistics, Fermi energy and density of static band theory of solids are briefly discussed in this chapter.

17.2 FERMI-DIRAC STATISTICS

The distribution of energy states in a semiconductor is explained by Fermi-Dirac statistics since it deals with the particles having half integral spin like electrons. Consider that the assembly of electrons as electron gas which behaves like a system

The McGraw·Hill Companies

of Fermi particles or Fermions. The Fermions obeying Fermi-Dirac statistics i.e., Pauli's exclusion principle.

Therefore, the probability function $f(E)$ of an electron occupying an energy level E is given by,

$$f(E) = \frac{1}{1+\exp\left[\left(E - E_F\right)/KT\right]} \tag{17.1}$$

where E_F is known as Fermi energy, K the Boltzmann constant and T the absolute temperature.

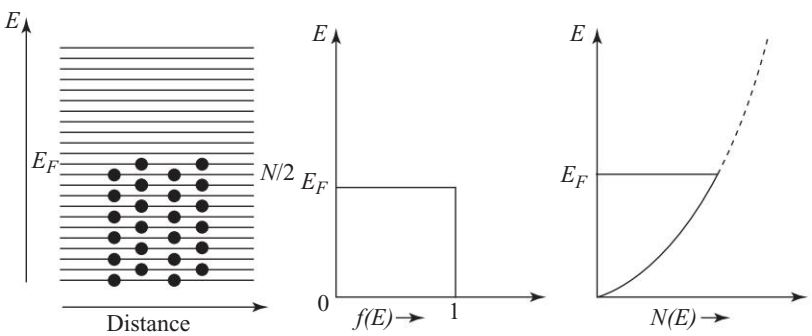
Eq.(17.1), the probability function $f(E)$ lies between 0 and 1. Hence, there are three possible probabilities namely:

$f(E) = 1$ 100 % Probability to occupy the energy level by electrons

$f(E) = 0$ No probability to occupy the energy levels by electrons and hence, it is empty

$f(E) = 0.5$ 50 % Probabilityof finding the electron in the energy level.

A detailed explanation for the above three cases are given by considering the temperature dependence of Fermi distribution function and its effects on the occupancy of energy level by electrons is shown in Fig. 17.1.



(a) *Distribution of electrons in conduction band at 0 K.*

(b) *Fermi Dirac distribution function at 0 K*

(c) *Electron distribution function*

Fig. 17.1 *Femi-Dirac distribution function*

Case I Probability of occupation at $T = 0\ K$, and $E < E_F$

Substituting the above condition in Eq. (17.1). We get,

$$f(E) = \frac{1}{1+e^{-\infty}}$$

$$= \frac{1}{1} = 1$$

Therefore, $f(E) = 1$ (17.2)

The McGraw·Hill Companies

Eq. (17.2), clearly indicates that at $T = 0$ *K*, the energy level below the Fermi energy level E_F is fully occupied by electrons leaving the upper levels vacant. Therefore, there is a 100 % probability that the electrons to occupy energy level below Fermi energy. Generally, a single energy band namely conduction band diagram as shown in Fig. 17.1 ais used to characterise the conductors.The conduction band have more energy levels than free electrons. Fig. 17.1 a indicates that conduction band is filled up to certain energy level marked as E_F, while the energy levels above the highest filled level is empty. Fermi energy level is defined as the upper most filled energy level in a conduction band at 0 *K*. In other words, the highest energy level occupied by an electron in a conductor at 0 *K* is the Fermi energy level.

Case II Probability of occupation at $T = 0$ *K* and $E > E_F$

Substituting the above condition in Eq. (17.1) We get,

$$f(E) = \frac{1}{1+e^{\infty}}$$

$$= \frac{1}{1+\infty}$$

or $$= \frac{1}{\infty} = 0$$

Therefore, $$f(E) = 0 \qquad (17.3)$$

Eq. (17.3) clearly indicates that at $T = 0$ *K*, the energy levels above the Fermi energy level E_F is unoccupied i.e., vacant. Therefore, there is a 0 % probability for the electrons to occupy the energy level above the Fermi energy level. Fig. 17.1 b shows that the variation of $f(E)$ for different energy values is a step function.

Case III Probability of occupation at $T = 0$ *K* and $E = E_F$

Substituting the above condition in Eq. (17.1) We get,

$$f(E) = \frac{1}{1+e^{0}}$$

$$= \frac{1}{1+1}$$

or $$= \frac{1}{2} = 0.5$$

Therefore, $$f(E) = 0.5 \text{ or } 1/2 \qquad (17.4)$$

The McGraw·Hill Companies

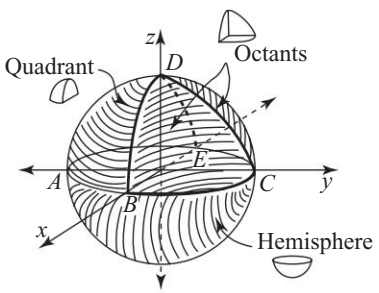
Eq. (17.4) clearly indicates that at T = 0 K, there is a 50 % probability for the electrons to occupy Fermi energy level. Therefore, the Fermi energy level is defined as the energy level at any temperature, the probability of electron is 50 % or ½. Similarly, it is also defined as E_F the average energy occupied by the electrons which participate in conduction process in conductor at temperature above 0 *K*.

Case IV At High temperature ($T > 0$ *K*) i.e., $KT >> E_F$ or $T = \infty$

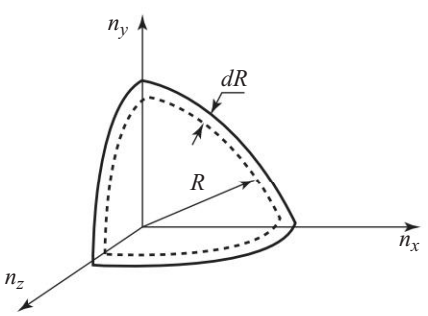
At higher temperature, electrons are excited above the Fermi energy level which is vacant. Therefore, most of the electrons are lying in the deep conduction band, without any disturbance. Since, the energy *KT* is not sufficient to make any transition to an unoccupied level. This energy is sufficient to make a jump to higher level those electrons lie in the energy level adjunct to Fermi level. At higher temperatures, the electron lost the quantum mechanical properties and the traditional classical Boltzmann distribution function.

17.2.1 Density of States

The energy distribution of electrons in a metal or semiconductor is determined using Fermi Dirac statistics. The n_x, n_y, n_z are quantum numbers used to construct the point in the sphere.



(a) *Division of a Sphere*



(b) *Octant*

Fig. 17.2 *Density of States*

The McGraw·Hill Companies

The ability of a metal to conduct electricity depends on the number of quantum states and the available energy levels of the electrons. Therefore, the determination of energy states for the electrons is essential. The numbers of quantum states present in a metal between the energy interval E and $E + dE$ per unit volume is known as density of states.

$$\text{Density of states } Z(E) = \frac{\text{No. of quantum states present between } E \text{ and } E + dE}{\text{Volume of the specimen}} \quad (17.5)$$

Consider that the sphere is further divided into number shells as shown in Fig. 17.2 a. Therefore, each shell is represented by a set of quantum numbers (n_x, n_y and n_z) and will have an associated energy. Let E be the energy of the point which is same for all points present on the sphere. Therefore, the radius of the sphere with energy E is

$$n^2 = n_x^2 + n_y^2 + n_z^2 \quad (17.6)$$

Consider that the unit volume represents one energy state. Therefore, number of energy states in a sphere of radius n is

$$= \frac{4}{3}\pi n^3 \quad (17.7)$$

Eq.(17.7) represents the total volume of the sphere. We know that the quantum number n_x, n_y, n_z takes only the positive integral values and hence, one has to take only one octant in the sphere i.e., 1/8 of the total volume of the sphere.

Therefore, the every states available within one octant of the sphere of radius n and its energy E is

$$= \frac{1}{8}\left[\frac{4}{3}\pi n^3\right] \quad (17.8)$$

Similarly, the energy states available within one octant of the sphere of radius $n + dn$ and its energy $E + dE$ is

$$= \frac{1}{8}\left[\frac{4}{3}\pi (n+dn)^3\right] \quad (17.9)$$

Therefore, the number of energy states available within the sphere of radius n and n+dn is obtained by finding the energy difference between the two energy levels namely E and E + dE.

$$N(E)\, d(E) = \frac{1}{8}\left[\frac{4}{3}\pi (n+dn)^3\right] - \frac{1}{8}\left[\frac{4}{3}\pi n^3\right]$$

$$= \frac{1}{8}\left(\frac{4\pi}{3}\right)\left[(n+dn)^3 - n^3\right]$$

Expanding $(n + dn)^3$, we get,

$$N(E)\, d(E) = \frac{\pi}{6}\left(n^3 + dn^3 + 3n^2 dn + 3n dn^2 - n^3\right) \quad (17.10)$$

Simplifying the above Eq. (17.10) using $(a+b)^3 = a^3 + b^3 + 3a^2b + 3ab^2$, we get,

$$N(E)\, d(E) = \frac{\pi}{6}\left(dn^3 + 3n^2 dn + 3ndn^2\right)$$

Neglecting the higher powers of dn i.e., dn^2 and dn^3, we get,

$$N(E)\, d(E) = \frac{\pi}{6}\left(3n^2 dn\right)$$

Simplify the above equation we get,

$$N(E)\, dE = \frac{\pi}{2} n^2 dn \tag{17.11}$$

(or)
$$N(E)\, dE = \frac{\pi}{2} n(ndn) \tag{17.12}$$

Consider a cubic metal piece with cube edge *l*. Therefore, the energy of electron within the cube is,

$$E = \frac{n^2 h^2}{8ml^2}$$

Rearranging the above equation, we get,

$$n^2 = \frac{8ml^2}{h^2} dE \tag{17.13}$$

(or)
$$n = \left[\frac{8ml^2 E}{h^2}\right]^{1/2} \tag{17.14}$$

The value of ndn is obtained by differentiating Eq. (17.13)

i.e.,
$$2ndn = \frac{8ml^2}{h^2} dE$$

(or)
$$ndn = \frac{8ml^2}{2h^2} dE \tag{17.15}$$

Substituting the value of *n* and ndn from the Eq. (17.14) and Eq. (17.15) in Eq. (17.12), we get,

$$N(E)dE = \frac{\pi}{2}\left(\frac{8ml^2 E}{h^2}\right)^{1/2}\left(\frac{8ml^2}{2h^2}\right) dE$$

Simplifying the above equation, we get,

$$N(E)\, dE = \frac{\pi}{4}\left(\frac{8ml^2}{h^2}\right)^{3/2} E^{1/2} dE \tag{17.16}$$

where l^3 is the volume of the metal piece.

The McGraw·Hill Companies

According to Pauli's exclusion principle, we know that two electrons with opposite spin occupy each state. Therefore, the number of energy states available for electron occupancy is

$$N(E)\,dE = 2 \times \frac{\pi}{4}\left(\frac{8m^2}{h^2}\right)^{3/2} E^{1/2} dE \qquad (17.17)$$

Eq.(17.17) is the effective number of energy states in a volume l^3 with energy between E and $E + dE$. The number of available energy states per unit volume $l^3 = 1$.

We know that,

Density of states $\quad Z(E)\,d(E) = \dfrac{N(E)\,d(E)}{V}$

Substituting the value of $N(E)\,d(E)$ and $l = 1$, in the above equation, we get,

Density of states $\quad Z(E)\,dE = \dfrac{\pi}{2}\left(\dfrac{8m}{h^2}\right)^{3/2} E^{1/2} dE \qquad (17.18)$

Eq. (17.18) is the density of energy states i.e., charge carrier in the energy interval E and $E + dE$. One can determine the carrier concentration in the metals and semiconductors using the above equation.

17.3 CARRIER CONCENTRATION (FREE ELECTRON DENSITY) IN METALS

The concentration of the carrier n_c i.e., number of electrons per unit volume in the given energy levels E and $E + dE$ is obtained by integrating the product of the density of states and the probability of occupancy $f(E)$.

Therefore, $\quad n_c = \int Z(E)\, f(E)\, dE \qquad (17.19)$

Substituting the values of $Z(E)$ and $f(E)$ in the above equation, we get,

$$n_c = \int \frac{4\pi}{h^3}(2m)^{3/2} E^{1/2} \frac{1}{1 + e^{(E - E_F)/KT}}\, dE \qquad (17.20)$$

Eq. (17.20) gives the carrier concentration in the given energy level $E + dE$.

17.4 EFFECT OF TEMPERATURE ON FERMI ENERGY FUNCTION

17.4.1 Determination of Fermi Energy at 0 K

In metals, at 0 K, the electrons occupy up to the Fermi level E_F. Therefore, $f(E) = 1$ for energy levels between $E = 0$ and $E = E_F$.

Substituting the above values in Eq. (17.20), we get,

$$n_c = \int \frac{4\pi}{h^3}(2m)^{3/2} E^{1/2} dE$$

The evaluation of integration in the above equation gives the concentration of electrons in a band in terms of Fermi energy level E_F. At $T = 0$, $E_F = E_{Fo}$ i.e., E_{Fo} is the highest energy level occupied by the energy at $T = 0$ *K*. The above integration is valid only $E < E_{Fo}$.

Therefore, integrating the above equation at $T = 0$ *K*, we get,

$$E_{Fo} = \left(\frac{h^2}{8m}\right)\left(\frac{3n}{\pi}\right)^{2/3} \tag{17.21}$$

(or) $$E_{Fo} = B\, n^{2/3} \tag{17.22}$$

where $B = \left(\frac{h^2}{8m_e}\right)\left(\frac{3}{\pi}\right)^{2/3}$ is a constant and is equal to 5.85×10^{-38} J.

Eq. (17.22) is used to determine the highest energy level i.e., Fermi energy level (E_{FO}) occupied by an electron at $T = 0$ *K* by knowing the value of the carrier concentrations. The value of E_{FO} for few important metals is given in Table 17.1.

Table.17.1 *Highest Energy level E_{FO} at 0 K*

Metals	*Values of n* ($\times 10^{28}$ m^{-3})	E_{FO}
Cu	8.491	7.06
Zn	13.153	9.45
Al	18.066	11.68

Mean energy of Electron at T = 0 K Consider that E_{ave} is the average energy of the electron at 0 K,

Therefore,

$$E_{ave} = \frac{1}{n}\int_o^{E_{Fo}} N(E)\,dE\; E \tag{17.23}$$

Substituting the values of $N(E)dE$ in the above equation, we get,

$$E_{ave} = \frac{1}{n}\int_o^{E_{Fo}} \frac{4\pi}{h^3}(2m)^{3/2} E^{1/2} E\, dE$$

(or) $$= \frac{1}{n}\frac{4\pi}{h^3}(2m)^{3/2}\int_o^{E_{Fo}} E^{3/2} dE \tag{17.24}$$

Integrating the term, we get,

$$E_{ave} = \frac{1}{n}\frac{4\pi}{h^3}(2m)^{3/2}\frac{2}{5}E_{FO}^{5/2}$$

(or) $$= \frac{1}{n}\frac{4\pi}{h^3}(2m)^{3/2}\frac{2}{5}E_{FO}^{3/2}E_{FO}$$

Substituting the values of E_{FO} from Eq. (17.21), in the above equation, we get,

$$E_{ave} = \frac{1}{n}\frac{4\pi}{h^3}(2m)^{3/2}\frac{2}{5}\left(\frac{h^2}{2m}\left(\frac{3n}{8\pi}\right)^{2/3}\right)^{3/2} E_{FO} \qquad (17.25)$$

Rearranging, the above equation, we get,

$$E_{ave} = \frac{1}{n}\frac{4\pi}{h^3}(2m)^{3/2}\ \frac{2}{5}\ \frac{h^3}{(2m)^{3/2}}\left(\frac{3n}{8\pi}\right) E_{FO} \qquad (17.26)$$

Simplifying the above equation, we get,

$$E_{ave} = \frac{3}{5}E_{Fo} \qquad (17.27)$$

Eq. (17.27) gives the average Fermi energy values occupied by electrons at 0 *K*. It is clear that Fermi energy of solids depends on density of electrons in the solid.

17.4.2 Fermi Energy of Electrons at Non Zero Temperature (T > 0 K)

When the temperature increases, i.e., $T > 0\ K$ and hence $(E - E_F) >> KT$. Therefore, the Fermi Dirac distribution function can be written as,

$$f(E) = e^{-(E-E_F)/KT} \text{ for } E - E_F > KT$$

Therefore, the carrier concentration of electrons

$$n = \int_{o}^{\text{top of the band}} f(E)\ N(E)\ dE \qquad (17.28)$$

Substituting the value of $f(E)$ and $N(E)$, we get,

$$n = \frac{4\pi}{h^3}(2m)^{3/2}\int_{o}^{\alpha} E^{1/2}\, e^{-(E-E_F)/KT} dE \qquad (17.29)$$

Substituting the value of n from the above equation (17.28) in Eq. (17.29), we get,

$$E_{ave} = \frac{1}{n}\int_{o}^{\alpha} E\, N(E)\, e^{-(E-E_F)/KT}\, dE \qquad (17.30)$$

Fermi energy at any temperature is obtained by evaluating the above equation,

i.e., $$E_F(T) = E_{Fo}\left[1 - \frac{\pi^2}{12}\left[\frac{kT}{E_{FO}}\right]^2\right] \qquad (17.31)$$

and

$$E_{avg}(T) = \frac{3}{5}\ E_{Fo}\left[1 - \frac{5\pi^2}{12}\left[\frac{kT}{E_{FO}}\right]^2\right] \qquad (17.32)$$

Eqs. (17.31) and (17.32) show that Fermi energy level decreases with increase in temperature, while average energy increase with increase in temperature. Thus,

the variation in Fermi energy level with temperature is very small and hence, it is neglegiable.

17.5 SIGNIFICANCE OF FERMI ENERGY

Fermi energy levels are used to explain the flow of electrons when two metals are in contact, p-type and n-type semiconductors and Seebeck effect. Let us discuss briefly about the significance of Fermi energy levels in the following heading:

17.5.1 Metal – Metal Contacts

Consider two different metals namely platinum (*Pt*) and molybdenum (*Mo*) respectively with work function 5.36 eV and 4.17 eV. The higher work function of Pt is higher than Mo which gives rise to higher Fermi energy level in Mo than Pt. When these metals are brought in contact as shown in Fig.17.3, the electrons from the higher energy level flows into lower energy level i.e., Mo to Pt. Therefore, the inner surface of *Pt* at the Pt-Mo metal contact region becomes more negative while the inner surface of *Mo* becomes positive. The flow of electrons takes place until two energy levels are equal. As a results, a contact potential is developed across the Pt-Mo contact region which is equal to 1.16 eV (5.36–4.2eV). The developed contact potential stops the further flow of electrons from Mo to Pt metals.

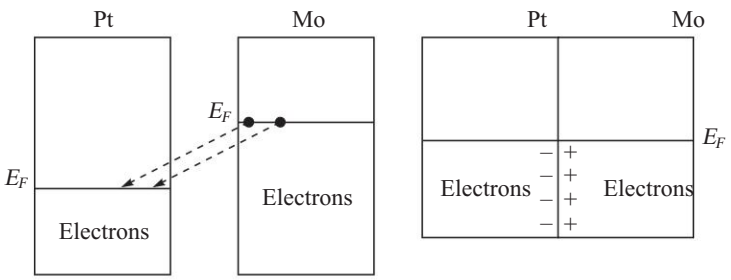


Fig. 17.3 *Pt-Mo metals contact -Contact Potential*

Similarly, the flow of electrons and holes in *pn* junction semiconductor are explained using the Fermi energy levels. When a *pn* junction is formed there is a difference in the Fermi energy levels between the *p*-type and *n*-type semiconductors. As a result there is a flow of electrons from *n*-type to p-type and there is a flow of holes from *p*-type to *n*-type semiconductors. The transfer of electrons and holes take place until the Fermi levels in both the semiconductors reaches same level.

17.5.2 Seebeck Effect and Thermocouple

Consider that the two ends of the metal are kept at the two different temperatures say 273 *K* and 373 *K* as shown in Fig.17.4. Therefore in view of the difference in the temperatures between the two ends, flow of electrons from hot end to cold end starts. Therefore more electrons are accumulated at the cold end and hence, it becomes more negative when compared to the hot end.Therefore, it develop a potential difference between the two ends of the metal, which is called as **Seebeck effect**. Once can use

the seebeck effect to measure the thermo emf across the metals which are kept at different temperatures using the thermocouple.

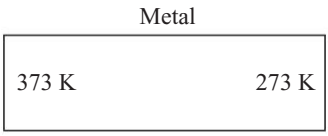


Fig. 17.4 *Seebeck effect*

Consider two copper wires having the same Fermi energy levels are connected at their ends. One end of the wire is kept as cold end i.e., at 273 K while the other end is kept as hot end i.e., at 373 K. A voltmeter connected across the hot and cold ends as shown in Fig. 17.5a is used to measure the thermo emf generated between the ends. There is no flow of electrons between the hot and cold ends, since the same copper wire is used having with same Fermi energy levels. Therefore, there is no thermo emf across the ends and hence voltmeter does not show any deflections.

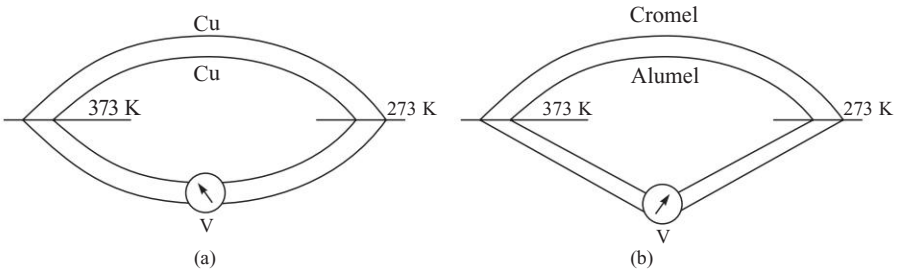


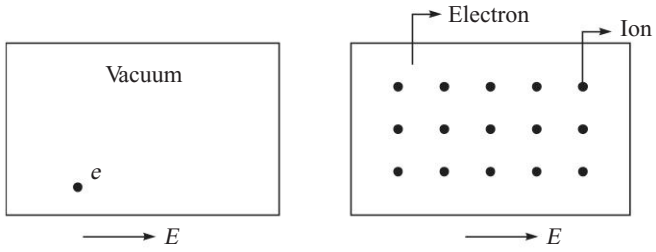
Fig. 17.5 *Thermocouple*

On the other hand, when two different metals having different Fermi energy levels say Cromel (eV) and Alumel (eV) metals are connected and kept at two different temperatures,it generate an emf as shown in Fig.17.5b.Therefore, electrons flow from hot to cold end and hence, the voltmeter shows a deflection. The cromel-alumel metal connects act as a thermocouple and is known as *K*-type thermocouple. The thermocouples are obtained by connecting two different materials having different Fermi levels. The different types of thermocouples and the materials used are given in Table 17.2.

17.6 EFFECTIVE MASS OF AN ELECTRON

Consider the influence of an external field E on the electron when it is in a vacuum as shown in Fig. 17.6(a).The external field results an increase in the force acquired by electron. According to Newton's law of motion, the force acquired by the electron is $F = ma$. Let F_{ext} be the force experienced by the electron when an external field is applied. Therefore, the acceleration of the electron

$$a = \frac{F_{ext}}{m} \tag{17.33}$$



(a) *Vacuum* (b) *Crystal*

Fig. 17.6 *Electron under the influence of an electric field*

On the other hand, when an external electric field E is applied to the crystal (Fig. 17.6b), the electron in the crystal acquires force both due to the external and internal fields. The field which exists in a crystal due to the periodic arrangement of atoms or ions is known as internal field. Let F_{int} be the internal field available in the crystal. Therefore, the acceleration exerted by the electron

$$a = \frac{F_{ext} + F_{\text{int}}}{m} \tag{17.34}$$

One cannot measure the internal force experienced by the electrons directly. Therefore, an another mathematical expression is used to determine the acceleration exerted by the electron by considering the internal force into the system and replacing the mass of an electron as the effective mass (m^*) of the electron as,

$$a = \frac{F_{ext}}{m^*} \tag{17.35}$$

In Eq. (17.35), the effective mass (m^*) of the electron incorporates all the internal forces exist in the crystal. The behaviour of effective mass is same as that of classical mechanics. The effective mass of some of the important metals are shown in Table 17.2. Metals like *Cu* and *Ag* show that their effective mass is nearly equal to the mass of the electrons. The effective masses of electrons in different metals are displayed in Table 17.2.

Table 17.2 *Effective masses of electrons in different metals*

Metals	m_e^* / m
Cu	1.010
Ag	0.990
Au	1.100
Bi	0.047
K	1.120
Li	1.280
Na	1.200
Ni	28.000
Pt	13.000
Zn	0.850

The McGraw·Hill Companies

Expression for the Effective Mass Consider an electric field of strength E is applied to the crystal. The force exerted by the electron in the crystal is

$$F = -eE$$
$$= ma \tag{17.36}$$

According to quantum concept, the electrons behaves as a wave and hence, the group velocity is

$$v_g = \frac{d\omega}{dk} \tag{17.37}$$

where ω is the angular frequency, k is known as wave vector or wave number and its value is equal to $2\pi/\lambda$.

We know that the energy of the electron

$$E = \hbar\omega \tag{17.38}$$

Differentiating the above equation, we get,

$$dE = \hbar\, d\omega$$

or $$d\omega = dE / \hbar$$

Substituting the value of $d\omega$ in the above equation, we get the group velocity

$$v_g = \frac{1}{\hbar}\frac{dE}{dk} \tag{17.39}$$

The acceleration exerted by the electron is obtained by differentiating the above equation *w.r.t.t*,

The acceleration exerted by the electrons $a = \frac{dv_g}{dt} = \frac{1}{\hbar}\frac{d}{dt}\left(\frac{dE}{dk}\right)$

or $$a = \frac{1}{\hbar}\frac{d^2E}{dk^2}\frac{dk}{dt} \tag{17.40}$$

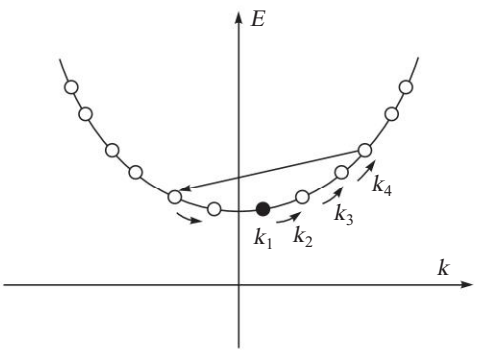


Fig. 17.7 *E-k diagram of conduction band*

One can determine the energy gained by the electron by considering the $E - k$ diagram of the conduction band as shown in Fig.17.7. When an electric field is applied, the electron gains energy and hence, it moves form one k value to another k value in the $E - k$ diagram. The accelerated electron coincides with lattice vibration and hence, random scattering of electrons takes place. The scattered electron takes into another k value as shown in Fig.17.7. Thus, the group velocity of the electron is determined from the gradient of the $E - k$ diagram. Let F_{ext} be the external force acting on the electron by the application of external field. The distance moved by the electron when it acquires energy externally is dt. Therefore, the energy gained by the electron

$$dE = F_{ext}\, v_g\, dt \tag{17.41}$$

Rearranging the above equation, we get,

$$F_{ext} = \frac{1}{v_g}\frac{dE}{dt} \tag{17.42}$$

Substituting the value of V_g from Eq. (17.39) in the above equation, we get,

$$F_{ext} = \hbar \frac{dk}{dt}$$

or

$$\frac{dk}{dt} = F_{ext}/\hbar \tag{17.43}$$

Substituting the value of $\frac{dk}{dt}$ from Eq. (17.43) in Eq. (17.40), we get,

$$a = \frac{F_{ext}}{\hbar^2\left[\frac{d^2E}{dk^2}\right]^{-1}} \tag{17.44}$$

Comparing the value of a with $a = F/m$, we get,

$$\frac{F}{m} = \frac{F_{ext}}{\hbar^2\left[\frac{d^2E}{dk^2}\right]^{-1}} = \frac{F_{ext}}{m^*} \tag{17.45}$$

where $m^* = \hbar^2\left[\frac{d^2E}{dk^2}\right]^{-1}$ known as the effective mass of the electron and it depends on the $E - k$ diagram.

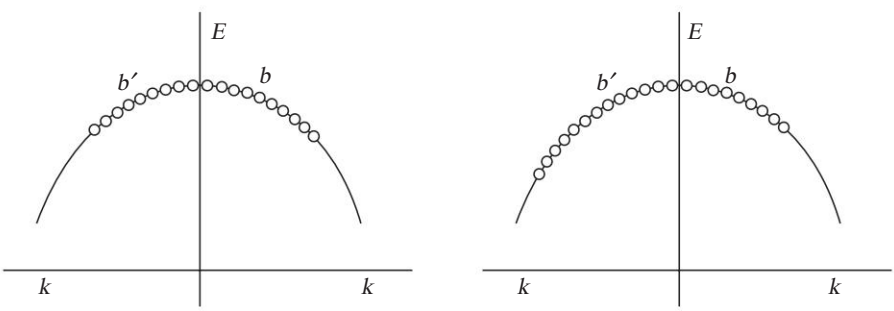
Consider an $E - k$ diagram for a valance band which is downward concave as shown in Fig.17.7. The effective mass of an electron, which is at the top of the valance band is negative. The negative sign of the effective mass indicates that the electrons are deaccelerating by the application of the external field. Generally, the effective mass of a crystal depends on curvature (d^2E / dk^2), on the $E - k$ diagram. Thus, the $E - k$ diagram of the crystal depends on the nature of bonding and the symmetry of

the crystal structure. Cyclotron resonance method is generally used to determine the effective mass of the crystal.

17.7 CONCEPT OF HOLE

One can explain the concept of hole by plotting a graph between energy of an electron E and wave vector k known as E-k diagram as shown in Fig.17.8. In Fig. 17.8a, all the states in both bands b and b' are filled by electron and hence, there is no vacancy state. On the hand in Fig.17.8b all the states are not filled in band b while there is a vacancy or empty state of an electron in band b. The absence of an electron or empty state is called as a hole.

(i) We know that the total wave vector in a filled energy band is zero. Let k be the wave vector of an electron in the band b and $-k$ be the wave vector of an electron in band b′. In Fig.17.8a, all the states are filled and hence, the total wave vector in a filled energy band is zero. However, in Fig.17.8b, there is once extra wave vector in band b' since there is an empty state in band b. This empty state or space is called as hole and its wave vector is k_h. Therefore, the wave vectors of the hole and electron are equal in magnitude but opposite in nature i.e., $k_h = -k_e$.



(a) *Filled states* (b) *Empty state*

Fig. 17.8 *E – k diagram of valence band*

(ii) The wave vectors of hole and electron are equal and opposite in magnitude. Therefore, the electron which is missing in the valence band is excited to conduction band.

(iii) The charge for a hole is positive i.e.,e^+, while the charge of an electron is negative i.e. e^-.

(iv) The energy of hole and electrons are equal but opposite in magnitude i.e., $\varepsilon_h\ (k_h) = \varepsilon_e\ (-k_e) = -\varepsilon_e\ (k_e)$.

(v) When an electric field is applied to a band, the negatively charged electrons moves opposite to the field direction (Fig. 17.9). On the other hand, the positively charged holes moves with same velocity as that of electron in the same direction of the field.

i.e., $v_h = v_e$

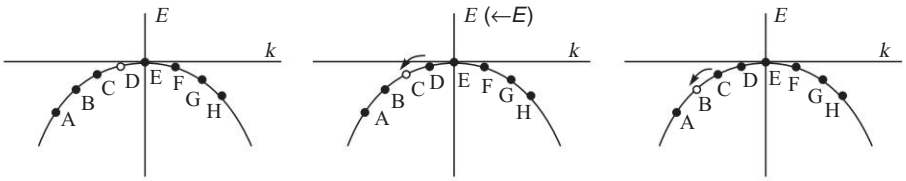


Fig. 17.9 *Motion of a hole—Applied field*

(vi) The effective mass of an electron at the top of the valence band is negative, while it is positive for hole due to its opposite charge as that of electron.

(vii) The equation of the motion of an electron is

$$\hbar \frac{dk_e}{dt} = -e\left(E + \bar{v}_e \, x \, \bar{B}\right) \tag{17.46}$$

Similarly, the equation for motion of a hole is,

$$\hbar \frac{dk_h}{dt} = e\left(E + \bar{v}_h \, x \, \bar{B}\right) \tag{17.47}$$

(viii) The equation for the current density for an electron is

$$J = -neV \tag{17.48}$$

Similarly, the equation for current density of a hole is

$$J = neV \tag{17.49}$$

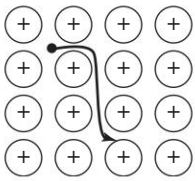
17.8 BAND THEORY OF SOLIDS – ORIGIN OF ENERGY GAP

The quantum free electron theory of metals assumes that the free electrons known as conduction electrons are fully responsible for the conduction mechanism.The above theory explains the phenomena like specific heat capacity, electrical conductivity, thermal conductivity, electrodynamics of metals, photoelectric effect, Compton effect etc. However, it fails to explain many other physical properties. This model fails to explain the distinction between metal, semiconductor and insulator.Similarly, it fails to explain the positive values of the Hall co-efficient. Therefore, the band theory of solids was proposed to overcome the failures of quantum free electron theory.

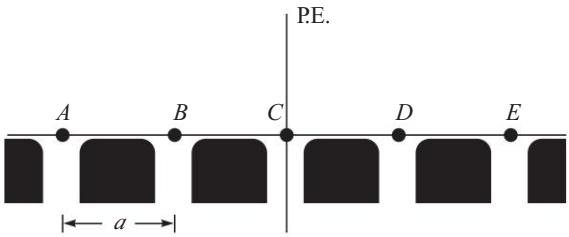
In order to explain the band theory of solids, cyclic or periodic boundary conditions is used to explain the free electrons. Consider the one dimensional periodic arrangement of positively charged ions inside a crystal as shown in Fig.17.10. The motion of the free electrons inside the positively charged ion core is shown in Fig.17.10a. Therefore, the wave function of the electron is

$$\psi(x) = A \, e^{jkx} \tag{17.50}$$

where k is the wave vector and A the amplitude.



(a) *Motion of electron*



(b) *Potential energy variation*

Fig. 17.10 *One dimensional periodic potential*

The potential energy of the freely moving electron at the positive ion site inside the crystal is zero while it is maximum at the middle. On the other hand, outside the crystal, the potential energy is infinite. Therefore, the potential energy of the moving electrons varies periodically with same period as that of crystal as shown in Fig.17.10b. Therefore, momentum (p) and energy (E) of the moving electron is respectively $\hbar k$ and $\left[(\hbar k)^2 / 2m\right]$.

When an electron moves in a periodic potential it undergoes similar effects such as reflection and diffraction as that of an ordinary wave. The electrons which satisfies the Bragg's condition for reflection during its motion in a periodic potential takes k vector.

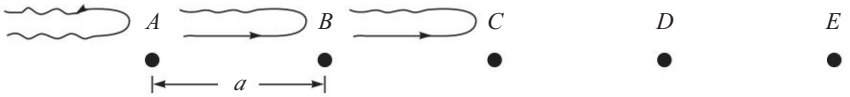


Fig. 17.11 *Reinforcement of electron waves*

Consider the arrangements of atoms in a periodic potential as shown in Fig.17.11. Let us consider electrons are moving along the positive x-direction in the periodic potential. When the electron wave incident on atom A, a part of the wave gets reflected back while, the remaining part of the wave gets transmitted through it. Similarly, when the part of the transmitted wave incident on atom B, a part of the wave gets reflected back while a part is transmitted through it and this process continues. All the reflected electron waves from atoms A, B, C etc. are reinforced when it satisfies the Bragg's condition. Therefore, the formation of standing wave pattern takes place due to reinforcement of waves when the path difference between the reflected waves from atoms A and B is $2a$. Therefore, interference patterns is obtained when

$$2a = n\lambda \tag{17.51}$$

where n is an integer and it takes the value 0, 1, 2, 3 etc and λ the wave length.

We know that wave vector $k = 2\pi/\lambda$, substituting the k value in the above equation, we get,

$$\text{Wave vector } k = n\frac{\pi}{a} \tag{17.52}$$

Eq. (17.41) show indicates that the standing wave patterns are formed due to the reinforcement of reflected beams whose wave vector is in the order of π/a.

Therefore, the equation for the wave function of the traveling wave is

$$\psi_+(x) = A\, e^{jkx} \tag{17.53}$$

Similarly, the wave function for the reflected wave is

$$\psi_-(x) = A\, e^{-jkx} \tag{17.54}$$

We know that the formation of the standing wave pattern takes place only when wave function $\psi_+(x)$ and $\psi_-(x)$ are reinforced. It is clear from Eq. (17.53) and Eq. (17.54), the reinforcement takes places in two different ways namely

$$\psi_1(x) = \psi_+(x) + \psi_-(x) \tag{17.55}$$

$$\psi_2(x) = \psi_+(x) - \psi_-(x) \tag{17.56}$$

Substituting $\psi_+(x)$ and $\psi_-(x)$ from respectively Eqns. (17.55) and (17.56) in the above equations, we get,

$$\psi_1(x) = Ae^{jkx} + Ae^{-jkx} = A\cos\frac{n\pi}{a}x \tag{17.57}$$

$$\psi_2(x) = Ae^{jkx} - Ae^{-jkx} = A\sin\frac{n\pi}{a}x \tag{17.58}$$

Let $\rho(\psi)$ be the probability density function or charge density of the standing wave pattern. Therefore, the probability density for the above two standing waves are

$$\rho(\psi_1) = |\psi_1|^2 = A^2\cos^2\frac{n\pi}{a}x \tag{17.59}$$

$$\rho(\psi_2) = |\psi_2|^2 = A^2\sin^2\frac{n\pi}{a}x \tag{17.60}$$

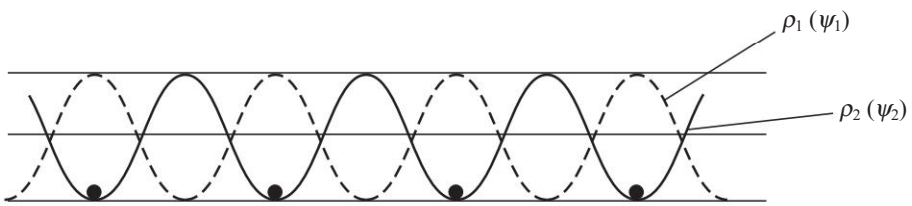


Fig. 17.12 *Piling of energy by charge density*

The McGraw·Hill Companies

The energies associated with standing wave of the above two wave function are used to determine the energy gap.

Brillouin Zones The energy of an electron moving in a periodic potential is limited within allowed regions or zones. One can determine the energy gap by finding the difference in the potential energies of the electrons piled with two different waves functions namely $\psi_1(x)$ and $\psi_2(x)$. Fig. 17.12 shows the piling of standing waves $\psi_1(x)$ and $\psi_2(x)$. The wave function $\psi_1(x)$ piles the maximum energy of the electrons at the positive ion cores. Therefore, the wave function $\psi_1(x)$ gives the potential energy of the standing wave.On the other hand, the wave function $\psi_2(x)$ piles the energy of the electrons at the middle of the positive ion cores. Thus, it increases the potential energy of the wave function than potential energy of the standing wave. Therefore, the difference between the potential energies of the electrons piled by these two wave functions $\psi_1(x)$ and $\psi_2(x)$ gives the energy gap i.e.,

Energy of electron $$E = \frac{(\hbar k)^2}{2m} \tag{17.61}$$

We know that, wave vector $k = \pm n\,(\pi/a)$ (17.62)

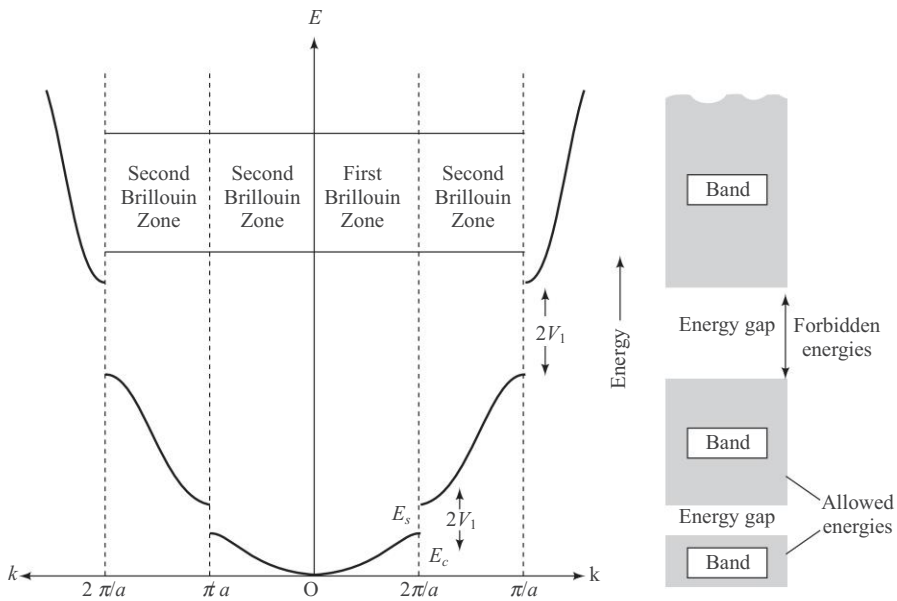


Fig. 17.13 *Energy gap—Brillouin zone boundaries*

The relationship between the energy and wave number for a one dimensional lattice is obtained by plotting a graph between total energy E and wave vector k as shown in Fig.17.13. Fig.17.13 demonstrates the origin of energy gap in electron energy at the Brillouin zone boundaries. The energy and wave vector graph shows discontinuities in the energy band at $k = \pm \pi/2, \pm \pi$ ... etc. It is clear from Fig.17.13

The McGraw·Hill Companies

that electron has an allowed energy value in the region or zone between $k = -\pi/a$ and $k = +\pi/a$. This zone is known as first Brillouin zone. Similarly, energy value lies in between the band $k = -\pi/a$ to $k = -\pi a$ and $k = +\pi/a$ to $k = +\pi$ a is known as second Brillouin zone. The break in between the first and second Brillouin zones is called as forbidden region or forbidden band or forbidden zone. Once can also define the higher order Brillouin zones.

Energy Gap One can determine the band gap energies by considering the interaction of the electron with the periodic potential of the lattice. Consider that an electron beam is incident on a two dimensional arrangement of lattice system as shown in Fig. 17.14. The incident electron gets reflection both at $k = \pm \pi/a$ and $k = \pm\frac{\sqrt{2}\,\pi}{a}$. When $k = \pm\pi/a$, the electron gets reflected from planes [0 1] and [1 0], while it gets reflected from plane [1 1] when $k = \pm\frac{\sqrt{2}\,\pi}{a}$.

We know that k

$$k \sin\theta = \pm\frac{n\pi}{a} \tag{17.63}$$

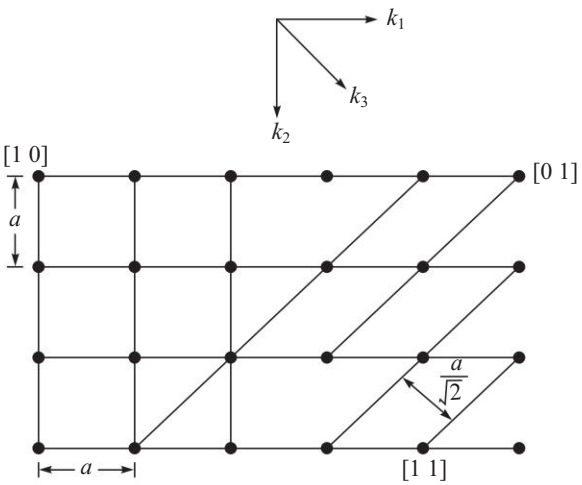
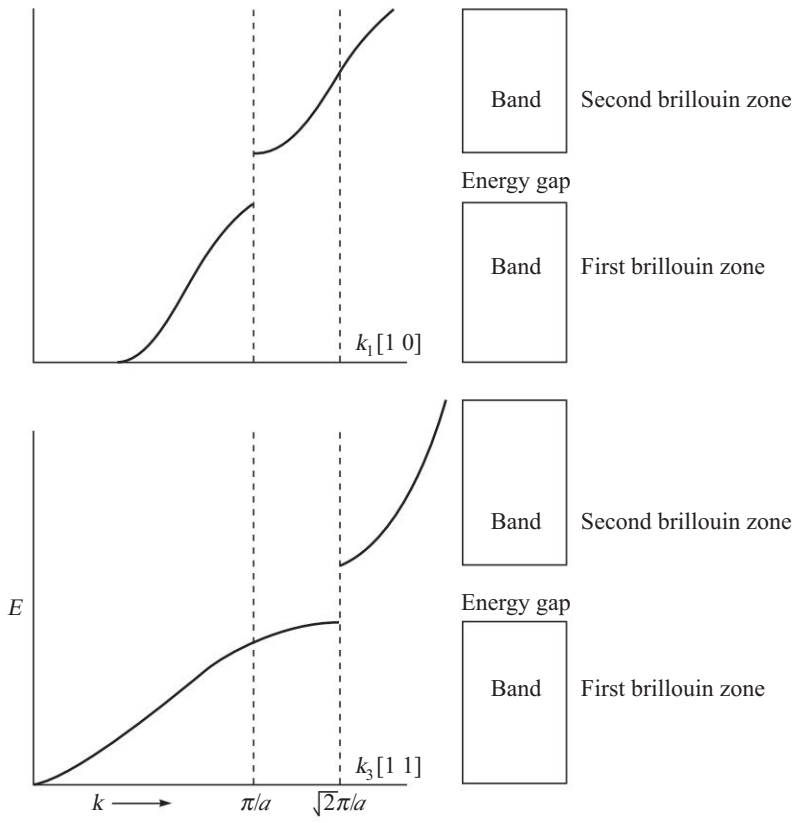


Fig. 17.14 *Reflection of wave vectors by* [0 1] *and* [1 1] *planes*

In order to determine the energy gap in the crystal lattice, the interaction of electron in all possible directions is to be considered. The minimum and maximum energy value in each band depends on the crystal direction. Therefore, the energy of the electron as a function of the crystal direction is shown in Fig.17.15.

For the [1 0] plane, E_1 is the maximum energy value while E_2 is the minimum energy value. However, for the [1 1] plane, E_1' is the maximum energy value which is higher than E_1 and E_2' is the lower energy value which is lower than E_2. Therefore, the energy gap $E_2' - E_1'$ is low when compared to $E_2 - E_1$.

The McGraw·Hill Companies

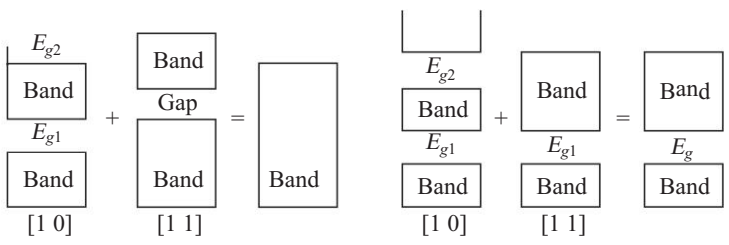


(a) *[1 0] plane* (b) *[1 1] plane*

Fig. 17.15 *E—k diagram*

Classification of Materials

The overlapping of the first and second Brillouin zone in a crystal lattice is shown in Fig. 17.16. The overlapping of the first Brillouin zone of [1 0] electron over the second Brillouin zone of [1 1] electron results a continuous distribution of energy.



(a) *Semiconductor or insulator of Metal* (b) *Metal*

Fig. 17.16 *Classification of materials*

Therefore, electron finds any energy (Fig. 17.16.a) and hence, the material is a metal. Similarly, the first Brillouin zone of [1 0] direction overlap with the first Brillouin zone of [1 1] direction results in an energy gap as shown in Fig. 17.16.b. Thus, the electron does not find any energy in between the bands and hence, the material is known as semiconductor or an insulator.

17.9 CONDUCTIVITY OF COPPER AND ALUMINUM

Material properties like electrical conductivity are widely used to select the materials for wide range of applications. Even though, copper and aluminum are monovalent metals, the resistivity of copper is higher than aluminum i.e., the electrical conductivity of *Cu* ($5.88 \times 10^7\ \Omega^{-1}\ m^{-1}$) at room temperature is lower than that of aluminum ($6.21 \times 10^7\ \Omega^{-1}\ m^{-1}$). We know that the electrical conductivity of a metal is,

$$\sigma = \frac{ne^2\tau_r}{m}$$

where n is the concentration of electrons and τ_r the relaxation time. μ is the mobility of the electrons and is equal to $\mu = e\ \tau_r/m$.

Therefore,

$$\text{Electrical conductivity } \sigma = ne\ \mu \tag{17.64}$$

The Table 17.3 displays some of the electrical properties of *Cu* and *Al*. The important properties of pure Cu such as highest conductivity next to silver, good mechanical strength, fairly resistant to corrosion and formability into wires or strips or rolled sheets helps to use for many electrical applications. The addition of metals like Fe (0.2%) or As (0.3%) reduces the its electrical conductivity nearly 50%. In view of the high resistivity of aluminum (1.6 times greater) than *Cu*, it replaces the Cu for many industrial applications. Manganin is an alloy of *Cu*, *Ni* and Manganese. Constantan is an alloy of Cu and *Ni*. It has constant resistance over a wide temperature range. For contact materials operating at high currents alloys of *Cu* like *Cu–W–*No is used. Solders, which are alloys having lower melting points than the metals to be connected. Alloys like *Sn–Pb–Cd* (melting point 145–180°*C*) and *Cu–Zn* alloy (hard solder) (melting point 825 – 860°*C*) are used for soldering.

Table 17.3 *Electronic properties of Cu and Al*

Metal	σ $(\Omega^{-1}\ m^{-1})$	n (m^{-1})	m^*/m_e	*Lattice Constant (a) A°*	E_F *eV*	μ_e $(m^2/\text{volt-sec.})$
Cu	5.88×10^7	8.5×10^{28}	1.01	3.61	7.00	4.3×10^{-3}
Al	6.21×10^7	5.85×10^{28}	0.99	4.05	11.8	7×10^{-3}

The McGraw·Hill Companies

17.10 EFFECT OF TEMPERATURE AND IMPURITY ON ELECTRICAL RESISTIVITY OF METALS (MATTHIESSEN'S RULE)

We know that the resistivity of the metals is due to the scattering of the conduction electrons. The resistivity of the metal is divided into two components namely ideal and residual resistivity. Generally in metals, the scattering takes place in the mechanism results in the component of resistivity namely ρ_{ph} and ρ_i. When the electrons are scattered by lattice vibration i.e., phonons free from all defeats give rise to ideal resistivity ρ_{ph}. The ideal resistivity is temperature dependent. On the other hand, when the electrons are scattered by the presence of the impurities and imperfection like dislocation vacancies, grain growth and boundaries, give rise to residual resistivity. The residual resistivity is independent of temperature and hence, the remaining does not be zero even at $T = 0\ K$. Therefore, the resistivity of a metal is

$$\text{Resistivity of metal } \rho = \rho_{ph} + \rho_i \tag{17.65}$$

Eq. (17.54) is known as Matthiessen's rule. According to Matthiessen's rule the total resistivity of a metal is the sum of resistivity due to phonons scattering which depends on temperature and the resistivity due to the scattering by impurities which is independent of temperature.

Assume that there is no scattering mechanism,

$$\rho_{ph} = \frac{m}{ne^2\tau_{ph}} \tag{17.66}$$

Simplify if there is known scattering by lattice vibration,

$$\rho_i = \frac{m}{ne^2\tau_i} \tag{17.67}$$

Substituting the ρ_{ph} and ρ_i in Eq. (4.1), we get,

$$\rho = \frac{m}{ne^2\tau_{ph}} + \frac{m}{ne^2\tau_i}$$

or

$$\rho = \frac{m}{ne^2}\left(\frac{1}{\tau_{ph}} + \frac{1}{\tau_i}\right) \tag{17.68}$$

Eq. (17.68) is used to explain the effect of temperature and impurities on electrical resistivity of metals. The temperature dependent resistivity of metal expects the superconducting metal is shown in Fig. 17.17.

At low temperature, the amplitude of lattice vibration is very small and hence, the scattered electrons also very less. Thus, it results a larger τ_{ph} and hence, the ρ_{ph} is almost equal to zero. Therefore, the total resistivity of the metal is equal to

$$\rho = \rho_i \tag{17.69}$$

It is clear from the Eq. (17.69), that the resistivity of a metal has a residual resistivity ρ_i at $T = 0\ K$ as shown in Fig. 17.17. On the other hand, when the temperature increases, the amplitude of lattice vibration increases and hence, the scattering of electrons by

the lattice also increases. Therefore, the resistivity of lattice vibration is added to total resistivity ($\rho_{ph} + \rho_i$) and hence, a linear increase in resistivity with increases in temperature is observed at low temperature as shown in Fig. 17.17.

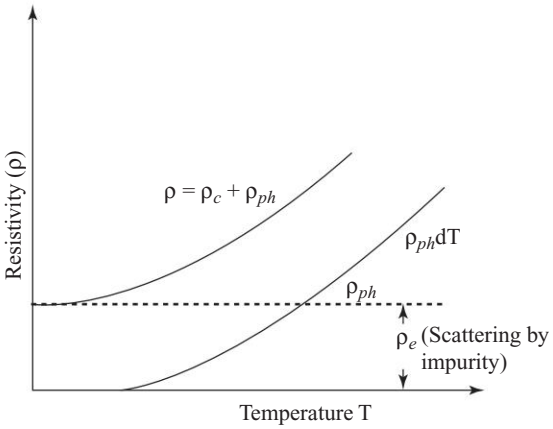


Fig. 17.17. *Temperature dependent resistivity of metal*

At high temperatures, the scattering effects due to phonons overcomes the effect due to scattering and impurities which give rise to the resistivity which is depend on lattice vibration only. As a result, at higher temperature the resistivity of the metals which depends purely on lattice vibrations varies exponentially than at low temperature. Similarly, at low temperature, the resistivity of metal depends on lattice vibration and residual resistivity and hence, shows a linear variation with temperature.

In a pure metal the scattering of electrons takes place only due to lattice vibrations. But in the case of impure metals, the scattering of electrons is both due to the lattice vibrations and impurity scattering. Therefore, the mean free time of electrons for impure metals is of two components namely, (i) the mean free time due to lattice vibration and (ii) the mean free time due to impurity atoms. The total mean free time is given by,

$$\frac{1}{\tau} = \frac{1}{\tau_T} + \frac{1}{\tau_I} \tag{17.70}$$

where τ_T is the mean free time due to thermal scattering (i.e., lattice vibrations) and τ_I the mean free time for electrons due to impurity scattering.

The drift mobility (μ_d) is also written as,

$$\frac{1}{\mu_d} = \frac{1}{\mu_T} + \frac{1}{\mu_I} \tag{17.71}$$

where μ_T is the mobility of electrons due to thermal scattering and μ_I the mobility of the electrons due to impurity scattering.

The electrical resistivity of an impure metal is written as,

$$\frac{1}{\rho} = \frac{1}{ne\mu_T} + \frac{1}{ne\mu_I} \tag{17.72}$$

$$\rho = \rho_T + \rho_I \tag{17.73}$$

The McGraw·Hill Companies

where ρ_T is the resistivity of the metal due to thermal scattering and ρ_I the resistivity of metal due to impurity scattering.The resistivity of the metal due to thermal scattering is known residual resistivity ρ_R.Hence, the resistivity of an impure metal is

$$\rho = \rho_R + \rho_I \tag{17.74}$$

The Eq. (17.74) is known as Matthiessen's rule.

Keypoints to Remember

- The particles like electrons with half integral spin obey Fermi Dirac statistics.
- Particle like electrons with half integral spin are known as Fermi particles or Fermions.
- The probability function of an electron occupying an energy level E is given by

$$f(E) = \frac{1}{1+\exp\left[(E-E)/KT\right]}$$

- Fermi energy level is the uppermost energy level in a conduction band at 0 K.
- At $T = 0$ K and $E > E_F$, the levels above Fermi levels is unoccupied i.e., vacant.
- The probability of electron to occupy the energy level above the Fermi energy level is Zero at $T = 0$ K and $E > E_F$.
- 50 % or ½ % is the probability for electrons to occupy Fermi energy level at $T = 0$ K and $E = E_F$.
- The number of quantum states present in a metal between any internal E and $E + dE$ per unit volume is known as density of states.
- Density of states $Z(E) = \dfrac{No.\text{ of quantum states present between } E \text{ and } E+dE}{\text{Volume of the specimen}}$
- The average volume of Fermi energy level $E_{ave} = 3/5\ E_{FO}$.

Solved Problems

Example 17.1 Calculate the Fermi energy of the following metals

	Cu	*Zn*	*Al*
ρ (g cm^{-3})	8.96	7.14	2.70
M_{at}	63.55	65.38	27.0

Given Data

For Cu

Density of Cu $= 8.96 \times 10^3$ kg m^{-3}
Atomic weight of Cu $= 63.55$

For Zn

Density of Zn $= 7.14 \times 10^3$ kg m^{-3}
Atomic weight of Zn $= 65.38$

For Al

Density of Al $= 2.7 \times 10^3$ kg m^{-3}
Atomic weight of Al $= 27.0$

Solution

(i) For Cu The concentration of electrons in *Cu*

$$n = \frac{\text{density} \times \text{Avogadro's constant} \times \text{no.of freee electrons per atom}}{\text{atomic weight}}$$

$$= \frac{8.96 \times 10^3 \times 6.022 \times 10^{26}}{63.55}$$

$$= 8.4905 \times 10^{28} \text{ m}^{-3}$$

Fermi energy at 0 *K*,

$$E_{FO} = \left(\frac{h^2}{8m_e}\right)\left(\frac{3n}{\pi}\right)^{2/3}$$

Substituting the values of *h*, *n* and m in the above equation, we get,

$$= \left(\frac{\left(6.626 \times 10^{-34}\right)^2}{8 \times 9.1 \times 10^{31}}\right)\left(\frac{3 \times 8.4905 \times 10^{28}}{\pi}\right)^{2/3}$$

$$= 1.129729 \times 10^{-18} \text{ J}$$

$$= 7.0608 \text{ eV}$$

(ii) For Zn The concentration of electrons in Zn

$$n = \frac{\rho N_A}{M_{at}} \times \text{no. of free electrons/atom}$$

$$= \frac{7.14 \times 10^3 \times 6.022 \times 10^{26} \times 2}{65.38}$$

$$= 13.152976 \times 10^{28} \text{ m}^{-3}$$

Fermi energy at 0 *K*,

$$E_{FO} = \left(\frac{h^2}{8m_e}\right)\left(\frac{3n}{\pi}\right)^{2/3}$$

Substituting the values of *h*, *n* and *m* in the above equation, we get,

$$= \left(\frac{\left(6.626 \times 10^{-34}\right)^2}{8 \times 9.1 \times 10^{31}}\right)\left(\frac{3 \times 13.153 \times 10^{28}}{\pi}\right)^{2/3}$$

$$= 1.51252397 \times 10^{-18}\ J$$

$$= 9.45 \text{ eV}$$

(iii) For Al The concentration of electrons in *Al*

$$n = \frac{\rho N_A}{M_{at}} \times \text{no. of free electrons/atom}$$

$$= \frac{2.7 \times 10^3 \times 6.022 \times 10^{23} \times 3}{27}$$

$$= 18.066 \times 10^{28}\ \text{m}^{-3}$$

Fermi energy at 0 *K*,

$$E_{FO} = \left(\frac{h^2}{8m_e}\right)\left(\frac{3n}{\pi}\right)^{2/3}$$

Substituting the values of *h*, *n* and m in the above equation, we get,

$$= \left(\frac{\left(6.626 \times 10^{-34}\right)^2}{8 \times 9.1 \times 10^{31}}\right)\left(\frac{3 \times 18.066 \times 10^{28}}{\pi}\right)^{2/3}$$

$$= 1.8689 \times 10^{-18}\ J$$

$$= 11.68 \text{ eV}$$

The Fermi energy of *Cu* is 7.0608 eV.

The Fermi energy of *Zn* is 9.45 eV.

The Fermi energy of *Al* is 11.68 eV.

Example 17.2 Calculate the density of states for Cu at the Fermi level for $T = 0\ K$, if (m*/m) = 1.01.

Given Data

Temperature = 0 K
The effective mass of electron in *Cu* = 1.01

Solution

The density of states is given by $Z(E) = \gamma\ E^{1/2}/2$

where $\lambda = 6.82 \times 10^{27}\ \text{m}^{-3}\ \text{eV}^{-3/2}$

Fermi energy at 0 *K*,

$$E_{FO} = \left(\frac{h^2}{8m_e}\right)\left(\frac{3n}{\pi}\right)^{2/3}$$

Substituting the values of *h*, *n* and *m* in the above equation, we get

$$= \left(\frac{(6.626 \times 10 - 34)^2}{8 \times 9.1 \times 10^{-31}}\right)\left(\frac{3 \times 8.4905 \times 10^{28}}{\pi}\right)^{2/3}$$

$$= 1.129729 \times 10^{-18}\ J$$

$$= 7.0608\ \text{eV}.$$

The value of E_F for *Cu* is 7.0608 eV.

Therefore,

$$Z(E) = \gamma\ E_F^{1/2}/2$$

$$= 3.39 \times 10^{27} \times \sqrt{7.0608}$$

$$= 9 \times 10^{27}\ \text{m}^{-3}$$

The density of states for *Cu* at the Fermi level for $T = 0\ K$ is $9 \times 10^{27}\ \text{m}^{-3}$

Example 17.3 The resistivity of Ni at 200 °C is 69 $n\Omega$m. If 80% Ni is alloyed with 20% chromium, the resistivity changes to 1120 $n\Omega$ m. Find the Nordheim's coefficient.

Given Data

Resistivity of *Ni* = 63 $n\Omega$ m

Resistivity of *Cr* = 129 $n\Omega$ m

Resistivity of 80% *Ni* + 20% *Cr* = 1120 $n\Omega$ m

Solution

The resistivity of alloy from Nordheim's relation is $\rho_I = C \times (1 - X)$

Where *X* is the atomic fraction of the solute atom.

Here, $X = 0.8$ therefore $\rho I = \text{C}\ X\ (1 - X)$

$1120 \times 10^{-9}\ \Omega\ \text{m} = \text{C} \times 0.8\ (1 - 0.8)$

$$\text{C} = \frac{1120 \times 10^{-9}}{0.8 \times 0.2}$$

$$= 7 \times 10^{-6}$$

The Nordheim's coefficient is $7 \times 10^{-6}\ \Omega\ \text{m}$

Example 17.4 Calculate the conductivity of Al at 25° C using the following data. Density of Al = 2.7 g cm^{-3}, atomic weight of Al = 27, and relaxation time of 10^{-14} s.

Given Data

Density of *Al* $= 2.7 \times 10^3$ Kg m^{-3}

Atomic weight of Al $= 27$

Relaxation time $= 10^{-14}$ s

Solution

The number of electron available per m^3 in *Al*,

$$n = \frac{\rho N_A}{M_{at}} \times \text{no. of free electrons per atom}$$

$$= \frac{2.7 \times 10^3 \times 6.022 \times 10^{26} \times 3}{27}$$

$$= 18.066 \times 10^{28} \text{ m}^{-3}$$

Conductivity of Al, $$n = \frac{ne^2\tau_r}{m}$$

$$= \frac{18.066 \times 10^{28} \times \left(1.6 \times 10^{-19}\right)^2 \times 10^{-14}}{9.11 \times 10^{-31}}$$

$$= 5.0823 \times 10^7 \text{ ohm m.}$$

The conductivity of *Al* is 5.0823×10^7 ohm m.

Example 17.5 Use Fermi distribution function to obtain the value of $F(E)$ for $E - E_F = 0.01$ eV at 200 *K*.

Given Data

The difference between Energy level to Fermi level $E - E_F = 0.01$ eV

$$= 0.01 \times 1.6 \times 10^{-19} J$$

$$= 1.6 \times 10^{-21} J$$

Temperature $T = 200$ *K*

Solution

We know that, the Fermi distribution function,

$$F(E) = \frac{1}{1 + e^{(E - E_F)/kT}}$$

Substitution the value of E-E_F and T, we get,

$$= \frac{1}{1+e^{(1.6\times10^{-21})/(1.38\times10^{-23}\times200)}} = \frac{1}{1+e^{0.5797}}$$

$$= 0.3589$$

The Fermi distribution function for energy E is 0.3589.

Example 17.6 Calculate Fermi energy and Fermi temperature in a metal. Fermi velocity of electron in the metal is 0.86×10^6 m s^{-1}.

Given Data

Velocity of electron	$v = 0.86 \times 10^6$ m s^{-1}
Mass of electron	$m = 9.1 \times 10^{-31}$ kg
Electronic charge	$e = 1.6 \times 10^{-19}$ C
Boltzmann's constant	$k = 1.38 \times 10^{-23}$ J K^{-1}

Solution

We know that, the Fermi energy

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

Substituting the values, we get,

$$= \frac{1}{2}\times9.11\times10^{-31}\times0.86\times10^6$$

$$= 3.368 \times 10^{-19} \text{ J}$$

We also know that, the Fermi temperature

$$T_F = \frac{E_F}{k}$$

Substituting the value, we get,

$$= \frac{3.368\times10^{-19}}{1.38\times10^{-23}}$$

$$= 2.43 \times 10^4 \text{ K}$$

Therefore, the Fermi temperature T_F is 2.43×10^4 K.

Example 17.7 Calculate the number of states lying in an energy interval of 0.01 eV above the Fermi level for a crystal of unit volume with Fermi energy $E_F = 3.0$ eV

Given Data

Mass of electron $m = 9.1 \times 10^{-31}$ kg

Energy interval $\Delta E = 0.01$ eV

Planck's constant $h = 6.63 \times 10^{-34}$ Js

Fermi energy $E_F = 3.0$ eV

Solution

Let us take E_1 is ground state,

i.e., $E_1 = 3.0$ eV

$= 3.0 \times 1.6 \times 10^{-19}$

$E_1 = 4.8 \times 10^{-19}$ J

E_2 is given by $E_2 = E_1 + \Delta E$

$= (3.0 + 0.01)$ eV

$= 3.01 \times 1.6 \times 10^{-19}$ J

$E_2 = 4.816 \times 10^{-19}$ J

Number of status per unit volume lying between E_1 and E_2 is given by

$$n = \int_{E_1}^{E_2} \frac{4\pi}{h^3} (2m)^{3/2} E^{1/2}\, dE$$

Here density of status Z (E) is unity

$$= \frac{4\pi}{h^3} (2m)^{3/2} \int_{E_1}^{E_2} E^{1/2} dE$$

Integrating the above equation, we get

$$= \frac{4\pi}{h^3} (2m)^{3/2} \left[\frac{2}{3} (E)^{3/2} \right]_{E_1}^{E_2}$$

Simplify the above equation, we get,

$$= \frac{4\pi}{h_3} (2m)^{3/2} \left[\frac{2}{3} (E)^{3/2} \right]_{E_1}^{E_2}$$

$$= \frac{4\pi}{h^3} (2\mathrm{m})^{3/2} . \frac{2}{3} \left[E_2^{3/2} - E_1^{3/2} \right]$$

Substituting the value we get,

$$= \frac{4\pi \times \left(2 \times 9.1 \times 10^{-31}\right)^{3/2}}{\left(6.63 \times 10^{-34}\right)^3} \times \frac{2}{3} \left[\left[(4.816)^{3/2} - (4.8)^{3/2} \right] \times 10^{-19} \right]^{3/2}$$

$$= 3.74 \times 10^{55} \times (1.108 \times 10^{-30})$$

$$= 4.14 \times 10^{25}$$

The number of states lying between the energy level is 4.14×10^{25}.

Example 17.8 Fermi temperature of a metal is 24600 K. Calculate the Fermi velocity.

Given Data

Fermi temperature of the metal

$$T_F = 24600 \text{ K}$$

Mass of an electron $m = 9.11 \times 10^{-31}$ kg

Solution

We know that, the Fermi velocity is

$$V_F = \sqrt{\frac{2kT_F}{m}}$$

Substituting the values, we get,

$$= \sqrt{\frac{2 \times 1.38 \times 10^{-23} \times 24600}{9.11 \times 10^{-31}}}$$

$$= 0.8633 \times 10^6 \, m \, s^{-1}$$

The Fermi velocity V_F is 0.8633×10^6 m s^{-1}

Example 17.9 Free electron density of aluminum is $18.10 \text{x} 10^{28}$ m^{-3}. Calculate its Fermi energy at 0 K. Planck's constant and mass of free electron are 6.62×10^{-34} J s and 9.1×10^{-31} kg.

Given Data

Electron density of aluminum $n = 18.10 \times 10^{28}$ m^{-3}

Planck's constant $h = 6.62 \times 10^{-34}$ J s

Mass of electron $m = 9.1 \times 10^{-31}$ kg

Solution

We know that, the Fermi energy at 0 K,

$$E_{F0} = \left(\frac{3n_c}{8\pi}\right)^{2/3} \frac{h^2}{2m}$$

Substituting the given values, we get,

$$= \left(\frac{3 \times 18.10 \times 10^{28}}{8 \times 3.14}\right)^{2/3} \times \frac{(6.62 \times 10^{-34})^2}{2 \times 9.1 \times 10^{-31}}$$

$$= 1.8689 \times 10^{-18} \text{ J}$$

$$= \frac{1.8689 \times 10^{-18}}{1.6 \times 10^{-19}}$$

$$= 11.68 \text{ eV}$$

The Fermi energy at 0 K is11.68 eV.

Objectives Type Questions

17.1 Probability function *f(E)* for an electron is

a) $f(E) = \dfrac{1}{1 - \exp\left[\left(E - E_f\right)/KT\right]}$ b) $f(E) = \dfrac{1}{1 + \exp\left[\left(E_f - E\right)/KT\right]}$

c) $f(E) = \dfrac{1}{1 + \exp\left[\left(E - E_f\right)/KT\right]}$ d) $f(E) = \dfrac{1}{1 - \exp\left[\left(E_f - E\right)/KT\right]}$

17.2 Fermi energy level is the _____ energy level in a conduction band at *0 K*.

17.3 Electrons occupy the Fermi energy level fully and leave vacant above E_F at T = 0 K, when

a) $E + E_F$ b) $E = E_F$

c) $E > E_F$ d) $E < E_F$

17.4 Fermi level is _____ at $T = 0$ K, $E < E_F$

17.5 Zero percent probability is the probability for the electrons to occupy the energy level above the Fermi energy level at T = 0 K and

a) $E + E_F$ b) $E = E_F$

c) $E > E_F$ d) $E < E_F$

17.6 Electrons occupy 50% or ½ of the Fermi energy level at _____ and _____

17.7 Density of states is equal to

a) $\dfrac{No.\text{ of quantum states present between } E \text{ \& } E < dE}{\text{Volume of the specimen}}$

b) $\dfrac{No.\text{ of quantum states present between } E \text{ \& } E > dE}{\text{Volume of the specimen}}$

c) $\dfrac{No.\text{ of quantum states present between } E \text{ \& } E - dE}{\text{Volume of the specimen}}$

d) $\dfrac{No.\text{ of quantum states present between } E \text{ \& } E + dE}{\text{Volume of the specimen}}$

17.8 The concentration of free electrons in metal is equal to

a) $\int \frac{4\pi}{h^3}(2m)^{3/2}E^{1/2}dE$ b) $\int \frac{4\pi}{h^3}(2m)^{3/2}E^{-1/2}dE$

c) $\int \frac{4\pi}{h^3}(2m)^{3/2}E^{5/2}dE$ d) $\int \frac{4\pi}{h^3}(2m)^{3/2}E^{3/2}dE$

17.9 The value of Fermi energy level at $T = 0$ K

a) $E_{Fo} = \left(\frac{\hbar^2}{8m}\right)\left(\frac{3n}{\pi}\right)^{2/3}$ b) $E_{Fo} = \left(\frac{h^2}{8m}\right)\left(\frac{3n}{\pi}\right)^{2/3}$

c) $E_{Fo} = \left(\frac{h^2}{8m}\right)\left(\frac{3\pi}{n}\right)^{2/3}$ d) $E_{Fo} = \left(\frac{h^2}{8\pi}\right)\left(\frac{3m}{n}\right)^{2/3}$

17.10 The average energy of the electrons at T = 0 K

a) $E_{ave} = 5/3\ E_{FO}$ b) $E_{ave} = 1/3\ E_{FO}$

c) $E_{ave} = 1/5\ E_{FO}$ d) $E_{ave} = 3/5\ E_{FO}$

17.11 Average Fermi energy level of electrons at non-zero temperature E_{ave} (T) is equal

a) $5/3\ E_{FO}$ b) $1/3\ E_{FO}$

c) $3/5\ E_{FO}$ d) $3/4\ E_{FO}$

KEYS

17.1 c) 17.2 Upper most filled
17.3 d) 17.4 Vacant
17.5 c) 17.6 T = 0 K, $E = E_F$
17.7 d) 17.8 a)
17.9 b) 17.10 d)
17.11 b)

Exercises

17.1 Evaluate the Fermi energy of alkali metals *Li*, *Na* and *K* using the following data:

	Li	*Na*	*K*
Density (kg m^{-3})	534	971	860
Atomic weight	6.939	22.99	39.202

17.2 Determine the electrical conductivity of Cu at 25ºC using the following data.

Density of Cu = 8960 Kg m^{-3}
Atomic weight of Cu = 63.55
Relaxation time of Cu = 10^{-14} s.

The McGraw·Hill Companies

Short Questions

17.1 Write the mathematical expression for the Fermi-Dirac statistics and explain all the terms.
17.2 What is Fermi Level and Fermi energy?
17.3 Explain the distribution of electrons at 0 K based on the quantum theory.
17.4 Draw Fermi distribution curve for 0 K and at any Temperature T K.
17.5 Define the density of states.
17.6 What do you mean by quantization of momentum?
17.7 Show that an electron's momentum in a potential well is quantized.
17.8 Deduce an expression for Fermi energy in terms of concentration of electrons at 0 K.
17.9 Obtain an expression for Fermi energy at T < 0K.
17.10 Explain the significance of Fermi energy in thermo emf measurements.
17.11 Explain the significance of Fermi energy in metal-metal contacts.
17.12 Write the expression for carrier concentration of metal.
17.13 What is meant by effective mass of an electron?
17.14 Explain the electrical conductivity of Cu.
17.15 What is a hole?

Descriptive Questions

17.1 Derive a mathematical expression for density of states.
17.2 Write short notes on the following
(i) Fermi-Dirac distribution
(ii) Fermi energy at $T = 0$ K and T > 0 K.
(iii) Significance of Fermi energy
17.3 What is meant by effective mass of an electron? Derive an expression for the effective mass of an electron.
17.4 What is hole? List out the properties of a hole.

The McGraw·Hill Companies

Chapter

18

Transport Properties of Semiconductors

OBJECTIVES

- To derive the carrier concentration, Fermi level and the conductivity of an intrinsic semiconductor
- To derive the carrier concentration and the Fermi energy level of the *p*-type and *n*-type extrinsic semiconductor
- To explain the Hall effect in a crystal and hence derive the Hall coefficient, Hall voltage, etc., of the crystal
- To discuss the experimental determination of Hall coefficients of a crystal with circuit diagram
- To discuss the important applications of Hall effect
- To derive the variation of carrier concentration, conductivity, Fermi level of the semiconducting materials as a function of temperature

18.1 INTRODUCTION

In metals, the electrons are fully responsible for electrical conduction. But in semiconductors, both electrons and holes are the charge carriers. The electrical conductivity in semiconductor is produced due to the movement of the holes and electrons. In this chapter, let us discuss about the concentration of carriers in intrinsic and extrinsic semiconductors, and Hall effect and its applications.

18.2 CARRIER CONCENTRATION IN AN INTRINSIC SEMICONDUCTOR

In a semiconductor, all the states in a valence band (VB) are occupied while in the conduction band (CB) they are empty at 0 K. Figure 18.1a shows the valence and conduction bands in a semiconductor at 0 K. At room temperature, due to thermal excitation, some electrons are available in the conduction band, as shown in Fig. 18.1b. In a semiconductor, both electrons and holes act as charge carriers. To find the concentration of carriers in an intrinsic semiconductor, first let us consider the conduction band alone.

In a conduction band, some lower energy states are filled with electrons, at room temperature. The concentration of electrons in the conduction band is given by

$$n = \int N(E)\, f(E)\, dE \tag{18.1}$$

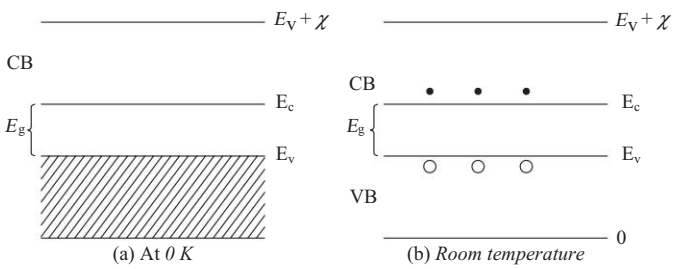


Fig. 18.1 *Energy-band diagram of a semiconductor*

where $N(E)$ is the number of electrons per unit energy per unit volume, and $f(E)$ the probability that a quantum state is occupied by the electrons. $N(E)$ is written as

$$N(E) = \frac{4\pi}{h^3}(2m_e^*)^{3/2} E^{1/2} \tag{18.2}$$

where m_e* is the effective mass of the electron.
The Fermi Dirac distribution function $f(E)$ is written as,

$$f(E) = \frac{1}{1 + \exp\left(\dfrac{E - E_F}{kT}\right)} \tag{18.3}$$

At room temperature, since $E - E_F > kT$, hence Eq. (18.3) can be written as,

$$f(E) = \frac{1}{\exp\left(\dfrac{E - E_F}{kT}\right)} \approx \exp\left[-\left(\frac{E - E_F}{kT}\right)\right] \tag{18.4}$$

From Eqs (18.2) and (18.4), Eq. (18.1) can be written as,

$$n = \int_{E_C}^{E_C + \chi} \frac{4\pi}{h^3}(2m_e^*)^{3/2}(E - E_C)^{1/2} \exp\left[-\left(\frac{E - E_F}{kT}\right)\right] dE \tag{18.5}$$

In Eq. (18.5), $E^{1/2}$ has been replaced by $(E - E_C)^{1/2}$, since the lower energy level in a conduction band is E_C. The above integral varies from E_C to $E_C + \chi$, where E_C is the lower energy level in the conduction band and χ the work function of the metal. Replacing $E_C + \chi$ by ∞, Eq. (18.5) can be written as,

$$n = \int_{E_C}^{\infty} \frac{4\pi}{h^3}(2m_e^*)^{3/2}(E - E_C)^{1/2} \exp\left[-\left(\frac{E - E_F}{kT}\right)\right] dE$$

$$= \frac{4\pi}{h^3}(2m_e^*)^{3/2} \int_{E_C}^{\infty} (E - E_C)^{1/2} \exp\left[-\left(\frac{E - E_F}{kT}\right)\right] dE \qquad (18.6)$$

The integral in Eq. (18.6) is evaluated by substituting $E = E_C + x$, $dE = dx$. The limit varies from 0 to ∞.

Therefore, the term

$$\int_{E_C}^{\infty} (E - E_C)^{1/2} \exp\left[-\left(\frac{E - E_F}{kT}\right)\right] dE$$

becomes

$$\int_0^{\infty} x^{1/2} \exp\left[\left(\frac{E_F - E_C - x}{kT}\right)\right] dx$$

$$= \exp\left(\frac{E_F - E_C}{kT}\right) \int_0^{\infty} x^{1/2} \exp\left[-\frac{x}{kT}\right] dx$$

To evaluate the integral, put $y = x/kT$, $\quad dy = dx/kT$
Therefore, the integral

$$\int_0^{\infty} (x)^{1/2} \exp\left(-\frac{x}{kT}\right) dx$$

becomes

$$(kT)^{3/2} \int_0^{\infty} (y)^{1/2} \exp(-y)\, dy$$

The integral $\int_0^{\infty} y^{n-1} \exp(-y) = \Gamma(n)$ is known as Gamma function. To evaluate this integral, the properties of the Gamma function like $\Gamma(1/2) = \sqrt{\pi}$ and $\Gamma(n+1) = n\Gamma(n)$ are used. The value of the above integral is

$$\int_0^{\infty} x^{1/2} \exp\left(\frac{-x}{kT}\right) dx = (kT)^{3/2} \frac{\pi^{1/2}}{2}$$

Substituting the value of the integral in Eq. (18.6), we get,

$$n = \frac{4\pi}{h^3}(2m_e^*)^{3/2}(kT)^{3/2} \frac{\pi^{1/2}}{2} \exp\left(\frac{E_F - E_C}{kT}\right)$$

$$= 2\left[\frac{2\pi m_e^* kT}{h^2}\right]^{3/2} \exp\left(\frac{E_F - E_C}{kT}\right) \tag{18.7}$$

Equation (18.7), can be written as,

$$n = N_C \exp\left(\frac{E_F - E_C}{kT}\right) \tag{18.8}$$

where $N_C = 2\left[\frac{2\pi m_e^* kT}{h^2}\right]^{3/2}$. Equation (18.8) gives the concentration of electrons in the conduction band.

The concentration of holes in the valence band is calculated by modifying Equation (18.1) as,

$$p = \int_{-\infty}^{\infty} N(E)[1 - f(E)]dE \tag{18.9}$$

where $f(E)$ represents the probability that a quantum state is occupied by the electrons. Therefore, $[1–f(E)]$ represents the probability that a quantum state is occupied by holes. The integral given in Eq. (18.9) varies from 0 to E_v, since 0 is the lowest energy value of the valence band and E_v the highest energy value of the conduction band. The number of electrons per unit energy per unit volume is given by Eq. (18.2) as,

$$N(E) = \frac{4\pi}{h^3}(2m_h^*)^{3/2} E^{1/2}$$

Since the maximum energy of the valence band is E_V, the above equation can be written as,

$$N(E) = \frac{4\pi}{h^3}(2m_h^*)^{3/2} (E_V - E)^{1/2} \tag{18.10}$$

Since $E < E_v$, the value of $[1 - f(E)]$ is obtained from Eq. (18.3):

$$1 - f(E) = 1 - \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)}$$

$$= \frac{1 + \exp\left(\frac{E - E_F}{kT}\right) - 1}{1 + \exp\left(\frac{E - E_F}{kT}\right)}$$

$$= \frac{\exp\left(\frac{E - E_F}{kT}\right)}{1 + \exp\left(\frac{E - E_F}{kT}\right)}$$

Since $E_F - E >> kT$, the exponential term in the denominator is negligible.

The McGraw·Hill Companies

Therefore, we can write

$$1 - f(E) = \frac{\exp\left(\dfrac{E - E_F}{kT}\right)}{1} \approx \exp\left(\frac{-(E_F - E)}{kT}\right) \tag{18.11}$$

From Eqs (18.11) and (18.10), Eq. (18.9) can be written as,

$$P = \int_{-\infty}^{E_V} \frac{4\pi}{h^3}(2m_h^*)^{3/2}(E_V - E)^{1/2} \exp\left(\frac{-(E_F - E)}{kT}\right) dE$$

The above integral varies from 0 to E_V, since 0 and E_V are the lowest and hightest energy values of the valence band, respectively.

$$= \frac{4\pi}{h^3}(2m_h^*)^{3/2} \int_0^{E_V} (E_V - E)^{1/2} \exp\left(\frac{-(E_F - E)}{kT}\right) dE \tag{18.12}$$

The value of the above integral is

$$\int_0^{E_V} (E_V - E)^{1/2} \exp\left(\frac{-(E_F - E)}{kT}\right) dE = (kT)^{3/2} \frac{\pi^{1/2}}{2} \exp\left(\frac{E_V - E_F}{kT}\right) \tag{18.13}$$

From Eqs (18.12) and (18.13), we get,

$$p = N_V \exp\left(\frac{E_V - E_F}{kT}\right) \tag{18.14}$$

where $$N_V = 2\left(\frac{2\pi m_h^* kT}{h^2}\right)^{3/2}$$

Equation (18.14) gives the concentration of holes in the valence band of a semiconductor.

18.2.1 Concentration in an Intrinsic Semiconductor

In an intrinsic semiconductor, the concentration of holes and electrons are equal, i.e., $n = p$. According to the mass action law, the product of the hole and the electron concentrations is equal to the square of the intrinsic concentrations, i.e.,

$$n\,p = n_i^2 \tag{18.15}$$

Substituting the values of n and p from Eqs (18.8) and (18.14) in Eq. (18.15), we get,

$$n_i^2 = N_C N_V \exp\left(\frac{(E_F - E_C)}{kT}\right) \exp\left(\frac{(E_V - E_F)}{kT}\right)$$

$$= N_C N_V \exp\left(\frac{E_V - E_C}{kT}\right)$$

$$= 4\left[\frac{2\pi kT}{h^2}\right]^3 (m_h^* \cdot m_e^*)^{3/2} \exp\left[-\left(\frac{E_C - E_V}{kT}\right)\right]$$

Substituting $E_C - E_V = E_g$ and the values of π, k, h^2 in the above equation, we get, Multiplying and dividing by m^3 in the above equation, we get,

$$n_i^2 = 4\left[\frac{2\pi \times 1.38 \times 10^{-23}}{(6.626 \times 10^{-34})^2}\right]^3 (m_h^* \cdot m_e^*)^{3/2}\, T^3 \exp\left[-\frac{E_g}{kT}\right]$$

$$= 4\left[\frac{2\pi \times 1.38 \times 10^{-23} \times 9.1 \times 10^{-31}}{(6.626 \times 10^{-34})^2}\right]^3 \left(\frac{m_h^* . m_e^*}{m^2}\right)^{3/2} T^3 \exp\left[-\frac{E_g}{kT}\right]$$

$$= 2.322 \times 10^{43} \left(\frac{m_h^* . m_e^*}{m^2}\right)^{3/2} T^3 \exp\left[-\frac{E_g}{kT}\right]$$

$$= AT^3 \exp(-E_g/kT) \qquad (18.16)$$

where $$A = 2.322 \times 10^{43} \left(\frac{m_h \cdot m_p}{m^2}\right)^{3/2}$$

The intrinsic concentration is given by

$$n_i = A^{1/2}\, T^{3/2} \exp\,(-E_g/2kT)$$

i.e., $$n_i = A_0\, T^{3/2} \exp(-E_g/2kT) \qquad (18.17)$$

where $$A_0 = A^{1/2} = 4.819 \times 10^{21} \left(\frac{m_h \cdot m_p}{m^2}\right)^{3/4}$$

Equation (18.17) gives the intrinsic carrier concentration in an intrinsic semiconductor.

18.2.2 Fermi Level in an Intrinsic Semiconductor

In an intrinsic semiconductor, the concentrations of electrons and holes are equal. Therefore, from Eqs (18.8) and (18.14), we get,

$$N_C \exp\left(\frac{E_F - E_C}{kT}\right) = N_V \exp\left(\frac{E_V - E_F}{kT}\right)$$

$$\frac{N_C}{N_V} = \exp\left(\frac{E_V - E_F}{kT}\right) \exp\left(\frac{-(E_F - E_C)}{kT}\right)$$

$$= \frac{E_V + E_C - 2E_F}{kT}$$

$$\ln \frac{N_C}{N_V} = \frac{E_V + E_C - 2E_F}{kT}$$

$$E_V + E_C - 2E_F = kT \ln\,(N_C/N_V)$$

The McGraw·Hill Companies

$$E_F = \frac{E_V + E_C}{2} - \frac{kT}{2} \ln \frac{N_C}{N_V}$$

If the effective masses of the holes and electrons are equal, then

$$E_F = \frac{E_C + E_F}{2} \tag{18.18}$$

Equation (18.18) shows that for an intrinsic semiconductor, the Fermi level lies at the middle of the energy gap, when the effective masses of the holes and electrons are equal.

18.3 CONDUCTIVITY OF SEMICONDUCTORS

The conductivity of a metal is given by,

$$\sigma = ne\mu \tag{18.19}$$

where n is the concentration of electrons and μ is the electron's mobility.

In a semiconductor, both the electrons and holes are the charge carriers. Therefore, Eq. (18.19) is modified as,

$$\sigma = ne\mu_e + pe\mu_h \tag{18.20}$$

where n and p are the concentrations of electrons and holes, respectively, and μ_e and μ_h are, respectively, the mobility of electrons and holes.

For an intrinsic semiconductor, since the concentrations of electrons and holes are equal, Eq. (18.20) can be written as,

$$\sigma = n_i e\,(\mu_e + \mu_h) \tag{18.21}$$

In a p-type semiconductor, since $p >> n$, Eq. (18.20) can be written as

$$\sigma = p\,e\,\mu_h \tag{18.22}$$

In an n-type semiconductor, since $n >> p$, Eq. (18.20) can be written as

$$\sigma = n\,e\,\mu_e \tag{18.23}$$

Equations (18.21) to (18.23) are used to find the conductivity of a semiconductor.

Substituting the value of n_i in Eq. (18.21), we get,

$$\sigma = 2\left(\frac{2\pi mkT}{h^2}\right)^{3/2}\left(\frac{m_e^* m_n^*}{m^2}\right)^{3/4} e^{-E_g/2kT}\, e^{(\mu_e+\mu_h)} \tag{18.24}$$

Equation (18.24) can be written as,

$$\sigma = Ae^{-\mathrm{Eg}/2\mathrm{kT}} \tag{18.25}$$

where

$$A = 2\left(\frac{2\pi mkT}{h^2}\right)^{3/2}\left(\frac{m_e^* m_h^*}{m^2}\right)^{3/4} e^{(\mu_e+\mu_h)}$$

From Eq. (18.25), the resistivity of the intrinsic semiconductor is

$$\frac{1}{\rho} = Ae^{-\mathrm{Eg}/2kT}$$

i.e., $$\rho = \frac{1}{A} e^{E_g/2kT} \tag{18.26}$$

Substituting the value of ($\rho = Ra/l$), in Equation (18.26), we get,

$$\frac{Ra}{l} = \frac{1}{A} e^{E_g/2kT}$$

$$R = \frac{l}{Aa} e^{E_g/2kT}$$

i.e. $$R = Ce^{\mathrm{E_g}/2\mathrm{kT}} \tag{18.27}$$

Here, $C = \frac{l}{Aa}$, where a is the area of cross-section and l the length of the specimen.

In Eq. (18.27), taking ln on both sides, we get,

$$\ln R = \ln C + E_g/2kT \tag{18.28}$$

Equation (18.28) is similar to equation of a straight line $y = mx + c$. In Eq. (18.28), $y = \ln R$, $x = 1/T$, $m = E_g/2k$ and $C = \ln C$. If a plot is drawn between ln R and $1/T$, the value of E_g can be determined from the slope of the straight line.

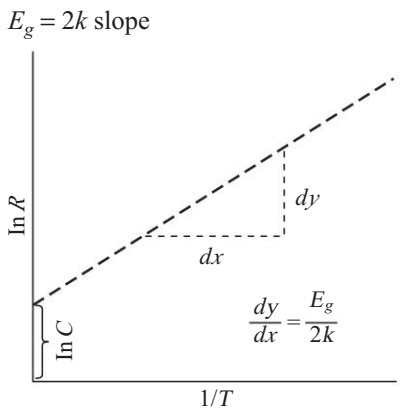


Fig. 18.2 *A plot of ln R versus* $\frac{1}{T}$ *for an intrinsic semiconductor*

Therefore, the energy gap of an intrinsic semiconductor is determined by drawing a plot between ln R and $1/T$.

Experimental determination of the band gap To determine the energy gap of a semiconductor, its resistance is measured at different temperatures using a post office box. Then, a plot is drawn between ln R and 1/T. The slope of the straight line is determined. The energy gap is determined using the relation, $E_g = 2k$ slope.

The McGraw·Hill Companies

18.4 EXTRINSIC SEMICONDUCTOR

Consider an intrinsic Si crystal is doped with pentavalent impurities such as *P*, As, Sb, etc., and the regular Si atom is replaced by these impurity atoms. Since these impurity atoms are having five valence electrons, four electrons occupy the regular lattice site and form covalent bonding. The one excess electron is loosely bound to the parent atom. At room temperature, the thermal excitation energy ($\approx kT$) produced is 0.025 eV. The excess of electron is lightly bound to the parent atom (~0.01 eV). Since, the thermal energy is sufficient to excite this electron, this electron can contribute to the conduction process. Since, the pentavalent impurity donate an electron, it is said to be a donor impurity. According to the energy band theory of solids, these excess electrons possess an energy level just below the conduction band. This energy level is known as the donor level. The donor level in an *n*-type semiconductor and the acceptor level in a *p*-type semiconductor are shown in Fig. 18.3. In Si, the difference between the donor level and the lowest energy value of the conduction band is nearly 0.03 eV. At room temperature, the thermal energy is sufficient to ionize donor impurities (N_D^+).

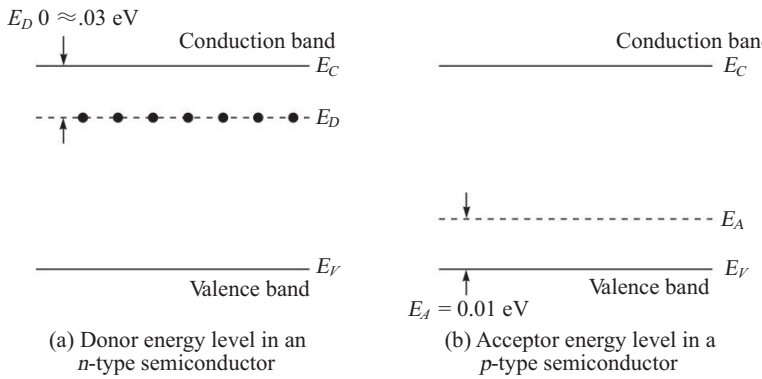


Fig. 18.3 *Donor and acceptor energy levels in an n-type and p-type semiconductors*

Consider an intrinsic semiconductor doped with trivalent impurities such as B, Al, Ga, etc. When a trivalent impurity is doped, it occupies the regular site of a Si atom. All the three electrons are shared by three different Si atoms. A covalent bond is not completed, because the trivalent impurity has only three electrons. Therefore, a hole is created due to the trivalent impurity doping. This hole is free to move within the crystal.

At 0 K, all the states in the valence band is completely occupied. The excess of hole produced due to the trivalent impurity doping occupies an energy level just above the valence band. This energy level is known as the acceptor level. For Ge, the acceptor level lies just 0.01 eV above the valence band and for Si, it lies just 0.01–0.04 eV above the valence band. Thus, a semiconductor has more positive charge when it is doped with acceptor impurities and is known as a *p*-type semiconductor.

18.5 *n*-TYPE SEMICONDUCTOR

A semiconductor doped with donor impurities is known as an *n*-type semiconductor. For example, if a tetravalent silicon atom is doped with pentavalent atoms such as As and Bi, it will become an *n*-type semiconductor.

18.5.1 Fermi Level

In an *n*-type semiconductor, the concentration of electrons in the conduction band is equal to the sum of the number of electrons excited from the valence band and the number of ionized donor impurities excited from the donor energy level. Let N_D and N_D^+ respectively be the concentration of electrons in the donor energy level and the ionized donor impurity concentration. Then, the number of electrons in the conduction band, N_e, is given by

$$N_e = N_h + N_D^+ \tag{18.29}$$

where N_h is the number of holes in the valence band.

The number of ionized donor impurity electron is given by

$$N_D^+ = N_D\,(1 - f(E_D)) \tag{18.30}$$

where $1 - f(E_D)$ is the probability of the absence of electrons in the donor level. The ionized donor impurity concentration is the concentration of the electron absent in the donor level.

The Fermi–Dirac distribution of function for donor energy level is given as,

$$f(E_D) = \frac{1}{1 + \exp\left[\dfrac{E_D - E_F}{kT}\right]}$$

Therefore,

$$1 - f(E_D) = \frac{\exp\left[\dfrac{E_D - E_F}{kT}\right]}{1 + \exp\left[\dfrac{E_D - E_F}{kT}\right]} \tag{18.31}$$

In an *n*-type semiconductor, the Fermi level lies between the donor energy level and the conduction band, hence $E_D - E_F$ is negative. Therefore, Eq. (18.31) can be written as,

$$1 - f(E_D) \approx \exp\left[\frac{E_D - E_F}{kT}\right] \tag{18.32}$$

From Eqs (18.30) and (18.32), the ionized donor impurity concentration can be written as,

$$N_D^+ = N_D \exp\left[\frac{E_D - E_F}{kT}\right] \tag{18.33}$$

At 0 K, all the states in the valence band are completely filled. Therefore, no electron is excited from the valence band to the conduction band. Hence, at $T = 0$ K, $N_h = 0$. Therefore, from Eq. (18.29), the concentration of electron in the conduction band can be written as,

$$N_e = N_D^+ \tag{18.34}$$

From Eq. (18.8), the value of N_e is written as,

$$N_e = N_C \exp\left[\frac{E_F - E_C}{kT}\right]$$

where
$$N_C = 2\left[\frac{2\pi m_c^* kT}{h^2}\right]^{3/2} \tag{18.35}$$

From Eqs (18.33), (18.34) and (18.35), we can write

$$N_C \exp\left[\frac{E_F - E_C}{kT}\right] = N_D \exp\left[\frac{E_D - E_F}{kT}\right]$$

or
$$\exp\left[\frac{E_F - E_C}{kT}\right] \exp\left[\frac{-(E_D - E_F)}{kT}\right] = \frac{N_D}{N_C}$$

or
$$\exp\left[\frac{E_F - E_C - E_D + E_F}{kT}\right] = \frac{N_D}{N_C}$$

In the above equation, taking log on both sides, we get,

$$\frac{2E_F - (E_C + E_D)}{kT} = \ln \frac{N_D}{N_C}$$

i.e.,
$$E_F = \frac{E_C + E_D}{2} + \frac{kT}{2} \ln \frac{N_D}{N_C} \tag{18.36}$$

At $T = 0$ K, Eq. (18.36) becomes

$$E_F = \frac{E_C + E_D}{2} \tag{18.37}$$

Equation (18.37) shows that the Fermi level lies at the middle of the donor level and the lower energy level of the conduction band.

18.5.2 Carrier Concentration

The number of electrons in the conduction band of an intrinsic semiconductor is given by,

$$n = N_C \exp\left[\frac{E_F - E_C}{kT}\right] \tag{18.38}$$

Substituting the value of E_F from Eq. (18.36) in Eq. (18.38), we get,

$$n = N_C \exp\left[\frac{E_C + E_D}{2kT} - \frac{E_C}{kT} + \frac{1}{2} \ln \frac{N_D}{N_C}\right]$$

$$= N_C \exp\left[\frac{E_D - E_C}{2kT}\right] \exp\left[\frac{1}{2} \ln \frac{N_D}{N_C}\right]$$

$$= N_C \exp\left[\frac{E_D - E_C}{2kT}\right] \left(\frac{N_D}{N_C}\right)^{1/2} \tag{18.39}$$

Substituting the values of N_C in Eq. (18.39), we get,

$$n = \left\{ 2\left[\frac{2\pi m_e^* kT}{h^2}\right]^{3/2} N_D \right\}^{1/2} \exp\left[\frac{E_D - E_F}{2kT}\right]$$

$$= (2\,N_D)^{1/2} \left[\frac{2\pi m_e^* kT}{h^2}\right]^{3/4} \exp\left[\frac{E_D - E_F}{2kT}\right] \qquad (18.40)$$

The Eq. (18.40) gives the concentration of electrons in an *n*-type semiconductor.

18.6 *p*-TYPE SEMICONDUCTOR

A semiconductor doped with acceptor impurities is known as a *p*-type semiconductor. For example, if Si is doped with B, Al, Ga, etc., it will become a *p*-type semiconductor.

18.6.1 Fermi Level

In a *p*-type semiconductor, the concentration of holes in the valence band is equal to the sum of the concentration of electrons available in the conduction band and the concentration of electrons available in the acceptor level, i.e., concentration of ionised acceptor impurities

$$p = N_e + N_A^+ \qquad (18.41)$$

At absolute zero temperature, all the states in the valence band are occupied and hence, no electron is available in the conduction band for electrical conduction. Therefore, at $T = 0$ K, $N_e = 0$. Hence, Eq. (18.41) can be written as,

$$p = N_A^+ \qquad (18.42)$$

That is, at 0 K, the hole concentration is equal to the concentration of ionized acceptor impurities. The ionized acceptor impurity is given by

$$N_A^+ = N_A f(E_A) \qquad (18.43)$$

where N_A is the acceptor concentration and $f(E_A)$ is the probability that a quantum state is occupied by an electron in the acceptor level. From the Fermi Dirac distribution function, $f(E_A)$ can be written as,

$$f(E_A) = \frac{1}{1 + \exp\left[\frac{E_A - E_F}{kT}\right]} \qquad (18.44)$$

From Eqs (18.43) and (18.44), we can write

$$N_A^+ = N_A \frac{1}{1 + \exp\left[\frac{E_A - E_F}{kT}\right]} \qquad (18.45)$$

Since, the acceptor level lies above the Fermi level, $E_A - E_F$ is positive. Therefore, the term 1 is neglected, compared to $\exp\left[\frac{E_A - E_F}{kT}\right]$. Therefore, Eq. (18.45) can be written as,

$$N_{As}^{+} = N_A \exp\left[\frac{-(E_A - E_F)}{kT}\right] \tag{18.46}$$

Substituting the value of $p \approx \left\{N_V \exp\left[\frac{E_V - E_F}{kT}\right]\right\}$ from Eq. (18.14), in Eq. (18.42) and hence, substituting the value of N^{+}_{A} in Eq. (18.46) we get,

$$N_V \exp\left[\frac{E_V - E_F}{kT}\right] = N_A \exp\left[\frac{-(E_A - E_F)}{kT}\right]$$

i.e., $$\exp\left[\frac{E_V - E_F + E_A - E_F}{kT}\right] = \frac{N_A}{N_V}$$

In the above equation, taking log on both sides, we get,

$$\frac{-2E_F}{kT} + \frac{E_V + E_A}{kT} = \ln\frac{N_A}{N_V}$$

or
$$E_F = \frac{E_V + E_A}{2} - \frac{kT}{2}\ln\frac{N_A}{N_V} \tag{18.47}$$

Equation (18.47) represents the Fermi level in a *p*-type semiconductor. At $T = 0$ K, Eq. (18.47) becomes

$$E_F = \frac{E_V + E_A}{2} \tag{18.48}$$

i.e., at zero degree Kelvin, the Fermi energy level lies at the middle of the top of the valence band and the acceptor energy level.

18.6.2 Carrier Concentration

The concentration of hole in the valence band in an intrinsic semiconductor is given by

$$p = N_V \exp\left[\frac{E_V - E_F}{kT}\right] \tag{18.49}$$

Substituting Eq. (18.47) in Eq. (18.49), we get,

$$p = N_V \exp\left[\frac{E_V}{kT} - \frac{E_V + E_A}{2kT} + \frac{kT}{2kT}\ln\frac{N_A}{N_V}\right]$$

$$= N_V \exp\left[\frac{E_V - E_A}{2kT}\right]\left(\frac{N_A}{N_V}\right)^{1/2}$$

$$= (N_A N_V)^{1/2} \exp\left[\frac{E_V - E_A}{2kT}\right] \tag{18.50}$$

Substituting the value of N_V in Eq. (18.50), we get,

$$p = (2N_A)^{1/2}\left[\frac{2\pi m_h^* kT}{h^2}\right]^{3/4} \exp\left[\frac{E_V - E_A}{2kT}\right] \quad (18.51)$$

Equation (18.51) represents the concentration of carriers in a *p*-type semiconductor.

18.7 HALL EFFECT

Let I_x be the current flowing through a specimen along the *x*-direction and B_Z be the transverse magnetic field applied along the *z*-direction. An electric field E_y is induced in a direction perpendicular to both the current and the magnetic field. This phenomenon is known as Hall effect.

18.7.1 Illustration

The origin of the Hall effect is easy to understand. This property confirms the particle nature of the charge carriers. Consider a semiconducting specimen, as shown in Fig. 18.4, or a metallic specimen, carrying a current I_x along the *x*-direction and a transverse magnetic field along the *z*-direction; then a force will be developed along the *y*-direction. The direction of force is easily noted from the corkscrew rule or right hand thumb rule. Due to the force, the charge carriers are forced downwards and they accumulate near the bottom face, i.e., face 1. If the specimen is a metal, since it has only one type of charge carriers, i.e., electrons, the electrons are forced down into the bottom surface. Therefore, the bottom surface becomes more negative compared to the upper surface. Hence, a potential difference is developed between the bottom and upper surfaces.

If the specimen is an *n*-type semiconductor, since the electrons are the charge carriers, the electrons are forced down into the bottom surface. Therefore, the upper face becomes positive, whereas the bottom surface becomes negative. In a *p*-type semiconductor, the bottom surface is occupied by holes and it becomes more positive compared to the upper surface. The potential difference between the upper and lower surfaces is known as *Hall voltage*.

Let I_x be the current applied through the specimen along the *x*-direction and B_Z be the magnetic field applied to the specimen along *z*-direction. Let *E* be the electric field intensity due to Hall effect. Then, at equilibrium,

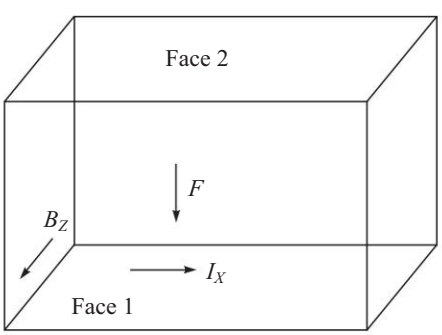


Fig. 18.4 *Hall effect*

The McGraw·Hill Companies

$$eE_H = Bev \tag{18.52}$$

where e is the electronic charge and v the velocity of the charge carriers.

Let V_H be the Hall voltage developed between the faces 1 and 2, then

$$E_H = V_H/d \tag{18.53}$$

From Eqs (18.52) and (18.53)

$$V_H = B\,dv \tag{18.54}$$

where d is the thickness of the specimen.

Let J be the current density and it is given by

$$J = nev \tag{18.55}$$

From Eqs (18.54) and (18.55), we get

$$V_H = Bd \times J/ne \tag{18.56}$$

The current density is also given by, $J = (I/A)$, where A is the area of the specimen. It is given by dw, where w is the width of the specimen. Therefore, $J = (I/dw)$. Substituting this value in Eq. (18.56), we get

$$V_H = \frac{Bd \cdot I}{ne \cdot dw}$$

or

$$V_H = \frac{B\,I}{ne\,w} \tag{18.57}$$

Equation (18.57) can be written by taking, $R_H = 1/ne$ (where R_H is the Hall coefficient) as,

$$V_H = \frac{B\,I \cdot R_H}{w}$$

i.e.,

$$R_H = \frac{V_H\,w}{B\,I} \tag{18.58}$$

Equation (18.58) gives the value of the Hall coefficient. By measuring I, B, V_H and w, the Hall coefficient is determined.

From the Hall coefficient, one can find the value of the concentration of the carriers using the relation, $R_H = 1/ne$. By knowing the concentration of the carriers, the mobility of the charge carriers is determined using the relation,

$$\sigma = n\,e\,\mu$$

or

$$\mu = R_H \sigma \tag{18.59}$$

where σ is the electrical conductivity of the material.

The Hall effect and the Hall mobility of selected metals are given in Table 18.1.

Table 18.1 *Hall effect and hall mobility of selected metals*

Metals	$n(\times 10^{28}\ m^{-3})$	$R_H\ (\times 10^{-11}\ m^3A^{-1}s^{-1})$	$\mu_H = \lvert\sigma R_H\rvert\ (\times 10^{-4}\ m^2v^{-1}s^{-1})$
Ag	5.85	–9.0	57
Al	18.06	–3.5	13
Au	5.90	–7.2	31
Cu	8.45	–5.5	32
Ga	15.3	–6.3	3.6
In	11.49	–2.4	2.9
Mg	8.60	–9.4	22
Na	2.5	–25	53

18.7.2 Experimental Determination of Hall Voltage

A rectangular slab of a semiconducting material of width w and breadth b is taken. A current of I_x ampere is applied to this sample along the x-direction, as shown in Fig. 18.5, by connecting the specimen with a battery B. A magnetic field B_z is applied to this specimen along the z-direction. A voltmeter is connected between the bottom and upper faces of the crystal.

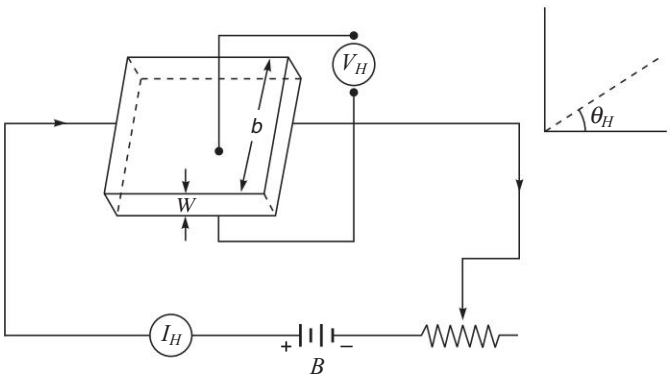


Fig. 18.5 *Experimental measurement of Hall voltage*

The Hall voltage measured is given by Eq. (18.58). Equation (18.58) is derived by assuming that all the particles travel with a mean drift velocity v. Actually, the current carriers have a random thermal distribution of speed. Taking into account this random distribution of speed, the Hall coefficient is redefined as $R_H = \frac{3\pi}{8}\left(\frac{1}{ne}\right) \approx \frac{3\pi}{8\rho}$, where ρ is known as the charge density and is equal to ne. Equation (18.59) can be written as,

$$\mu = \left(\frac{3\pi}{8}\right)\frac{\sigma}{ne} \tag{18.60}$$

Since, the charge of the electron is negative, for an n-type material, Eq. (18.60) can be written as,

$$\mu_e = -\left(\frac{3\pi}{8}\right)\frac{\sigma}{ne} \tag{18.61}$$

The McGraw·Hill Companies

and for hole, since the charge is positive and hence, Eq. (18.60) can be written as,

$$\mu_h = \left(\frac{3\pi}{8}\right)\frac{\sigma}{ne} \tag{18.62}$$

The Hall angle (θ_H) is defined as,

$$\tan\theta_H = E_H/E_x \tag{18.63}$$

where E_H is the Hall voltage measured per unit thickness of the specimen. From Eq. (18.53),

$$E_H = \frac{V_H}{d} = \frac{BI_x}{newd} = \frac{BJ_x}{ne} = \frac{Bnev_x}{ne} = Bv_x \tag{18.64}$$

From Eqs (18.63) and (18.64),

$$\tan\theta_H = \frac{Bv_x}{E_x} \tag{18.65}$$

Thus, by measuring the Hall voltage and the conductivity of the specimen simultaneously, one can determine the mobility, concentration, Hall coefficient and Hall angle.

18.7.3 Applications of the Hall Effect

Some of the applications of the Hall effect are given below.

(i) It is used to find whether the given semiconductor is *n*-type or *p*-type.

(ii) It is used to find the concentration of carriers and hence, the mobility of carriers.

(iii) Equation (18.57) shows that the Hall voltage is the product of two independent variables like the current and the magnetic field. Therefore, this property can be used to carry out the multiplication process and to measure the power dissipation in a load, where the load current and the voltage are multiplied.

(iv) The Hall effect is ideally suited to measure the magnetic field in many applications. Using a typical Hall effect magnetometer, the magnetic field can be measured in the range of 10 μT to 1 T full scale. This value is comparable to the earth's magnetic field of ~50 μT and to that of a strong magnet, typically ~1 T. The manufacturers use different semiconductors depending upon the application and the accuracy needed.

(v) Hall effect semiconducting devices are used as sensors to sense magnetic fields. The linear Hall effect sensor devices from Texas Instruments are TL170–173 and from Sprague or Allegro Microsystems is the UG3500–3600 series. The sensor TL173 is capable of providing a Hall voltage of 15 mV per mT of applied magnetic field.

(vi) Hall effect is also used in magnetically activated electronic switches. Hall effect electronic switches are used as non-contacting keyboards and panel switches. The applications of Hall effect switches range from ignition systems to speed controls, position detectors, alignment controls, brushless dc motor commutators, etc.

18.8 VARIATION OF ELECTRICAL CONDUCTIVITY WITH TEMPERATURE

The concentration of charge carriers in *n*-type and *p*-type semiconductors are derived by assuming the temperature as room temperature, $n \approx N_D$ and $p \approx N_A$. If the temperature is raised, the conductivity of the extrinsic semiconducting materials vary with two factors: (i) variation of carrier concentration with, temperature, and (ii) variation of drift mobility with temperature.

18.8.1 Variation of Carrier Concentration with Temperature

Consider an *n*-type semiconductor with N_d donors per unit volume. Let $N_d >> n_i$, where n_i is the intrinsic carrier concentration. At very low temperatures, all the valence bands are completely occupied by the electrons and the donor carriers. Therefore, no electron is available in the conduction band for the electrical conduction. This is depicted in Fig. 18.6a.

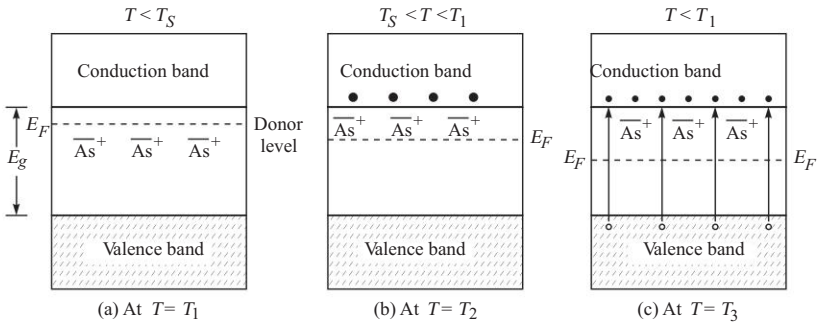


Fig. 18.6 *Concentration of electrons at different temperatures: (a) At a very low temperature, (b) At a temperature, which is sufficient to excite donor electron, and (c) At a temperature, which is sufficient to excite valence electrons*

If the temperature is increased gradually, some of the donor electrons are ionized and their electrons are excited to the conduction band, as shown in Fig. 18.6b. At room temperature, all the donor electrons are excited to the conduction band. If the temperature is further increased, there is no increase in concentration of electrons in the conduction band, since the electrons in the conduction band require a minimum energy of $\Delta E = E_c - E_v$, which is greater than the ionization energy $[\Delta E_i = E_c - E_d\ (<< E_g)]$ of the donor electrons. The electron concentration at very low temperature is given by

$$n = \left(\frac{1}{2} N_C\ N_d\right)^{1/2} \exp\left(-\frac{\Delta E}{2kT}\right) \tag{18.66}$$

Equation (18.66) is similar to the intrinsic concentration of carriers,

$$n_i = (N_C N_V)^{1/2}\ \exp\left(-\frac{E_g}{2kT}\right) \tag{18.67}$$

In Eq. (18.66), a factor ½ is introduced because the donor occupation statistics is different from the usual Fermi Dirac distribution function. Equation (18.66) is applicable only to find the donor concentration excited from the donor level to the conduction band at very low temperature. But Eq. (18.67) is used to find the concentration of carriers excited from the valence band to the conduction band.

The McGraw·Hill Companies

If the temperature is further increased, the electrons present in the valence band acquire energy and hence, breaks the bond. The electrons dislodged from the covalent bond are excited to the conduction band due to thermal agitation.

The dependence of the electron concentration on temperature is classified into the three categories given below:

(i) Low-temperature range
(ii) Medium temperature range, and
(iii) High-temperature range

Let us discuss the above three points briefly.

(i) Low-Temperature Range ($T < T_s$) At very low temperatures, all the donor levels and the valence bands are occupied by electrons. No electron is available in the conduction band. If the temperature is slightly raised, the donor electrons are ionized and they are excited to the conduction band. The ionization of donor electrons continues until the temperature Ts, known as the saturation temperature, is reached and hence, all the donor electrons are ionised. The concentration of electrons in this range is given by Eq. (18.66) and is known as ionization range.

(ii) Medium-Temperature Range ($T_s < T < T_i$) If the temperature is further increased beyond the saturation temperature, there is no increase in concentration of electrons in the conduction band, until a temperature required to liberate the electrons in the valence band is reached. Since the electrons in the valence band requires an energy of $\Delta E = E_C - E_V$, which is very high compared to the donor ionization energy $\Delta E_i = E_C - E_D$, this temperature range is said to be the extrinsic range.

(iii) High-Temperature Range ($T > T_i$) When the temperature becomes sufficient to liberate electrons from the valence band, the concentration of electrons in the conduction band increases very sharply. The reason is, at this temperature, more and more number of electrons get liberated and they are excited to the conduction band. This region is said to be the intrinsic range.

If a plot is drawn between ln (n) and $1/T$, a curve as shown in Fig. 18.7 is obtained. The three regions can be distinctly seen in Fig. 18.7. Since there is a sharp increase in the concentration of electrons at high temperature, concentration is given as,

$$n \propto T^{3/2} \exp\left(\frac{-E_g}{2kT}\right) \tag{18.68}$$

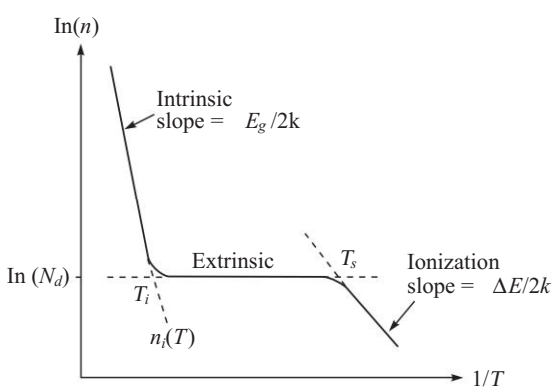


Fig. 18.7 *A plot of ln (n) versus* $\frac{1}{T}$ *for n-type semiconductor*

The McGraw·Hill Companies

This shows that there is a domination of the exponential part over $T^{3/2}$ in the high temperature region. In the extrinsic region, $n \approx N_D$ and is practically independent of temperature. The temperature dependence of the intrinsic concentration of Si, Ge and GaAs are shown in Fig. 18.8. From the slope of this curve, one can determine the band gap of the semiconducting materials.

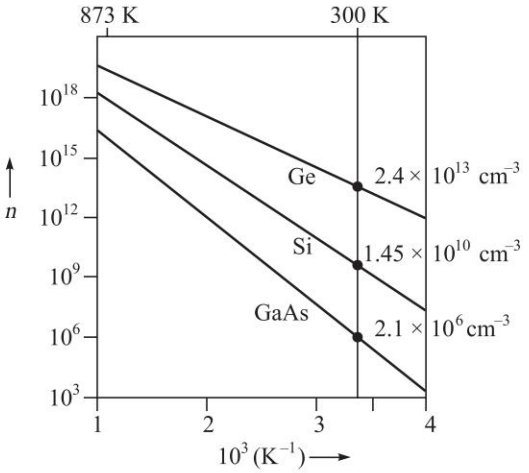


Fig. 18.8 *Temperature dependence of intrinsic concentration of Si, Ge and GaAs*

18.8.2 Variation of Drift Mobility with Temperature

The drift mobility varies with two distinctly different temperature variations. At high temperature region, the drift mobility is limited by scattering from lattice vibrations. At high temperature, $\mu \alpha T^{-3/2}$, i.e., the mobility decreases with the increase in temperature. At low temperature, the scattering of electrons by ionized impurities is the major mobility limiting mechanism and $\mu \propto T^{3/2}$.

The mobility is given by the expression,

$$\mu = \frac{e\tau_r}{m_e^*} \tag{18.69}$$

where e is the charge of the electron, τ_r the mean free time, and m_e^* the effective mass. The mean free time is given by

$$\tau_r = \frac{1}{Sv_{th}N_s} \tag{18.70}$$

where S is the surface area of the scatterer, v_{th} the mean speed of the electrons and N_s the number of scatterer per unit volume. The surface area of the scatterer is πr^2. As the temperature increases, the amplitude of the lattice vibration also increases. Therefore, $r^2 \propto T$ behaviour is observed.

The mean speed of the electron is obtained using the relation,

$$\frac{1}{2} m v_{th}^2 = \frac{3}{2} k_B T$$

i.e.,

$$v_{th} \propto T^{1/2} \tag{18.71}$$

The McGraw·Hill Companies

Thus, the mean free time is given as,

$$\tau_r = \frac{1}{(\pi r^2) v_{\text{th}} N_s} \propto \frac{1}{T(T^{\frac{1}{2}})} \propto \frac{1}{T^{\frac{3}{2}}}$$

i.e., $$\tau \propto T^{-3/2} \tag{18.72}$$

i.e., at high temperature, the drift mobility varies with $T^{-3/2}$.

At low temperature, the ionized impurity scattering is dominant over thermal scattering. Consider that at a low temperature, an electron is scattered by an ionized impurity, such as As^+. As an electron is scattered by an ionized impurity, the magnitude of the kinetic energy of the electron is less than the potential energy.
i.e., $KE < |PE|$. Therefore, the PE is given by

$$PE = \frac{e^2}{4\pi\varepsilon_0\varepsilon_r r_c}$$

The critical radius corresponding to the electron, when it is just scattered, is r_c which is when $KE \approx |PE(r_c)|$.

Therefore, $$\frac{3}{2} k_B T = \frac{e^2}{4\pi\varepsilon r_c}$$

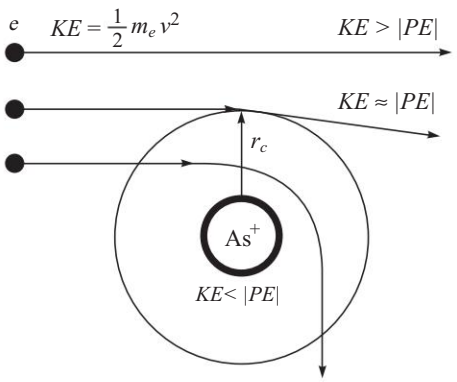


Fig. 18.9 *Scattering of an electron by an ionized impurity*

i.e., $$r_c = \frac{e^2}{6\pi k_B T \varepsilon}$$

The surface area, $$S = \pi r^2 = \frac{\pi e^4}{(6 k_B T \varepsilon \pi)^2}$$

i.e., $$S \propto \frac{1}{T^2} \tag{18.73}$$

From Eqs (18.70), (18.71) and (18.73), we get,

$$\tau_1 \propto \frac{1}{T^{-2}(T^{1/2})\, N_{\text{I}}} \tag{18.74}$$

where N_{I} is the concentration of ionized impurities. The ionized impurity scattering limited mobility from Eq. (18.69) can be written as,

$$\tau_1 \propto \frac{T^{3/2}}{N_1} \tag{18.74}$$

Equation (18.74) shows that the drift mobility due to impurity scattering is directly proportional to $T^{3/2}$. The combined drift mobility using Mathiessen's rule can be written as,

$$\frac{1}{\mu_d} = \frac{1}{\mu_I} + \frac{1}{\mu_L} \tag{18.75}$$

where μ_d , μ_L and μ_l are the drift mobility, the mobility due to lattice scattering and the mobility due to impurity scattering, respectively. The variations of drift mobility with temperature and doping concentration are shown in Fig. 18.10 and Fig. 18.11, respectively.

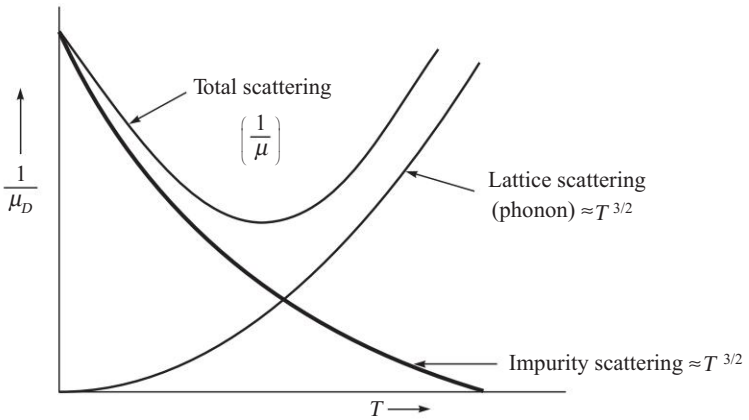


Fig. 18.10 *A plot of* $\left(\frac{1}{\mu_d}\right)$ *versus temperature (T)*

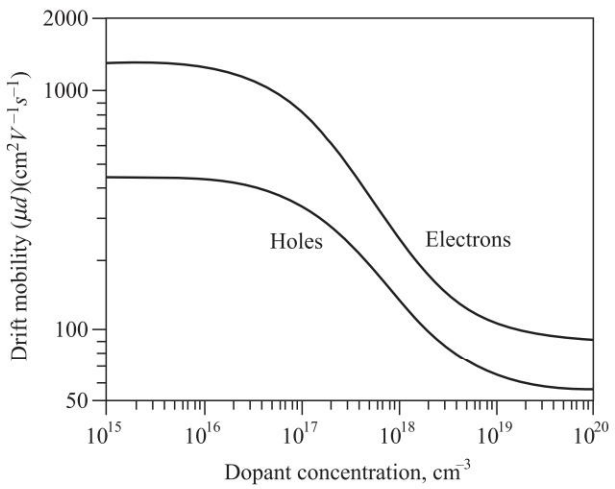


Fig. 18.11 *Variation of drift mobility with doping concentration in Si for electrons and holes*

The McGraw·Hill Companies

18.8.3 Variation of Conductivity with Temperature

The variation of electrical conductivity with temperature has three distinct regions namely: (i) ionization range, (ii) extrinsic range and (iii) intrinsic range. The variation of ln (σ) versus ($1/T$) is shown in Fig. 18.12.

In the ionization range, the concentration of carriers is given by

$$n = \left(\frac{1}{2} N_C N_D\right)^{\frac{1}{2}} \exp\left[\frac{-(E_C - E_D)}{2kT}\right] \tag{18.76}$$

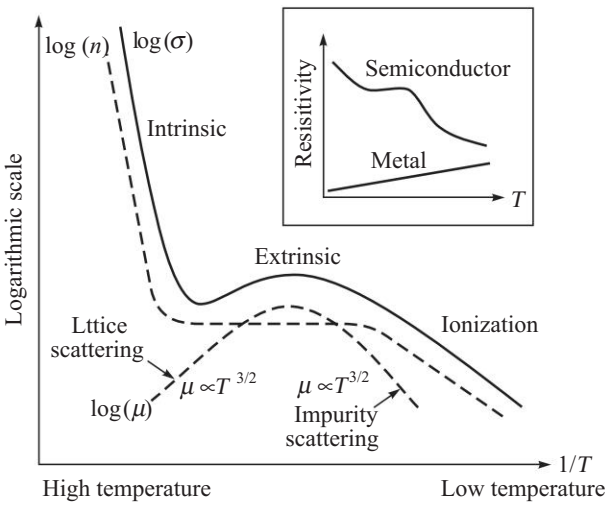


Fig. 18.12 *Variation of conductivity, concentration and mobility with temperature*

and the electrical conductivity is also proportional to the concentration of carriers; it dominates the temperature dependence of conductivity.

In the intrinsic region, the conductivity is given by

$$\sigma_I = n_i e(\mu_e + \mu_h)$$

In the above equation, the intrinsic concentration varies with temperature and hence, the conductivity also varies with temperature. In the extrinsic region, the concentration of carriers is $n \approx N_d$ for n-type material and the conductivity varies with the temperature dependence of the drift mobility. The temperature dependence of conductivity, concentration and mobility are shown in Fig. 18.12.

18.9 VARIATION OF FERMI LEVEL WITH TEMPERATURE IN EXTRINSIC SEMICONDUCTOR

18.9.1 *n*-type Semiconductor

The Fermi level for an n-type semiconductor is given by

$$E_f = \left(\frac{E_C + E_D}{2}\right) + \frac{kT}{2} \ln\left[\frac{N_D}{N_C}\right] \tag{18.77}$$

Substituting $T = 0$ K in the above equation, we get, $E_f = \left(\frac{E_C + E_D}{2}\right)$. This shows that in an *n*-type semiconductor, the Fermi level lies in between the conduction band and the donor level at zero degree Kelvin. As the temperature increases from 0 K, the Fermi level falls. At higher temperature, it falls below the donor level and hence, it approaches the centre of the forbidden gap. The Fermi level lying at the middle of the energy gap at very high temperature shows that the material is behaving as an intrinsic semiconductor. As the donor concentration is increased, the Fermi level will move up. The variation of Fermi level with temperature is shown in Fig. 18.13.

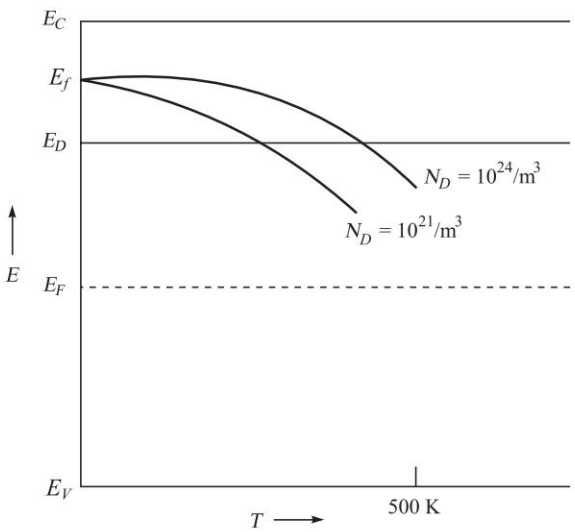


Fig. 18.13 *Variation of Fermi level in an n-type semiconductor with temperature*

18.9.2 *p*-type Semiconductor

The Fermi level in a *p*-type semiconductor is given by

$$E_f = \left(\frac{E_A + E_V}{2}\right) + \frac{kT}{2} \ln\left[\frac{N_V}{N_A}\right] \tag{18.78}$$

The variation of Fermi level with temperature in a *p*-type semiconductor is shown in Fig. 18.14. As $T = 0$ K, the above equation reduces to $E_F = \left(\frac{E_A + E_V}{2}\right)$ i.e., the Fermi level lies at the middle of the acceptor level and the valence band at 0 K. If the temperature is increased, the Fermi level rises. If the temperature is sufficiently high, the Fermi level increases above the acceptor level. At very high temperature, it lies at middle of the energy gap. This shows that at very high temperatures, the *p*-type material behaves as an intrinsic material. As the acceptor concentration is increased, the Fermi level shifts downwards as shown in Fig. 18.14.

The McGraw·Hill Companies

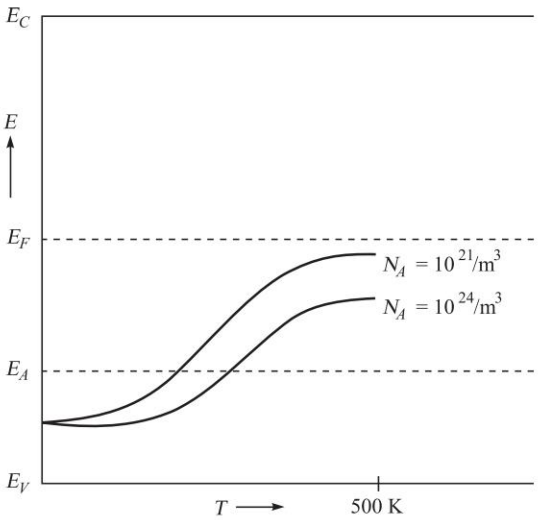


Fig. 18.14 *Variation of Fermi level with temperature in a p-type material*

Keypoints to Remember

- In semiconductors, electrons and holes are the charge carriers for conductors.
- In semiconductors, all the states in the valence band are occupied while in the conduction band, they are empty at 0 K.
- Fermi distribution function

$$f(E) = \frac{1}{1 - \exp\left(\frac{E_F - E_C}{RT}\right)}$$

- The concentration of electrons in the conduction band of a semiconductor is equal to

$$n = N_c \exp\left(\frac{E_F - E_C}{RT}\right)$$

- The concentration of holes in the valence band of a semiconductor is equal to

$$p = N_v \exp\left(\frac{E_V - E_F}{RT}\right)$$

- The concentration of intrinsic carrier concentration in an intrinsic semiconductor is $n_i = A_o T^{3/2} \exp\left(-\frac{E_g}{2KT}\right)$

The McGraw·Hill Companies

- In an intrinsic semiconductor, when the effective mass of holes and electrons are equal, then Fermi energy level lies at the middle of energy gap,

$$E_F = \frac{E_C + E_V}{2}$$

- A semiconductor doped with acceptor impurities has more positive charges and is known as a *p*- type semiconductor.
- A semiconductor doped with donor impurities is known as an *n*-type semiconductor.
- In an *n*-type semiconductor, the Fermi energy level lies at the middle of the donor level and the lower energy level of the conduction band.

$$\text{i.e., } E_F = \frac{E_C + E_D}{2}$$

- In a *p*-type semiconductor, the Fermi energy level lies at the middle of the top of the valence band and the acceptor energy level

$$\text{i.e., } E_F = \frac{E_V + E_A}{2}$$

- When a transverse magnetic field is applied to a current carrying conductor in the *z*-direction, an electric field is induced in a direction perpendicular to both the current and magnetic field. This phenomenon is known as Hall effect.
- The Hall voltage developed on the specimen is equal to $V_H = E_H \times d$, where E_H is the electric field intensity due to Hall effect and d, the thickness of specimen.
- According to Mathiessen, the combined drift mobility is equal to $\frac{1}{\mu_d} = \frac{1}{\mu_l} + \frac{1}{\mu_L}$, where μ_l and μ_L are the mobility due to impurity scattering and lattice scattering.

Solved Problems

Example 18.1 Calculate the intrinsic carrier concentration, intrinsic conductivity and resistivity of Ge at 300 K using the following data: $\mu_e = 0.4\ m^2\ V^{-1}\ s^{-1}$, $\mu_h = 0.2\ m^2\ V^{-1}\ s^{-1}$, $E_g = 0.7$ eV, $m_e^* = 0.55\ m_0$, and $m_n^* = 0.37\ m_0$.

Given Data

Temperature (T) = 300 K

Electron mobility (μ_e) = $0.4\ m^2\ V^{-1}\ s^{-1}$

Hole mobility (μ_h) = $0.2\ m^2\ V^{-1}\ s^{-1}$

The McGraw·Hill Companies

Band gap (E_g) $= 0.7$ eV

Electron effective mass (m_e*) $= 0.55\ m_0$

Hole effective mass (m_h*) $= 0.37\ m_0$

Solution The intrinsic concentration is given by

$$n_i = 2\left(\frac{2\pi kT}{h^2}\right)^{3/2} (m_h^* m_e^*)^{3/4} \exp\left[\frac{-E_g}{2kT}\right]$$

$$= 2\left(\frac{2\pi \times 1.38 \times 10^{-23} \times 300}{(6.626 \times 10^{-34})^2}\right)^{3/2} (0.55 \times 0.37)^{3/4} \times (9.1 \times 10^{-31})^{3/2}$$

$$\exp\left[\frac{-0.7 \times 1.6 \times 10^{-19}}{1.38 \times 10^{-23} \times 300}\right]$$

$$= 1.352 \times 10^{13}\ \text{m}^{-3}$$

The intrinsic conductivity $(\sigma) = n_i\, e\, (\mu_e + \mu_n)$

$= 1.298 \times 10^{-6}$

The intrinsic resistivity $(\rho) = 1/\sigma$

$= 0.77 \times 10^{6}$

The intrinsic concentration $(n_i) = 1.352 \times 10^{13}\ \text{m}^{-3}$

The intrinsic conductivity $(\sigma) = 1.298 \times 10^{-6}\ \Omega^{-1}\ \text{m}^{-1}$

The intrinsic resistivity $(\rho) = 0.77 \times 10^{6}\ \Omega\ \text{m}$

Example 18.2 A *n*-type Si wafer has been doped uniformly with antimony (Sb) atoms and the doped Si has the donor concentration of 10^{16} per cm^3. Calculate the Fermi energy with respect to the Fermi energy in intrinsic Si (For Si, $n_i = 1.45 \times 10^{10}\ \text{cm}^{-3}$).

Solution Doping of Sb converts Si into *n*-type material with $N_d = 10^{16}\ \text{cm}^{-3}$.

For Si, $n_i = 1.45 \times 10^{10}\ \text{cm}^{-3}$

$$N_d > n_i$$

Therefore, we have $N_d = 10^{16}\ \text{cm}^{-3}$

For intrinsic Si,

$$n_i = N_c \exp\left[\frac{-(E_C - E_{Fi})}{kT}\right] \quad \text{(i)}$$

For doped Si,

$$N_d = N_c \exp\left[\frac{-(E_C - E_{Fd})}{kT}\right] \quad \text{(ii)}$$

Dividing Eq. (i) by Eq. (ii), we get,

$$\frac{N_d}{n_i} = \exp\left[\frac{(E_{Fd} - E_{Fi})}{kT}\right]$$

$$E_{Fd} - E_{Fi} = kT \ln\left(\frac{N_d}{n_i}\right) = 0.348 \text{ eV}$$

The Fermi energy with respect to E_F in intrinsic Si = 0.348 eV.

Example 18.3 Find the resistance of an intrinsic germanium rod which is 1 cm long, 1 mm wide and 1 mm thick at 300 K. For germanium $n_i = 2.5 \times 10^{19}$ /m^3, $\mu_e = 0.39$ m^2 V^{-1} s^{-1}, $\mu_h = 0.19$ m^2 V^{-1} s^{-1} at 300 K.

Given Data

Intrinsic carrier concentration	$n_i = 2.5 \times 10^{19}$ /m^3
Electron mobility,	$\mu_e = 0.39$ m^2 V^{-1} s^{-1}
Hole mobility	$\mu_h = 0.19$ m^2 V^{-1} s^{-1}
length of the rod	$l = 1 \times 10^{-2}$ m

Solution

We know that the electrical conductivity of an intrinsic semiconductor (Germanium) is

$$\sigma = n_i e (\mu_e + \mu_h)$$

Substituting the values, we get,

$$= 2.5\times10^{19} \times 1.6\times10^{-19} \times (0.39+0.19)$$

$$= 2.32$$

Therefore, the resistance,

$$R = \frac{l}{\sigma A}$$

where A is area of cross-section and given by product of width and thickness.

Substituting the given values, we get,

$$R = \frac{1\times10^{-2}}{2.32\times(1\times10^{-3}\times1\times10^{-3})} = 4310\ \Omega$$

The resistance of germanium is 4310 Ω.

Example 18.4 The intrinsic carrier density is 1.5×10^{16} / m^{-3}. If the electron and hole mobilities are 0.13 and 0.05 m^2 V^{-1} s^{-1}, calculate the electrical conductivity.

Given Data

Carrier concentration,	$n_i = 1.5 \times 10^{16}$ m^{-3}
Electron mobility,	$\mu_e = 0.13$ m^2 V^{-1} s^{-1}
Hole mobility,	$\mu_h = 0.05$ m^2 V^{-1} s^{-1}

The McGraw·Hill Companies

Solution

We know that the electrical conductivity is

$$\sigma = n_i e[\mu_e + \mu_h]$$

Substituting the values, we get,

$$= 1.5 \times 10^{16} \times 1.6 \times 10^{-19} [0.13 + 0.05]$$
$$= 4.32 \times 10^{-4}\ \Omega^{-1}\ \text{m}^{-1}$$

The electrical conductivity of the intrinsic semiconductor is $4.32 \times 10^{-4}\ \Omega^{-1}\ \text{m}^{-1}$.

Example 18.5 Find the intrinsic resistivity of Ge at room temperature of 300 K if the carrier density is $2.15 \times 10^{13}/\text{cm}^3$

Given data

Mobility of electron $\mu_e = 3900\, cm^2\, V^{-1}\, s^{-1}$

Mobility of hole $\mu_n = 1900\, cm^2$

Carrier density $n_i = 2.15 \times 10^{13}\ \text{cm}^{-3}$

Solution

We know that the conductivity of Ge is

$$\sigma_i = e(\mu_e + \mu_h) n_i$$

Substituting the values, we get,

$$= 1.6 \times 10^{-19} \times (3900 + 1900) \times 2.15 \times 10^{13}$$
$$= 1.6 \times 10^{-19} \times 5800 \times 2.15 \times 10^{13}$$
$$= 2.32 \times 10^{-2}\ \text{ohm}^{-1}\ \text{cm}^{-1}$$

We also know that the intrinsic resistivity

$$\rho_i = \frac{1}{\sigma_i}$$

Substituting the values, we get,

$$= \frac{1}{2.32 \times 10^{-2}}$$
$$\rho_i = 43\ \Omega\ \text{cm}$$

Example 18.6 In a *p*-type germanium $n_i = 2.1 \times 10^{19}\ \text{m}^{-3}$ and density of boron is 4.5×10^{23} atoms m^{-3}. The electron and hole mobilities are 0.4 and 0.2 $\text{m}^2\text{V}^{-1}\text{s}^{-1}$ respectively. What is its electrical conductivity before and after the addition of boron atoms?

Given data

Carrier concentration $n_i = 2.1 \times 10^{19}\ \text{m}^{-3}$

Electron mobility $\mu_e = 0.4 m^2 V^{-1} s^{-1}$

Hole mobility $\mu_h = 0.2\, m^2\, V^{-1}\, s^{-1}$

Density of boron is equal to density of hole $p = 4.5 \times 10^{23}\ \text{m}^{-3}$

Solution

We know that the conductivity of the Ge is

$$\sigma = n_1 e(\mu_e + \mu_h)$$

Substituting the values, we get,

$$= 2.1\times10^{19} \times 1.6\times10^{-19}(0.4+0.2)$$

$$= 2.016\Omega^{-1}m^{-1}$$

After doping with boron, the electrical conductivity due to extrinsic (p-type) semiconductor is

$$\sigma_\gamma = p\, e\, \mu_p$$

Substituting the values, we get,

$$= 4.5 \times 10^{23} \times 1.6 \times 10^{-19} \times 0.2$$

$$= 1.44\times10^{4}\,\Omega^{-1}m^{-1}$$

Therefore, the electrical conductivity of boron doped semiconductor is $1.44 \times 10^4\ \Omega^{-1}\ m^{-1}$.

Example 18.7 Find the resistance of a 1 cm^3 pure Si crystal. What is the resistance when the crystal is doped with arsenic with a doping concentration of 1 in 10^9 (Given: $n = 5 \times 10^{28}\ \text{m}^{-3}$, $n_i = 1.45 \times 10^{13}\ \text{m}^{-3}$, $\mu_e = 1.35\ \text{m}^2\ \text{V}^{-1}\ \text{s}^{-1}$, $\mu_h = 0.45\ \text{m}^2\ \text{V}^{-1}\ \text{s}^{-1}$).

Given Data

Atomic concentration in Si $= 5 \times 10^{28}\ \text{m}^{-3}$

Intrinsic concentration (n_i) $= 1.45 \times 10^{13}\ \text{m}^{-3}$

Electron mobility (μ_e) $= 1.35\ \text{m}^2\ \text{V}^{-1}\ \text{s}^{-1}$

Hole mobility (μ_h) $= 0.45\ \text{m}^2\ \text{V}^{-1}\ \text{s}^{-1}$

Solution The intrinsic conductivity is

$$\sigma_i = n_i\, e\, (\mu_e + \mu_h)$$

$$= 4.18 \times 10^{-6}\ \Omega^{-1}\ \text{m}^{-1}$$

The resistivity is $\rho = 1/\sigma = (RA/L) = 2.39 \times 10^{-5}\ \Omega\ \text{m}$

Therefore, the resistance is

$$R = L/\sigma A = 2.39 \times 10^7\ \Omega$$

The McGraw·Hill Companies

If the crystal is doped with 1 in 10^9 arsenic, then the donor concentration is given by

$$N_d = \frac{N_{si}}{10^9} = \frac{5\times10^{28}}{10^9} = 5 \times 10^{19}\ \text{m}^{-3}$$

The concentration of hole is

$$p = \frac{N_i^2}{N_d} = \frac{\left(1.45\times10^{16}\right)^2}{5\times10^{19}} = 4.2 \times 10^6\ \text{m}^{-3}$$

Therefore, the conductivity is $\sigma = ne\mu_e = N_d\, e\, \mu_e$

$$= 10.8\ \Omega^{-1}\ \text{m}^{-1}$$

The resistance is $R = L/\sigma A = 9.26\ \Omega$

Example 18.8 Mobilities of electrons and holes in a sample of intrinsic Ge at 300 K are 0.36 $\text{m}^2\ \text{V}^{-1}\ \text{s}^{-1}$ and 0.17 $\text{m}^2\ \text{V}^{-1}\ \text{s}^{-1}$, respectively. If the resistivity of the specimen is 2.12 Ω m, compute the forbidden energy gap for Ge, $m_e^* = 0.5\ m_0$ and $m_h^* = 0.37\ m_0$.

Given Data

Temperature (T)	= 300 K
Resistivity (ρ)	= 2.12 Ω m
Electron mobility (μ_e)	= 0.36 $\text{m}^2\ \text{V}^{-1}\ \text{s}^{-1}$
Hole mobility (μ_h)	= 0.17 $\text{m}^2\ \text{V}^{-1}\ \text{s}^{-1}$

Solution The conductivity is given by,

$$\sigma = 1/\rho = \frac{1}{2.12} = 0.471698\ \Omega^{-1}\ \text{m}^{-1}$$

The conductivity is also given by,

$$\sigma = n_i\, e\, (\mu_e + \mu_h)$$

$$n_i = \frac{\sigma}{e\left(\mu_e + \mu_h\right)} = \frac{0.471698}{1.6\times10^{-19}(0.36+0.17)}$$

$$= 5.56247 \times 10^{18}$$

The intrinsic concentration of carriers is given by

$$n_i = (N_C\, N_V)^{1/2} \exp\left(\frac{-E_g}{2k_BT}\right)$$

where

$$N_C = 2\left(\frac{2\pi k_BT}{h^2}\right)^{3/2} \left(m_e^*\right)^{3/2}$$

$$= 2\left(\frac{2\pi\times1.38\times10^{-23}\times300}{\left(6.626\times10^{-34}\right)^2}\right)^{3/2} \left(0.5\times9.1\times10^{-31}\right)^{3/2}$$

$$= 8.852 \times 10^{24}$$

$$N_V = 2\left(\frac{2\pi k_B T}{h^2}\right)^{3/2}\left(m_h^*\right)^{3/2}$$

$$= 2\left(\frac{2\pi \times 1.38\times10^{-23}\times 300}{\left(6.626\times10^{-23}\right)^2}\right)^{3/2}\left(0.37\times9.1\times10^{-31}\right)^{3/2}$$

$$= 5.635 \times 10^{24}$$

The energy gap of an intrinsic semiconductor is given by

$$E_g = 2kT \ \text{In}\left[\frac{\left(N_C N_V\right)^{1/2}}{n_i}\right]$$

$$= 2kT \times 14.054$$

$$E_g = 0.727 \text{ eV}$$

The band gap of Ge is 0.7 eV.

Example 18.9 For an intrinsic semiconductor with gap width $E_g = 0.7$ eV, calculate the concentration of intrinsic charge carriers at 300 K assuming that $m = m_e^* = m_0$ (rest mass of electron).

Given Data

Reset mass of electron, $m_0 = m^* = 9.1 \times 10^{-31}$ kg

Planck's constant $h = 6.62 \times 10^{-34}$ Js

Boltzmann's constant, $k = 1.38 \times 10^{-23}$ JK^{-1}

Band-gap energy $E_g = 0.7$ eV

Electronic charge $e = 1.6 \times 10^{-19}$ C

Temperature, $T = 300$ K

Solution

We know that the intrinsic carrier concentration

$$n_i = 2 \times \left(\frac{2\pi m_e^* kT}{h^2}\right)^{3/2} e^{-E_g/2kT}$$

Substituting the values, we get,

$$= \left(\frac{2\times3.14\times9.1\times10^{-31}\times1.38\times10^{-23}\times300}{(6.626\times10^{-34})^2}\right)^{3/2} \exp\left[\frac{-E_g}{KT}\right]$$

$$= 2.5087 \times 10^{25} \ \exp\left[\frac{-E_g}{KT}\right]$$

The McGraw·Hill Companies

$$= 2.5087 \times 10^{25} \exp\left[\frac{-0.7 \times 1.6 \times 10^{-19}}{2 \times 1.38 \times 10^{-23} \times 300}\right]$$

$$= 2.5087 \times 10^{25} \times 1.335 \times 10^{-6}$$

$$= 33.49 \times 10^{18}$$

Therefore, the carrier concentration of an intrinsic semiconductor is $= 33.49 \times 10^{18}$ m^{-3}.

Example 18.10 The energy gap of silicon is 1.1 eV. Its electron and hole mobilities at room temperatures are 0.48 and 0.013 $m^2V^{-1}s^{-1}$. Evaluate the carrier concentration and its electrical conductivity.

Given Data

Energy gap,	$E_g = 1.1$ eV
Electron mobility,	$\mu e = 0.48\ m^2V^{-1}s^{-1}$
Hole mobility,	$\mu_h = 0.013\ m^2\ V^{-1}s^{-1}$
Temperature,	$T = (27^0C + 273) = 300$ K
Boltzmann's constant	$k = 1.38 \times 10^{-23}$ JK^{-1}

Solution

We know that the intrinsic carrier concentration

$$n_i = 2\left(\frac{2\pi m_e^* kT}{h^2}\right) e^{(-E_g/2kT)}$$

Substituting the given values, we have

$$n_i = 2 \times \left(\frac{2 \times 3.14 \times 9.1 \times 10^{-31} \times 1.38 \times 10^{-23} \times 300}{\left(6.626 \times 10^{-34}\right)^2}\right)^{3/2} \times \exp\left(\frac{-1.1 \times 1.6 \times 10^{-19}}{2 \times 1.38 \times 10^{-23} \times 300}\right)$$

$$n_i = 2 \times \left(\frac{2.3626 \times 10^{-50}}{4.39038 \times 10^{-67}}\right)^{3/2} \times \exp\left(\frac{-1.76 \times 10^{-19}}{828 \times 10^{-23}}\right)$$

$$= 2 \times [5.3942 \times 10^{16}]^{3/2} \times exp^{(-21.256)}$$

$$= 2 \times 1.2528 \times 10^{25} \times 5.8697 \times 10^{-10}$$

$$n_i = 1.4707 \times 10^{16}\ m^{-3}$$

We know that the electrical conductivity,

$$\sigma_i = n_i e\ (\mu_e + \mu_n)$$

Substituting the values, we get,

$$= 1.4707 \times 10^{16} \times 1.6 \times 10^{-9} (0.48 + 0.013)$$

$$\sigma_i = 1.160 \times 10^{-3} \ \Omega^{-1} \ \text{m}^{-1}$$

Therefore, the electrical conductivity of silicon is $1.160 \times 10^{-3} \ \Omega^{-1} \ \text{m}^{-1}$.

Example 18.11 For a silicon semiconductor with a band gap of 1.12 eV, determine the position of the Fermi level at 300 K if $m_e^* = 0.12m_0$ and $m_h^* = 0.28m_0$.

Given Data

Band gap, $E_g = 1.12$ eV

Effective mass of electron $m_e^* = 0.12m_0$

Effective mass of hole $m_h^* = 0.28m_0$.

Temperature T = 300 K

Solution

We know that the Fermi energy is

$$E_F = \frac{E_g}{2} + \frac{3\ kT}{4}\left(\log_e \frac{m_h^*}{m_e^*}\right)$$

Substituting the values. we get,

$$E_F = \frac{1.12}{2} + \frac{3\times1.38\times10^{-23}\times300}{4\times1.6\times10^{-19}}\left(\log_e \frac{0.28\,m_0}{0.12\,m_0}\right)$$

$$= 0.56 + \frac{3}{4}\times0.0258 \times \log_e 2.333$$

$$= 0.56+0.01935\times0.8473=0.56+0.016$$

$$= 0.576 \text{ eV}$$

Therefore, the Fermi energy silicon at 300 K is 0.576 eV.

Example 18.12 Suppose that the effective mass of holes in a material is 4 times that of electrons. At what temperature would the Fermi level be shifted by 10% from the middle of the forbidden energy gap? Given $E_g = 1$ eV.

Given Data

Energy gap $E_g = 1$ eV

$= 1.6 \times 10^{-19}$ J

At temperature *T* $m_h^* / m_e^* = 4$

Solution

We know that the Fermi energy is

$$E_F = \frac{E_c + E_v}{2} + \frac{3kT}{4}\log\left(\frac{m_h^*}{m_e^*}\right)$$

The McGraw·Hill Companies

For O K, E_F is lies between the band gap. Hence, E_F is given by

$$E_F = E_v + 0.5 \text{ eV}$$

$$(E_v + 0.5) \times 1.6 \times 10^{-19} = \frac{E_c + E_v}{2}$$

$$[\because T = 0]$$

Let Fermi level shift by 10% at temperature T, i.e., 0.1 eV.

i.e, $$(E_v + 0.6) \times 1.6 \times 10^{-19} = \frac{E_v + E_0}{2} + \frac{3kT}{4} \log 4$$

Subtracting the above equation, we get,

$$0.16 \times 10^{-19} = \frac{3kT}{4} \log 4$$

Simplifying, we get,

$$T = \frac{4 \times 0.16 \times 10^{-19}}{3 \times 1.38 \times 10^{-23} \times \log 4}$$

$$= \frac{0.6408 \times 10^{-19}}{5.739 \times 10^{-23}} = 116 \text{ K}$$

Therefore, the temperature at watch Fermi level is shifted 10% is 116 K.

Example 18.13 The following data are given for an intrinsic Ge at 300 K. Calculate the conductivity of the sample (Given: $n_i = 2.4 \times 10^{19}$ m^{-3}, $\mu_e = 0.39$ m^2 V^{-1} s^{-1}, $\mu_h = 0.19$ m^2 V^{-1} s^{-1})

Given Data

Intrinsic concentration (n_i) $= 2.4 \times 10^{19}$ m^{-3}
Electron mobility (μ_e) $= 0.39$ m^2 V^{-1} s^{-1}
Hole mobility (μ_h) $= 0.19$ m^2 V^{-1} s^{-1}

Solution The intrinsic conductivity is $\sigma_i = n_i e (\mu_e + \mu_p)$

$$= 2.4 \times 10^{19} \times 1.6 \times 10^{-19} (0.39 + 0.19)$$
$$= 2.22 \ \Omega^{-1} \text{ m}^{-1}$$

The conductivity of Ge at 300 K $\sigma_i = 2.22$ W^{-1} m^{-1}.

Example 18.14 In an *n*-type semiconductor, the Fermi level lies 0.3 eV below the conduction band at 300 K. If the temperature is increased to 330 K, find the new position of the Fermi level.

Solution The concentration of carrier electrons at 300 K is given by

$$N_{i300} = N_C \exp\left[\frac{(E_C - E_{F300})}{kT}\right] \quad \text{(i)}$$

$$= N_C \exp\left[\frac{\left(-0.3\times1.6\times10^{-19}\right)}{k\times300}\right]$$

$$N_{i330} = N_C \exp\left[\frac{-\left(E_C - E_{F300}\right)}{kT}\right] \quad \text{(ii)}$$

Taking $N_{C300} \approx N_{C330}$ from Eqs (i) and (ii) we get,

$$\exp\left[\frac{-0.3\times1.6\times10^{-19}}{k\times300}\right] = \exp\left[\frac{-E_C - E_{F300}}{k\times330}\right]$$

$$\frac{0.3}{300} = \frac{E_C - E_{F300}}{330}$$

$$E_C - E_{F330} = 0.33 \text{ eV}$$

At 330 K, the Fermi energy level lies 0.33 eV, below the conduction band.

Example 18.15 The conductivity of Ge at 20 °C is 2 Ω^{-1} m^{-1}. What is its conductivity at 40 °C? $E_g = 0.72$ eV.

Solution The conductivity is given by

$$\sigma_1 = n_i\, e\, (\mu_e + \mu_m)$$

where

$$n_i = 2\left[\frac{2\pi kT}{h^2}\right]^{3/2} (m_e m_p)^{3/4} \exp\left[\frac{-E_g}{2kT}\right]$$

At 20 °C and 40 °C

$$n_{i20} = 2\left[\frac{2\pi k\times293}{h^2}\right]^{3/2} (m_e m_p)^{3/4} \exp\left(\frac{-E_g}{2k\times293}\right) \quad \text{(i)}$$

$$n_{i40} = 2\left[\frac{2\pi k\times313}{h^2}\right]^{3/2} (m_e m_p)^{3/4} \exp\left(\frac{-E_g}{2k\times313}\right) \quad \text{(ii)}$$

Dividing Eq. (ii) by (i), we get

$$\frac{n_{i40}}{n_{i20}} = \left(\frac{313}{293}\right)^{3/2} \exp\left[\frac{E_g}{2k}\left(\frac{1}{293} - \frac{1}{313}\right)\right] = 2.743$$

i.e.

$$\sigma_{i20} = n_{i20}\, e(\mu_e + \mu_m)$$

$$\sigma_{i40} = n_{i40}\, e(\mu_e + \mu_m)$$

i.e.

$$\frac{\sigma_{i20}}{\sigma_{i40}} = \frac{n_{i20}}{n_{i40}}$$

$$\sigma_{i40} = \sigma_{i20}\,\frac{n_{i40}}{n_{i20}} = 2\times2.743$$

The McGraw·Hill Companies

$$= 5.486\ \Omega^{-1}\ m^{-1}$$

The conductivity of Ge at 40 °C is 5.486 Ω^{-1} m^{-1}.

Example 18.16 The energy gap of Si is 1.1 eV. The average electron effective mass is 0.31 *m*, where *m* is the free electron mass. Calculate the concentration in the conduction band of Si at RT, $T = 300$ K. Assume that $E_F = E_{g/2}$.

Given Data

Energy gap of Si = 1.1 eV
Effective mass of electron = 0.31 m
Temperature = 300 K
$E_F = E_g/2$

Solution For Si, the intrinsic concentration is given by

$$n_i = 2\left[\frac{2\pi kT m_e^*}{h^2}\right]^{3/2} \exp\left[\frac{-(E_C - E_V)}{2kT}\right]$$

Substituting the given values, we get

$$n_i = 2\left[\frac{2\pi \times 1.38\times10^{-23} \times 300 \times 0.31 \times 9.1\times10^{-31}}{(6.626\times10^{-34})^2}\right]^{3/2} \exp\left[\frac{-E_g}{2kT}\right]$$

$$n_i = 4.32166 \times 10^{24} \exp\left[\frac{-1.1\times1.6\times10^{-19}}{2\times1.38\times10^{-23}\times300}\right]$$

$$= 2.5367 \times 10^{15} \text{ electrons per } m^3$$

The intrinsic concentration of Si at 300 K is 2.5367×10^{15} electrons per m^3.

Example 18.17 The Hall coefficient and conductivity of Cu at 300 K have been measured to be -0.55×10^{-10} m^3 $A^{-1}s^{-1}$ and 5.9×10^7 Ω^{-1} m^{-1}, respectively. Calculate the drift mobility of electrons in copper.

Given Data

Hall coefficient of Cu = -0.55×10^{-10} m^3 A^{-1} s^{-1}

Conductivity of Cu = 5.9×10^7 Ω^{-1} m^{-1}

Temperature = 300 K

Solution Drift mobility $= -|R_H\,\sigma|$

$$= 0.55 \times 10^{-10} \times 5.9 \times 10^7$$

$$= 3.2 \times 10^{-3}\ m^2\ V^{-1}\ s^{-1}$$

The drift mobility is given by,

$$\mu_d = 3.2 \times 10^{-3}\ m^2\ V^{-1}\ s^{-1}$$

Example 18.18 Using the electron drift mobility from the Hall effect measurement ($\mu_d = 3.2 \times 10^{-3}\ m^2V^{-1}s^{-1}$), calculate the concentration of conduction electrons in copper, and determine the average number of electrons contributed to the free electron gas per copper atom in a solid (Given: $\sigma = 5.9 \times 10^7\ \Omega^{-1}\ m^{-1}$)

Solution The conductivity of an *n*–type semiconductor is given by

$$\sigma = ne\mu$$

Therefore, $$n_i = \frac{\sigma}{e\mu}$$

$$n_i = \frac{5.9\times10^7}{1.6\times10^{-19}\times3.2\times10^{-3}}$$

$$= 1.15 \times 10^{29}\ m^{-3}$$

The concentration of free electron in pure Cu,

$$n = \frac{\text{Avogadro's constant} \times \text{density} \times \text{number of free electrons per atom}}{\text{Atomic weight}}$$

$$= \frac{6.022\times10^{23}\times8900\times1\times10^3}{63.5}$$

$$= 8.44 \times 10^{28} \text{ electrons per } m^3.$$

The average number of electrons contributed per Cu atom

$$= \frac{\text{Concentration of electrons in an } n\text{-type semiconductor}}{\text{Concentration of electrons in pure Cu atom}}$$

$$= \frac{1.15\times10^{29}}{8.44\times10^{28}} = 1.36$$

Therefore, the average number of electrons contributed per Cu atom is one.

Example 18.19 Find the Hall coefficient and electron mobility of germanium for a given sample (1 cm length, 5 mm breadth, 1 mm thickness). A current of 5 milliamperes flows from a 1.35-volt supply and develops a Hall voltage of 20 nillivolts across the specimen in a magnetic field of 0.45 Wb / m_2.

Given Data

Current through the specimen $I = 5 \times 10^{-3}$ A

Voltage across the specimen $V = 1.35$ V

Length of the sample $L = 1 \times 10^{-2}$ m

Breadth of the sample $b = 5 \times 10^{-3}$ m

Thickness of the sample $t = 1$ mm or 1×10^{-3} m

Area of the sample $a = 5 \times 10^{-6}$ m^2

Hall voltage $V_y = 20 \times 10^{-3}$

Magnetic field $H = 0.45$ Wb m^{-2}

Solution

We know that the resistivity of the Ge sample is

$$\rho = \frac{Ra}{l}$$

$$\frac{V}{I} = \frac{1.35}{5\times10^{-3}}$$

Substituting the values of $R = V/I$, we get,

$$S = \frac{Va}{lI}$$

$$\rho = \frac{1.35\times5\times10^{-6}}{5\times10^{-3}\times1\times10^{-2}} = 0.135\ \Omega m$$

We know that the Hall field, is

$$E_y = \frac{V_y}{Thickness}$$

$$= \frac{20\times10^{-3}}{1\times10^{-3}}$$

$$= 20\ Vm^{-1}$$

Current density given by,

$$J_x = \frac{Current}{Area\ of\ cross\ section}$$

Substituting the values, we get,

$$J = \frac{5\times10^{-3}}{5\times10^{-6}}$$

$$= 1\times10^{3}\ Am^{-2}$$

$$\frac{1}{ne} = \frac{E_y}{HJ_x}$$

$$= \frac{20}{0.45 \times 10^3}$$

$$= 0.044\, m^3\, C^{-1}$$

We know that the Hall coefficient is

$$R_H = \frac{3\pi}{8} \times \frac{1}{ne}$$

Substituting the values, we get,

$$= \frac{3 \times 3.14}{8} \times 0.044$$

$$= 0.0524$$

$$= 0.0524\, m^3\, C^{-1}$$

We know that the electron mobility is,

$$\mu_e = \frac{R_H}{\rho}$$

$$= \frac{0.0524}{0.135} = 0.39$$

$$= 0.39\ m^2 V^{-1} s^{-1}$$

$\therefore$ Therefore, the mobility of the Ge sample is $0.39\ m^2 V^{-1} s^{-1}$

Example 18.20 A 2.0-cm wide and 1.0-mm thick copper strip is placed in a magnetic field with B = 1.5 weber / m^2 perpendicular to the strip. Suppose a current of 200 A is set up in the strip. What Hall potential difference would appear across the strip? Given N = 8.4×10^{28} electrons/m^3.

Given Data

Current flowing $I_x = 200$ A

Applied magnetic field $H_z = 1.5$ Wb m^{-2}

Number of electrons per unit volume $n = 8.4 \times 10^{28}$ electrons m^{-3}

Thickness of the strip $d = 1.0 \times 10^{-3}$ m

Solution

Hall potential, $V_y = \frac{I_x H_Z}{net}$

Substituting the values, we get,

$$V_y = \frac{200 \times 1.5}{8.4 \times 10^{28} \times 1.6 \times 10^{-19} \times 1.0 \times 10^{-3}}$$

The McGraw·Hill Companies

$$= 2.2\times10^{-5} \quad = 22 \times 10^{-6}$$

Therefore, the Hall potential difference appearance between the ship is 22 μ *v.*

Example 18.21 The Hall coefficient of a specimen of a doped silicon is found to be 3.66×10^{-4} m^3/C. The resistivity of the specimen is 8.93×10^{-3} Ω m. Find the mobility and density of the charge carriers

Given Data

Hall coefficient of the specimen $R_H = 3.66 \times 10^{-4}\ m^3\ C^{-1}$

Resistivity of the specimen $\rho = 8.93 \times 10^{-3}\ \Omega\ m$

Solution

We know that the carrier concentration is $n_h = \dfrac{1}{R_H e}$

Substituting the values, we get,

$$= \frac{1}{3.66\times10^{-4}\times1.6\times10^{-19}}$$

$$= 1.708 \times 10^{22}\ m^{-3}$$

The carrier concentration of silicon doped specimen is 1.708×10^{22} m^{-3}

We also know that, the mobility of carrier is,

$$\mu_n = \frac{R_H}{\rho}$$

Substituting the values, we get,

$$\mu_h = \frac{3.66 \times 10^{-4}}{8.93 \times 10^{-3}}$$

$$\mu_h = 0.04099\ m^2\ V^{-1}\ s^{-1}$$

The mobility of silicon doped specimen is 0.04099 $m^2\ V^{-1}\ s^{-1}$.

Example 18.22 An *n*-type semiconductor specimen has Hall coefficient $R_H = 3.66 \times 10^{-11}\ m^3\ A^{-1}\ s^{-1}$. The conductivity of the specimen is found to be $112 \times 10^7\ \Omega^{-1}\ m^{-1}$. Calculate the charge carrier density n_e and electron mobility at room temperature.

Given Data

Hall coefficient, $R_H = 3.66 \times 10^{-11}\ m^3\ A^{-1}\ s^{-1}$

Conductivity, σ $= 112 \times 10^7\ \Omega^{-1}\ m^{-1}$

Solution The Hall coefficient is given by

$$R_H = \frac{3\pi}{8}\frac{1}{ne}$$

$$\therefore \quad n = \frac{3\pi}{8}\frac{1}{R_{H^e}}$$

$$= \frac{3\pi}{8}\frac{1}{3.66\times10^{-11}\times1.6\times10^{-19}}$$

$$= 2\times10^{29}\ \text{m}^{-3}$$

$$\sigma = ne(\mu_e + \mu_h)$$

For *n*-type semiconductor, $\sigma = ne\mu_e$

$$\mu_e = \sigma/ne$$

$$\mu_e = \frac{112\times10^{7}}{1.6\times10^{-19}\times2\times10^{29}}$$

$$= 0.035\ \text{m}^2\ \text{V}^{-1}\ \text{s}^{-1}$$

The concentration of electrons is $2\times10^{29}\ \text{m}^{-3}$.

The electron mobility at room temperature = $0.035\ \text{m}^2\ \text{V}^{-1}\ \text{s}^{-1}$.

Example 18.23 A current of 50 A is established in a slab of copper, 0.5 cm thick and 2 cm wide. The slab is placed in a magnetic field *B* of 1.5 T. The magnetic field is perpendicular to the plane of the slab and to the current. The free electrons concentration in Cu is $8.4\times10^{28}\ \text{m}^{-3}$. What will be the magnitude of the Hall voltage across the width of the slab.

Given Data

Current (i)	= 50 A
Magnetic field (*B*)	= 1.5 T
Thickness of the slab	$= 0.5\times10^{-2}$
Width of the slab (*d*)	$= 2\times10^{-2}$
Concentration of electrons (*N*)	$= 8.4\times10^{28}\ \text{m}^{-3}$

Solution The Hall voltage is given by,

$$V_H = \frac{BI}{new}$$

$$= \frac{1.5\times50}{8.4\times10^{28}\times1.6\times10^{-19}\times2\times10^{-2}}$$

$$= 2.79\times10^{-7}$$

The Hall voltage is 2.79×10^{-7} V.

The McGraw·Hill Companies

Objectives Questions

18.1 In a semiconductor all the states at 0 K in valence band are _____.

18.2 All the states at 0 K in the conduction band of a semiconductor are empty. (True/false)

18.3 The value of concentration of electrons in the conduction band is equal to

a) $n = N_c \exp\left(\frac{E_F - E_C}{kT}\right)$ b) $n = N_c \exp\left(\frac{E_F + E_C}{kT}\right)$

c) $n = N_c \exp\left(\frac{E_F - E_C}{kT}\right)$ d) $n = N_c \exp\left(\frac{E_F + E_C}{kT}\right)$

18.4 The concentration of holes in the valence band is equal to

a) $p = N_v \exp\left(\frac{E_V + E_F}{2kT}\right)$ b) $p = N_v \exp\left(\frac{E_V - E_F}{3kT}\right)$

c) $p = N_v \exp\left(\frac{E_V + E_F}{kT}\right)$ d) $p = N_v \exp\left(\frac{E_V - E_F}{kT}\right)$

18.5 The intrinsic carrier concentration in an intrinsic semiconductor is equal to

a) $n_i = A_o T^{3/2} \exp\left(\frac{E_g}{3kT}\right)$ b) $n_i = A_o T^{3/2} \exp\left(-\frac{E_g}{3kT}\right)$

c) $n_i = A_o T^{3/2} \exp\left(\frac{E_g}{2kT}\right)$ d) $n_i = A_o T^{3/2} \exp\left(-\frac{E_g}{2kT}\right)$

18.6 The the Fermi energy level in an intrinsic semiconductor is equal to _____.

18.7 In an intrinsic semiconductor, the concentrations of electrons and holes are not equal. (True/False).

18.8 The thermal excitation energy produced by an electron in-extrinsic semiconductors at room temperature is equal to _____.

18.9 The acceptor energy level for Ge lying above the valence band is equal to _____.

18.10 In *n*-type semiconductors, the value of Fermi energy level is equal to

a) $E_F = \frac{E_C + E_V}{2}$ b) $E_F = \frac{E_C + E_D}{2}$

a) $E_F = \frac{E_C + E_A}{2}$ d) $E_F = \frac{E_C - E_D}{2}$

The McGraw·Hill Companies

18.11 The concentration of electrons in an *n*-type semiconductor is equal to

a) $n = (N_D)^{1/2}\, 2\left[\frac{2\pi m_e * kT}{h^2}\right]\exp\left(\frac{E_D - E_F}{kT}\right)$

b) $n = (N_D)^{1/2}\, 2\left[\frac{2\pi m_e * kT}{h^2}\right]\exp\left(\frac{E_D + E_F}{kT}\right)$

c) $n = (2N_D)^{1/2}\, 2\left[\frac{2\pi m_e * kT}{h^2}\right]\exp\left(\frac{E_D - E_F}{kT}\right)$

d) $n = (2N_D)^{1/2}\, 2\left[\frac{2\pi m_e * kT}{h^2}\right]\exp\left(\frac{E_D + E_F}{kT}\right)$

18.12 The Fermi energy level in a *p*-type semiconductor is equal to

a) $E_F = \frac{E_V + E_A}{2}$ b) $E_F = \frac{E_V - E_c}{2}$

c) $E_F = \frac{E_V - E_A}{2}$ d) $E_F = \frac{E_V + E_C}{3}$

18.13 The carrier concentration in a *p*-type semiconductor is equal to

18.14 The mobility of charge carrier is σ = _____.

18.15 The hall voltage of a specimen is equal to

a) $V_H = E_H / d$ b) $V_H = E_H \times d$

a) $V_H = 2E_H / d$ d) $V_H = 2E_H \times d$

18.16 According to Mathiessen's rule, combined drift mobility is equal to

a) $\frac{1}{\mu_d} = \frac{2}{\mu_l} - \frac{1}{\mu_L}$ b) $\frac{1}{\mu_d} = \frac{2}{\mu_l} + \frac{1}{\mu_L}$

c) $\frac{1}{\mu_d} = \frac{1}{\mu_l} - \frac{1}{\mu_L}$ d) $\frac{1}{\mu_d} = \frac{1}{\mu_l} + \frac{1}{\mu_L}$

18.17 In semiconductors, the electrical conductivity is proportional to concentration of carrier. (True/False)

KEY

18.1 Occupied
18.2 True
18.3 a)
18.4 d)
18.5 d)
18.6 $\frac{E_V + E_C}{2}$
18.7 False
18.8 0.025 eV
18.9 0.01 eV
18.10 b)
18.11 c)
18.12 a)
18.13 $(2N_a)^{\frac{1}{2}}\left[\frac{2\pi M_h kT}{h^2}\right]^{3/4}\exp\left(\frac{E_V - E_A}{2kT}\right)$
18.14 n e μ
18.15 b)
18.16 d)
18.17 True

The McGraw·Hill Companies

Exercises

18.1 Calculate the intrinsic conductivity of Si at 300 K.

$\mu_e = 0.135\ m^2\ V^{-1}\ s^{-1}$
$\mu_h = 0.048\ m^2\ V^{-1}\ s^{-1}$
$E_g = 1.11\ eV$
$m_{e^*} = 0.26\ m_0$
$m_{n^*} = 0.39\ m_0$

18.2 Calculate the intrinsic carrier concentration of GaAs at 300 K. Given that the electron effective mass is 0.07 m, the hole effective mass is 0.56 m and its energy gap is 1.4 eV.

18.3 The resistance of a CdSe crystal is 10 Ω cm at 300 K. Find its resistance at 350 K. The band gap of CdSe is 1.74 eV.

18.4 The resistance of a semiconductor at different temperatures is given below:

T (K)	*ρ (Ω m)*
300	89
340	165
380	365
420	1012
450	9658

Find the band gap of that semiconducting material.

18.5 SnSe is a *p*-type semiconductor. Its Fermi level lies just 0.2 eV above the valence band at 300 K. Find the position of the Fermi level at 370 K.

Short Questions

18.1 What is meant by Fermi level?
18.2 What is meant by Fermi temperature?
18.3 Distinguish between the intrinsic and extensic semiconductor?
18.4 State mass action law.
18.5 What is an intrinsic semiconductor?
18.6 What is an extrinsic semiconductor?
18.7 What is an *n*-type semiconductor?
18.8 What is a *p*-type semiconductor?
18.9 What is meant by donor energy level?
18.10 What is meant by acceptor level?
18.11 Write down the expression for carrier concentration for a *p*-type semiconductor

The McGraw·Hill Companies

18.12 Write down the expression for carrier concentration for a *n*-type semicondictor?
18.13 Define the term conductivity of a semiconductor.
18.14 Define the term mobility of a semiconductor.
18.15 What is Hall effect?
18.16 Explain the origin of Hall effect.
18.17 What is Hall coefficient?
18.18 What is meant by Hall voltage?
18.19 Mention the applications of Hall effect.

Descriptive Questions

18.1 (a) Derive the expressions for concentration of holes in a valence band and that of electrons in a conduction band for an intrinsic semiconductor.
(b) Deduce an expression for intrinsic concentration using mass action law.
18.2 Deduce the mathematical expressions for the concentration of electrons and hole respectively in the conduction band and valence band for an intrinsic semiconductor and hence show that the Fermi level lies at the middle for an intrinsic semiconductor.
18.3 Obtain an expression for the conductivity of an *n*-type semiconductor and hence explain the determination of energy gap.
18.4 Derive an expression for Fermi level in *n*-type semiconductor and hence obtain an expression for carrier concentration of an *n*-type semiconductor.
18.5 Derive an expression for Fermi level in a *p*-type semiconductor and hence obtain an expression for concentration of hole in a *p*-type semiconductor.
18.6 (a) What is Hall effect?
(b) Obtain an expression for Hall coefficient.
(c) Explain the experimental technique used to find Hall coefficient.
(d) List the advantages of Hall effect.
18.7 Write briefly, about the temperature dependence of carrier concentration and Fermi level in an extrinsic semiconductor.

Chapter

19

Superconducting Materials

OBJECTIVES

- To explain the superconducting phenomena and its applications
- To discuss the important properties of superconducting materials
- To explain the diamagnetic property of superconductors
- To explain the BCS theory for superconducting materials
- To discuss the development of high temperature superconductors
- To discuss the important applications of superconductors

19.1 INTRODUCTION

The helium gas was liquefied at 4.2k by Kamerlingh Onnes in 1908. Liquid He has a temperature of 4.2 K. Further, he studied the properties of Hg at very low temperatures. He found that the resistivity of Hg suddenly dropped to zero at 4.2 K (liquid He temperature). At 4.2 K, the resistivity of Hg is in the order of 10^{-5} Ω cm, i.e., at 4.2 K, Hg is converted into a superconducting material.

For a normal conductor, the relation between temperature and resistivity is shown in Fig. 19.1.

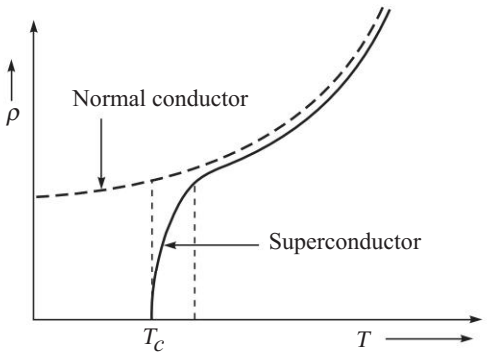


Fig. 19.1 *Temperature versus resistivity*

At very low temperatures, a normal conductor has some resistivity. But for a superconductor, the resistivity is suddenly dropped to zero at very low temperatures. This is shown in Fig. 19.1.

Let T_c be the transition temperature. It is defined as the temperature at which a normal conductor is converted into a superconductor. It is also known as critical temperature. Nowadays, superconduction can be achieved even at higher transition temperatures.

For semiconductors, T_c varies from 0.3 K (GeTe) to 1.25 K (NbO); for metals, T_c varies from 0.35 K (Hafnium) to 9.22 K (Niobium); and for alloys, from 18.1 K (Nb_3Sn) to 22.65 K (Nb_3 Ge).

Good superconducting materials like zinc and lead are not good electrical conductors. Good electrical conductors like copper and gold are not good superconductors. Addition of impurities destroy the superconducting property. A normal conductor is brought into a superconducting state by increasing its pressure.

19.2 GENERAL PROPERTIES OF SUPERCONDUCTING MATERIALS5

A theoretical explanation for superconductivity was given by Bardeen, Cooper and Schrieffer in 1957 and hence, the theory is known as BCS theory.
At the transition temperature, the following changes are observed.

(i) The electrical resistivity drops to zero.
(ii) The magnetic flux lines are expelled from the material.
(iii) There is a discontinuous change in the specific heat.
(iv) Further, there are also small changes in the thermal conductivity and the volume of the materials.

19.2.1 Electrical Resistance

The electrical resistance of a superconducting material is in the order of 10^{-5} Ω cm. The ratio of the resistance of a material in the superconducting state (ρ_s) to the resistance of the same material in the normal state (ρ_n) is less than 10^{-5},

i.e., $$\frac{\rho_s}{\rho_n} < 10^{-5} \qquad (19.1)$$

19.2.2 Magnetic Property

When the superconducting materials are subjected to a large value of magnetic field, it will result in the destruction of the superconducting property. The minimum field required to destroy the superconducting property is given by

$$H_c = H_0\left[1 - \frac{T^2}{T_c^2}\right] \qquad (19.2)$$

where H_0 is the field required to destroy the superconducting property at 0 K, H_c the minimum field required to destroy the superconducting property at T K and T_c the transition temperature of the material.

The McGraw·Hill Companies

The magnetic properties of the material can be represented graphically as shown in Fig. 19.2.

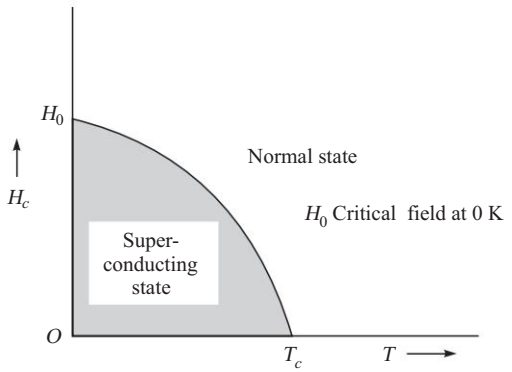


Fig. 19.2 *Magnetic field versus temperature*

19.2.3 Diamagnetic Property

When a normal conducting material is placed in a magnetic field of flux density B, the lines of forces penetrate through the material, as shown in Fig. 19.3a. On the other hand, if the material is cooled for superconductivity, the magnetic lines of forces are ejected from the material, as shown in Fig. 19.3b. A diamagnetic material also repels the magnetic field. Therefore, a superconducting material behaves as a perfect diamagnetic material. This behaviour was first observed by Meissner and hence, this property is known as *Meissner effect*.

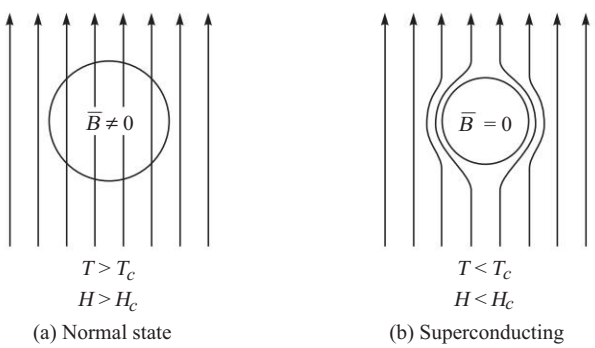


Fig. 19.3 *Meissner effect*

19.2.4 Effect of Electric Current

The application of a large value of electric current to a superconducting material destroys the superconducting property. Consider a coil of wire wound on a superconductor as shown in Fig. 19.4. Let i be the current flowing through the wire. The application of the current induces a magnetic field. Thus, the induced magnetic field in the conductor destroys the superconducting property.

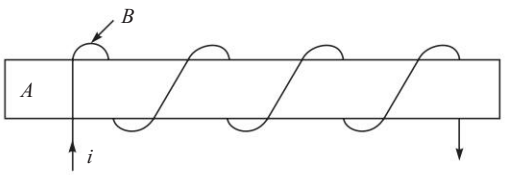


Fig. 19.4 *Effect of current in a superconductor*

The induced critical current (I_c) required to destroy the superconducting property is given by

$$I_c = 2\pi r\, H_c \tag{19.3}$$

where H_c is the critical magnetic field required and r the radius of the superconductor.

19.2.5 Effect of Pressure

Certain materials are brought into the superconducting state by increasing the pressure. For example, cesium is a normal conductor at atmospheric pressure. While increasing the pressure of Cs, it is converted into a superconductor at 110 kbar ($T_c = 1.5$ K).

19.2.6 Isotopic Effect

The presence of isotopes slightly changes the transition temperature of the superconductor. The atomic mass of Hg varies from 199.5 to 203.4. Due to the variation in atomic mass, the transition temperature of isotopes of Hg varies from 4.185 to 4.146 K. Maxwell showed that the transition temperature is inversely proportional to the square root of the atomic mass of the isotope of a single superconductor.

i.e., $$T_c \propto \frac{1}{M^{\alpha}}$$

$$M^{\alpha}\, T_c = \text{constant} \tag{19.4}$$

where α is a constant equal to 1/2 and M the atomic weight.

19.3 TYPES OF SUPERCONDUCTORS

Superconductors are classified into two types. They are type I superconductors and type II superconductors. Type I superconductors are known as soft superconductors and type II superconductors are known as hard superconductors.

19.3.1 Type I Superconductors

Type I superconductors behave as perfect diamagnetic materials and obey the Meissner effect. Figure 19.5 shows the relation between the magnetisation produced and the applied magnetic field for type I superconductors.

A negative sign is introduced in the magnetisation value to represent the diamagnetic property of the superconductor. The material produces a repulsive force up to the critical field H_c. Therefore, due to the repulsive force, it does not allow the magnetic

The McGraw·Hill Companies

field to penetrate through it. Hence, the material behaves as a superconductor. At H_c, the repulsive force is zero and hence, the material behaves as a normal conductor. Sn, Hg, Nb, V, $C_{0.1}$ $T_{0.3}$ $V_{0.6}$ are some examples of type I superconductors.

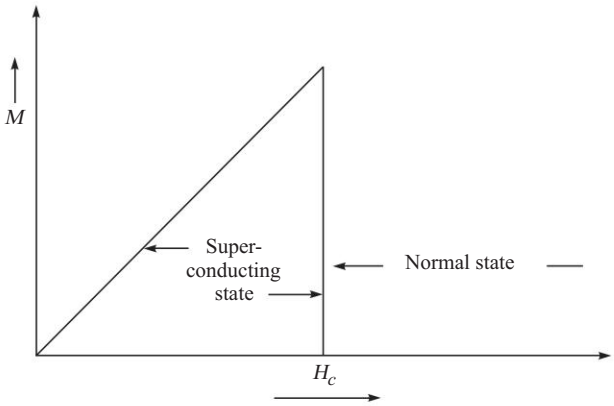


Fig. 19.5 *Type I superconductor*

19.3.2 Type II Superconductors

Type II superconductors do not perfectly obey the Meissner effect. These materials behave as a perfect superconductor up to H_{c1}. Above H_{c1}, the repulsive force decreases, resulting in decrease in the magnetisation M and hence, the magnetic flux starts to penetrate through the material. The magnetic field penetrates up to the value H_{c2}. In the region up to H_{c2}, the material behaves as a superconductor, as shown in Fig. 19.6.

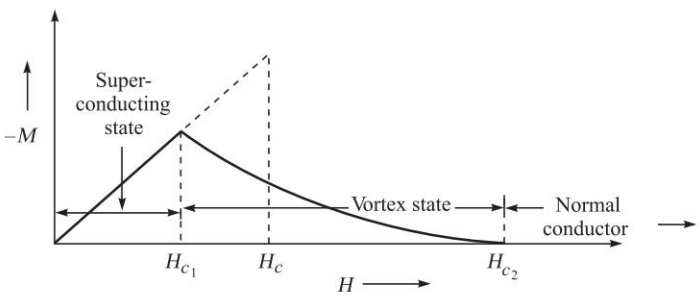


Fig. 19.6 *Type II superconductor*

Let H_{c1} and H_{c2} be the lower and upper critical fields. This region is known as vortex state or mixed state. Above H_{c2}, the materials behave as normal conductors. Examples for type II superconductors are Nb_3 Sn, Nb_3 Ge, YBa_2 Cu_3 O_7.

19.4 BARDEEN, COOPER AND SCHRIEFFER (BCS) THEORY

The theory of superconductivity has been developed in several levels. In 1957, Bardeen, Cooper and Schrieffer proposed a microscopic theory known as BCS theory. The BCS theory explains most of the phenomena associated with superconductivity in a natural

The McGraw·Hill Companies

manner. This theory involves the electron interaction through phonon as mediators. The main idea behind the BCS theory is the experimental results of the two effects, namely, isotope effect and variation of specific heat with temperature. Let us discuss briefly the main postulates of the BCS theory and its major accomplishments.

In 1950, Froblich and Bardeen showed the existence of self-energy of electrons accompanied by virtual phonons when it moves through a crystal lattice. This means that electrons travelling in a solid interact with lattice vibrations by the virtue of electrostatic forces between them. The oscillator distortion of the lattice is quantised in terms of phonons. This interaction is called electron-phonon interaction, which leads to scattering of electrons and hence, causes a change in the electrical resistivity. The resistivity is sensitive with temperature, particularly in the low temperature region, since the number of phonon increases with temperature.

It is assumed from the BCS theory that the electron-phonon interaction produces an attractive interaction between the two electrons. For example, an electron of wave vector K emits a virtual phonon, which is absorbed by an electron K^1. Thus, K is scattered as $K–q$ and K^1+q, as shown in Fig. 19.7. The resulting electron–electron interaction depends on the relative magnitude of the electronic energy change and phonon energy. Once the phonon energy exceeds the electronic energy, the interaction becomes attractive interaction (V_{ph}). Thus, for attractive interaction, the wave vector and spin are represented as $K\uparrow$ and $K\downarrow$. Therefore, the two electrons interacting attractively in the phonon field are called cooper pair and the same is shown in Fig. 19.7.

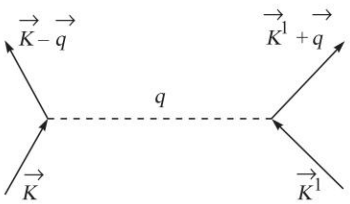


Fig. 19.7 *Electron–phonon interaction*

19.5 ELECTRON–PHONON INTERACTION

On the other hand, these electrons, due to their charge, also repel each other with an interaction V_c, but Bardeen, Cooper and Schrieffer assumed that in the superconducting phase, V_{ph} dominates over V_c. Therefore, the net interaction is attractive

$$V = -V_{ph} + V_c < 0 \tag{19.5}$$

The energy of the pair of electrons in the bond structure is less than in the free state, where the electron exist separately. The difference in the energy between the two states is known as binding energy of the Cooper pair. The electron lattice electron interaction is stronger than electron coulomb interaction when the temperature is less than critical temperature. The Cooper pair is completed at $T = 0$ K and is completely broken when it reaches the critical temperature T_c.

The difference in the energy between the free state of the electron and the paired state appears as the energy gap at the Fermi surface. At the Fermi surface, the normal

The McGraw·Hill Companies

electron states are above the energy gap, while superconducting states are below the energy gap. At absolute 0 K, the pairing is completed and hence, the difference in energy of free and paired electron states, that is normal and superconducting electron states is maximum. At $T = T_c$, the pairing is dissolved and hence, energy gap reduces to zero. Thus, across the energy gap, there are many excited states for the superconducting Cooper pairs.

BCS theory has solved the problem of the electron energy when the attractive interaction V is present and the expression for critical temperature is given by

$$T_c = 1.14\ \theta_D \exp\left[-1/VN\left(E_F\right)\right] \tag{19.6}$$

where θ_D is the Debye temperature, $N(E_F)$ the density of the electron states at the Fermi surface and V the net attractive interaction. Equation (19.6) also accounts for isotopic effect since θ_D is proportional to $M^{-1/2}$. The BCS theory explains the existence of Meissner effect, coherence length and the energy gap parameters.

19.6 HIGH TEMPERATURE SUPERCONDUCTORS

Very few superconductors have been discovered between 1911 and 1986. The observed highest T_c had gone from 4 K to about 23 K in the above period and the work was accomplished with liquid helium. Later, the materials with T_c up to 40 K, 93 K and 125 K have been discovered, respectively, in the years 1983, 1987 and 1988. The new materials with T_c crossing 77 K is more important, since the cooling is accomplished by liquid nitrogen instead of liquid helium. The above trend confirms that it might be possible to develop the superconductors at room temperature. The superconductors having high T_c values are called as high temperature superconductors (HTSC).

Generally, the oxide compound ($BaPb_xBi_{1-x}O_3$) with pervoskite structure (Sleight, 1975) shows T_c at about 13 K. Muller and Bendorz have tried to investigate the occurrence of superconductivity with other oxide compounds and found a T_c of about 35 K in solid solutions of $La_{2-x}\ Ba_x\ CuO_4$. Chu, in 1987, showed that another compound $YBa_2Cu_3O_{7-x}$ became superconducting at 90 K. In 1988, various groups working in USA and Japan discovered thallium and bismuth group of high T_c oxide compounds. Later, HTSCs were also achieved by the substitution of Pb and Tl compounds.

The HTSC compounds are normally represented by simplified notations such as 1212, 1234, etc. These notations are based on the number of atoms on each metal element. For example, the compound $YBa_2Cu_3O_3$ is represented in the simplified notation as 123, based on the number of metal elements. Some of the HTSC components along with transition temperature and notations are given in Table 19.1.

The crystal structure of the superconductors given in Table 19.1, namely, $La_{2-x}M_xCuO_4$ (M = Sr, Ba) among the oxide superconductors, is shown in Fig. 19.8. The alkaline earth elements Ca, Sr and Ba can be substituted on La site and is represented by M.

Thus, it is clear that the high T_c superconductors are not metal or intermetallic compounds, but they are oxides of copper in combination with other elements. As a result of the combinations, the HTSC materials become brittle and it is easy to form wires and tapes. The HTSC wires/tapes provides a transmission of electrical power over a long distance without any resistive losses.

Table 19.1 *Structure and properties of cuperate superconductors*

Notations	*Chemical formula*	*Structure*	*a* (Å)	*b* (Å)	*c* (Å)	T_C (K)
214	$La_{1.82}Sr_{0.18}CuO_4$	Tetragonal	3.77	–	13.25	38
123	$Yba_2Cu_3O_7$	Orthorhombic	3.81	3.88	11.63	90
Tl-1212	$TlBa_2CaCu_2O_7$	Tetragonal	3.83	–	19.68	80
Tl-1223	$TlBa_2Ca_2Cu_3O_9$	Tetragonal	3.84	–	15.88	105
Tl-1234	$TlBa_2Ca_3Cu_4O_{11}$	Tetragonal	3.85	–	12.10	120
Tl-2201	$TlBa_2CuO_6$	Tetragonal	3.86	–	23.24	15
Tl-2212	$Tl_2Ba_2Ca_2Cu_3O_{10}$	Tetragonal	3.86	–	29.39	108
Tl-2223	$Tl_2Ba_2Ca_2Cu_3O_{10}$	Tetragonal	3.85	–	35.60	125
Tl-2234	$Tl_2Ba_2Ca_3Cu_4O_{12}$	Tetragonal	3.85	–	42.00	105
BI-2201	$Bi_2Sr_2CuO_6$	Orthorhombic	5.36	5.37	24.62	6
BI-2212	$Bi_2Sr_2CaCu_2O_8$	Orthorhombic	5.50	5.41	30.78	81
BI-2223	$Bi_2Sr_2Ca_2Cu_3O_{10}$	Orthorhombic	5.40	5.41	37.18	110

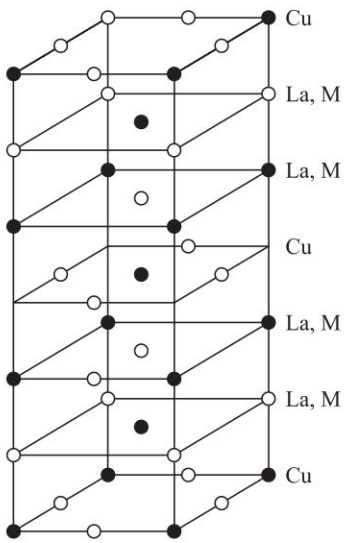


Fig. 19.8 *Tetragonal crystal structure of $La_{2-x}M_xCuO_4$ (M = Sr, Ba)*

19.7 APPLICATIONS

Due to the vast development of superconducting materials, they find a large range and diversity of applications in different fields. Following are some of the important applications.

19.7.1 Superconducting Magnets

Normally, when current flows through a coil, it generates magnetic fields. If the coil is replaced by a superconducting material, it generates a large magnetic field. In the

The McGraw·Hill Companies

conventional high field electromagnets, the effect of joule heating is a major problem, while in superconducting materials, the current flows without any resistive loss and joule heating.

Therefore, high critical magnetic field superconductors are used for these applications. The most widely used superconducting materials are type II superconductors such as Nb–Ti and Nb_3Sn compounds. The important applications of superconducting magnets are NMR, medical diagnostics and spectroscopy, magnetic leviation, magnetic shielding, etc.

19.7.2 Josephson Effect

In 1962, Brian Josephson predicted the flow of current in an insulator (oxide), which is sandwiched in between two superconductors and a dc voltage is applied as shown in Fig. 19.9. The current flowing through the device has both ac and dc components. The ac component exists only during the existence of the applied voltage V, while the dc component persists even after the removal of applied voltage. This effect is known as Josephson effect. The frequency of the ac components is $\gamma = 2$ eV/h irrespective of the material used, and hence the value of e/h can be determined more precisely.

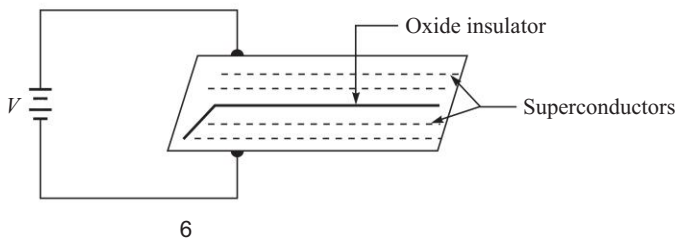


Fig. 19.9 *Josephson effect*

19.7.3 Superconducting Quantum Interface Device (SQUID)

SQUID is a magnetometer, which involves the super current properties of the Josephson junctions. A typical SQUID system used for sensing magnetic field is shown in Fig. 19.10. The low temperature superconductors are being used for fabricating this SQUID. The main applications of SQUID are to detect a small fractional change in flux, geological layers in different minerals, to detect NMR signals at low temperature, etc.

19.7.4 Maglev

Maglev is the acronym of magnetic levitation. Maglev a system which is used to run the vehicle levitated from the guideway i.e., the rail tracks of conventional railways. The principle behind maglev is the electromagnetic forces between the superconductivity magnets. Maglev system consists of 8 figured cavitation coils and sidewalk on the guide ways. The schematic representation of the maglev system is shown in Fig. 19.11.

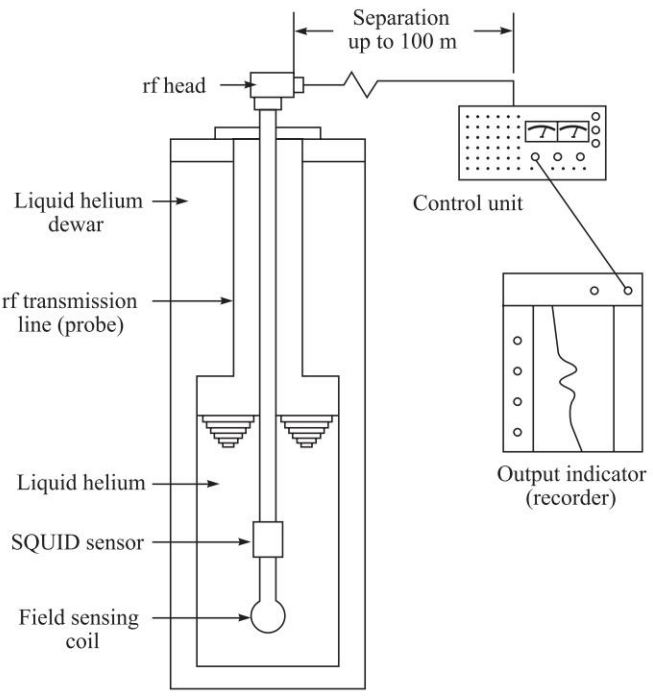
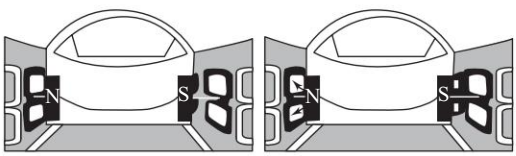
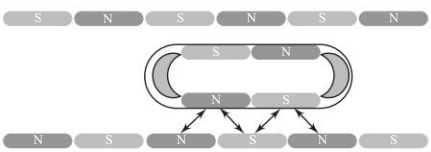


Fig. 19.10 *SQUID system for sensing external magnetic field*



(a) *Maglev– later guidance*



(b) *Maglev– Propulsion*

Fig. 19.11 *Maglev system*

The guide ways for the maglev is similar to the rail tracks of the conventional railways. Consider that an onboard superconductivity magnets passes at a high speed with several centre meters below the centre of 8- figured coils. An electric current is induced within the levitation coils. Thus, the levitation coil act as electromagnet and hence, produces an electromagnetic forces which push the on-board superconducting

The McGraw·Hill Companies

magnet to upward. Therefore, the maglev system is levitated and hence, the onboard vehicle is levitated from the ground. The levitation coils which are facing each other are connected through a loop under the guide way. When the on board vehicle is levitated laterally, an electric current is induced in the loop and hence, creates a repulsive and attractive forces. The repulsive force is acting on the levitation coil which is near the onboard vehicle while attractive force acting on the levitation coil side further apart from the on board vehicle. Therefore, running on board vehicle is always located at the centre of guideway.

The onboard superconductivity magnet is accelerated employing the repulsive force and attractive force induced between the magnets. The propulsion coils which are fitted on the sidewall on both sides of guide tube is used to accelerate onboard vehicle. The acceleration is achieved by giving a sufficient energy (three phase a.c.) to the propulsion coils from the substation. Therefore, the on board vehicle is accelerated.

19.7.5 Other Applications

(i) Superconducting electric generators are small in size compared to conventional electric generators. Superconducting generators produce more power compared to ordinary generators.
(ii) Superconducting materials are used as no loss transmission lines.
(iii) In Japan, superconducting materials are used to levitate a train above its rail.
(iv) Superconducting materials are used as a storage device in computers.
(v) Superconducting materials are used as relay in switching circuits.

Cryotron It consists of two superconducting materials A and B as shown in Fig. 19.12.

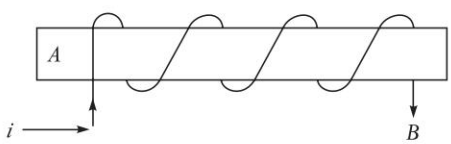


Fig. 19.12 *Cryotron*

Let H_{ca} and H_{cb} be the critical fields of the materials A and B, respectively. Let $H_{ca} < H_{cb}$. Consider that a current i is passed through the material B. The current induces some magnetic field H. If H lies between H_{ca} and H_{cb}, then the induced field will destroy the superconducting property of the material A. Hence, the resistivity increases and the contact is broken.

Keypoints to Remember

- When the resistivity of a material is suddenly dropped to zero at very low temperature. *At this zero resistivity*, the material will conduct electricity. These materials are known as superconductors

- The minimum field required to destroy superconductivity is $H_C = \left[1 - \frac{T^2}{T_c^2}\right]$, where H_0 is the field required to destroy the superconducting property at 0 K and T_C, the transition temperature of the material.
- A superconducting material behaves as a perfect diamagnetic material. When the material is cooled for superconductivity, the magnetic lines of forces are ejected from the material. This property is known as Meissner effect.
- A large value of electric current applied to a superconducting material destroys the superconducting properties.
- The transition temperature is inversely proportional to the square root of atomic mass of isotope of a single superconductor, i.e., $T_C \alpha \frac{1}{\mu^\alpha}$, where μ is the atomic mass of the isotope and α, a constant equal to 0.5.
- Type I superconductors are perfect diamagnetic materials and obey the Meissner effect.
- Type II superconductors are not perfect diamagnetic materials and hence, do not obey the Meissner effect.
- When a dc potential is applied to an oxide insulator sandwiched between superconductors, the ac component of current exists only during the existence of applied voltage, while the dc component of current persists even after removal of current. This phenomenon is known as Josephson effect.

Solved Problems

Example 19.1 Prove that superconductors are perfectly diamagnetic.

Solution For a magnetic material, the flux density is given by

$$B = \mu_0 (M + H)$$

For a superconducting material, $B = 0$, i.e. $0 = \mu_0 (M + H)$

or $$M + H = 0$$

$$M = -H$$

$$\frac{M}{H} = \chi = -1$$

For a diamagnetic material, the susceptibility is negative. This equation proves that a superconductor behaves as a perfect diamagnetic material.

Example 19.2 Superconducting tin has a critical temperature of 3.7 K at magnetic filed and a critical field of 0.0306 Tesla Find the critical field at 2 K.

Given Data

Critical temperature $T_c = 3.7$ K
Critical temperature $H_0 = 0.0306$ T

Solution

We know that the critical field

$$H_c = H_0 \left[1 - \frac{T^2}{T_c^2}\right]$$

Substituting the values, we get,

$$= 0.0306 \left[1 - \frac{(2)^2}{(3.7)^2}\right]$$

$$= 0.02166 \text{ T}$$

Therefore, the critical field at 2 K is 0.02166 T.

Example 19.3 Calculate the critical current for a wire of lead having a diameter of 1 mm at 4.2 K. Critical temperature for lead is 7.18 K and $H_c\,(0) = 6.5 \times 10^4$ A m^{-1}

Given Data

Critical temperature $T_c = 7.18$ K
Critical magnetic field $H_0 = 6.4 \times 10^4$ A m^{-1}

Solution

We know that the critical field

$$H = H_0 \left[1 - \frac{T^2}{T_c^2}\right]$$

Substituting the values, we get,

$$= 6.5 \times 10^4 \times \left[1 - \left(\frac{4.2}{7.18}\right)^2\right]$$

$$= 42.758 \times 10^3 \text{ A m}^{-1}$$

Therefore, the critical current through the wire is

$$I_c = 2\pi r H_c$$

Substituting the values, we get

$$I_c = 2 \times 3.14 \times 0.5 \times 10^{-3} \times 42.758 \times 10^3$$

$$I_c = 134.26 \text{ Amp}$$

The critical current through a wire of lead is 134.26 A.

Example 19.4 The critical temperature for a metal with isotopic mass of 199.5 is 4.185 K. Calculate the isotopic mass if the critical temperature falls to 4.133 K

Given Data

Critical temperature of lead $T_{c_1} = 4.185$ K

Critical magnetic field at 0K $T_{c_2} = 4.133$ K

Diameter of the lead at 4.2 K $r = 1 \times 10^{-2}$ m

Solution

We know that the transition temperature for metal with isotopic mass is

$$T_{c2} = T_{c1}\left[\frac{M_1}{M_2}\right]^{\frac{1}{2}}$$

Substituting the values, we get,

$$= 4.184 \times \left[\frac{199.5}{203.4}\right]^{\frac{1}{2}}$$

$$= 4.144 \text{ K}$$

The critical temperature for metal with isotopic mass of 203.4 is 4.144 K.

Objectives Questions

19.1 The resistivity of liquid helium drops to zero at _____ K.

19.2 A normal metal is converted into a superconductor at a temperature known as _____.

19.3 The T_c values of NbO, hafmium and niobium are respectively _____, _____ and _____ K.

19.4 The acronym of BCS theory is _____.

19.5 When a material is cooled below superconducting temperature, the magnetic lines of forces penetrat through the material. (True/False)

19.6 The current required to destroy the superconducting property is equal to

a) $I_C = 2\,\pi\, r\, H_O$ b) $I_C = 2\pi\, H_C$

c) $I_C = 2\,\pi\, r\, H_C$ d) $I_C = 4\,\pi\, \text{r}\, H_C$

19.7 The isotropic effect for a single superconductor is

a) $m^{\alpha}\, T_C = \text{Constant}$ b) $m^{\alpha}\, T_C = 0$

c) $m^{\alpha}\, T_C = 2$ d) $m^{\alpha}\, T_C = 3$

The McGraw·Hill Companies

19.8 Type I superconductors do not obey the Meissner effect. (True/False)
19.9 The intermediate state between a superconducting and normal conductor is _____ state.
19.10 Type II superconductors are perfect diamagnetic materials. (True/False)
19.11 BCS theory is used to explain _____
19.12 According to BCS theory, the expression for critical temperature is equal to

a) $T_C = 1.14\ \theta_D \exp\left(\frac{-1}{VN(E_F)}\right)$ b) $T_C = 1.24\ \theta_D \exp\left(\frac{-1}{VN(E_F)}\right)$

c) $T_C = 1.34\ \theta_D \exp\left(\frac{-1}{VN(E_F)}\right)$ d) $T_C = 1.44\ \theta_D \exp\left(\frac{-1}{VN(E_F)}\right)$

19.13 BCS theory does not explain the following
a) Meissner effect b) Coherence effect
c) Isotopic effect b) Energy gap parameters
19.14 The value of T_C for a high-temperature superconductor is low. (True/False)
19.15 Expand HTSC.
19.16 Expand SQUID.

KEY

19.1 4.2
19.2 Critical temperature T_C
19.3 1.25,0.35 and 9.22
19.4 Bardeen, Copper and Schrieffer
19.5 False
19.6 c)
19.7 a)
19.8 False
19.9 Vortex
19.10 False
19.11 d)
19.12 a)
19.13 c)
19.14 false
19.15 High-temperature superconductors
19.16 Superconducting quantum interface device

Short Questions

19.1 What are superconductors?
19.2 Explain Meissner effect.
19.3 What are type I and type II superconductors?
19.4 What are it propertics of the super conductor?
19.5 What is meant by isotopic effect?
19.6 What are Cooper pairs?
19.7 What is a high-temperature superconductor?
19.8 State Josephson effect.

The McGraw·Hill Companies

19.9 What is SQUID?
19.10 What are the applications of SQUID?
19.11 Mention some of the oxide superconductors.

Descriptive Questions

19.1 What are superconductors? Explain the occurrence of superconductivity. What are the different types of superconductors? Explain them in detail.
19.2 Explain the properties of superconducting materials in detail.
19.3 What is BCS theory? Enumerate the important results of BCS theory.
19.4 Describe the applications of superconductors in various fields like superconducting magnets and SQUID.

Chapter

Magnetic Materials

20

OBJECTIVES

- To understand the principle, theory, properties and applications of magnetic materials
- To study the different types of magnetic materials along with suitable examples
- To derive the necessary theory for diamagnetism, paramagnetism, ferromagnetism and anti-ferromagnetism
- To discuss the experimental methods used to study the paramagnetic properties of materials
- To discuss the structure, property and applications of ferromagnetic materials
- To study the hard and soft magnetic materials and the energy product of magnetic materials
- To study the magnetic materials used in storage devices and their applications

20.1 INTRODUCTION

The materials which strongly attract a piece of iron are known as *magnetic materials* or *magnets*. The magnetic property of a material arises due to the magnetic moment or magnetic dipole of materials. Materials which are magnetised by the application of an external magnetic field are known as *magnetic materials*. Similarly, materials which are not magnetised due to the application of an external magnetic field are known as *non-magnetic materials*. The magnetism of materials is responsible for magnetic moment of materials. When a current flows through a conductor, it produces a magnetic moment along the axis of the coil. Similarly, the electrons revolving around the nucleus leads to an *orbital magnetic moment*. The *spin magnetic moment* arises due to the spin ($\pm 1/2$) of electrons. Magnetic materials are more important in terms of potential practical applications like magnetic storage and biomedical uses. General classes of magnetic materials are diamagnetic, paramagnetic, ferromagnetic, antiferromagneitc and ferrimagnetic materials.

In this chapter, the basic parameters and classification of magnetic materials are discussed in the first few sections. The different types, theories, properties and applications of magnetic materials are discussed in following sections.

The McGraw·Hill Companies

20.2 MAGNETIC PARAMETERS

The magnetic property of materials depends on the degree of magnetisation. The parameters such as magnetisation, magnetic susceptibility and magnetic permeability are used to characterise the magnetic materials. Let us discuss the important magnetic parameters which are used to characterise the magnetic materials.

(i) Magnetic Dipole The magnetic dipoles, generally known as north and south poles, commonly exist in magnetic materials. The magnetic dipoles are not separate poles unlike an electric dipole.

(ii) Magnetic Field Strength The magnetic field strength H at any point in a magnetic field is the force experienced by a unit north pole placed at that point. Its unit is A m^{-1}.

(iii) Magnetic Induction or Flux Density The magnetic induction or flux density B in any material is defined as the number of lines of force through a unit area of cross-section perpendicularly.

Therefore,

$$\text{Magnetic induction } B = \frac{\phi}{A} \tag{20.1}$$

where A is the area of cross-section and ϕ, the magnetic force. The unit for B is Wb m^{-2}.

(iv) Magnetic Dipole Moment Consider that m is the magnetic pole strength and $2l$, the length of magnet. The magnetic dipole moment is equal to the product $m \times 2l$.

$$\text{Magnetic dipole moment } \mu_m = m \times 2l \tag{20.2}$$

The magnetic dipole moment is a vector quantity.

(v) Magnetisation or Intensity of Magnetisation Magnetisation, or intensity of magnetisation, is the measure of magnetism of magnetic materials and is defined as the magnetic moment per unit volume.

$$M = \frac{M_m}{V} \tag{20.3}$$

where V is the volume. The unit for M is A m^{-1}.

(vi) Magnetic Susceptibility Magnetic susceptibility (χ) is used to explain the magnetisation of materials. It is defined as the ratio of magnetisation to the magnetic field strength.

$$\text{Magnetic susceptibility } \chi = \frac{M}{H} \tag{20.4}$$

The McGraw·Hill Companies

(vii) Magnetic Permeability The magnetic permeability μ is defined as the ratio of amount of magnetic density B to the applied magnetic field intensity H. It is used to measure the magnetic lines of forces penetrating through a material.

$$\mu = \frac{B}{H} \tag{20.5}$$

Rearranging the above equation, we get,

$$\text{Magnetisation } B = \mu_0 (M + H) \tag{20.6}$$

where μ_0 is the permeability of free space and is equal to $4\pi \times 10^{-7}$ H m^{-1}

Relation between Magnetic Parameters

Relative permeability of a material is defined as the ratio of permeability of the medium to free space,

i.e., $$\text{Relative permeability } \mu_r = \frac{\mu}{\mu_0} \tag{20.7}$$

Substituting the values of $\mu = B/H$, we get,

$$\mu_r = \frac{B}{\mu_0 H} \tag{20.8}$$

Substituting B value from Eq. (20.6) in the above equation, we get,

$$\mu_r = \frac{\mu_0 (M + H)}{\mu_0 H}$$

$$\mu_r = \frac{M}{H} + \frac{H}{H} \tag{20.9}$$

Therefore, relative permeability $$\mu_r = 1 + \chi \tag{20.10}$$

It is clear that the magnetic parameters are analogous to dielectric parameters. The comparisons of magnetic and dielectric parameters are given in Table 20.1.

20.3 BOHR MAGNETON

The magnetic moment of an atomic particle is represented by unit ampere metre^2. The magnetic moment of an atomic particle is very low, and it is usually represented by a unit known as Bohr magneton. The value of one Bohr magneton is,

$$1 \text{ Bohr magneton} = \frac{eh}{4\pi m} \tag{20.11}$$

Substituting the values of e, m and h, we get,

$$1 \text{ Bohr magneton } (\beta) = \frac{1.6\times10^{-19}\times6.626\times10^{-34}}{4\pi\times9.1\times10^{-31}}$$

$$1\beta = 9.27\times10^{-24} \text{ A m}^2 \qquad (20.12)$$

The magnetic moment of an electron is 1 Bohr magneton.

Table 20.1 *Comparison of magnetic and dielectric parameters*

S. No.	*Dielectric parameters*	*Magnetic parameters*
1.	Electric field	Magnetic field
2.	Electric displacement	Magnetic induction
3.	Electric dipole moment	Magnetic dipole moment
4.	Permittivity	Permeability
5.	Polarisation	Magnetisation
6.	Non-polar dielectric	Diamagnetic
7.	Polar dielectric	Paramagnetic
8.	Ferroelectric hysteresis	Ferromagnetic hysteresis
9.	Remanent polarization	Remanent magnetisation
10.	Coercive electric field	Coercive magnetic field
11.	Ferroelectric domain	Magnetic domain
12.	Ferroelectric Curie temperature	Ferromagnetic Curie temperature
13.	Converse piezoelectric effect	Magnetostriction effect

20.4 CLASSIFICATION OF MAGNETIC MATERIALS

Magnetic materials are classified into two categories based on existence of dipole moment and the response of the magnetic material to external magnetic fields.

(i) Diamagnetic materials—no permanent magnetic moment
(ii) Paramagnetic, ferromagnetic, antiferomagnetic and ferrimagnetic materials—possess permanent magnetic moment.

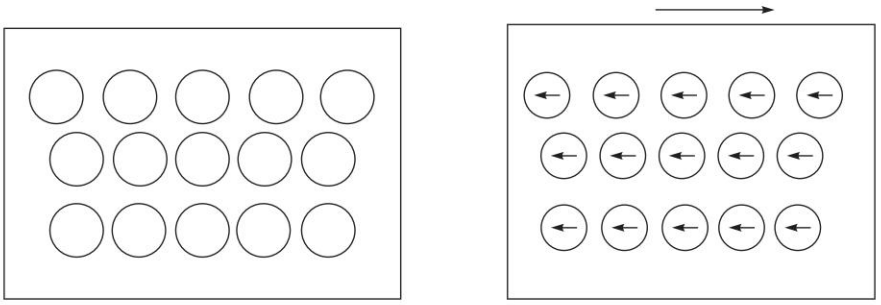
Generally, diamagnetic and paramagnetic materials are known as non-magnetic materials, due to poor response to an external magnetic field. Ferromagnetic, antiferomagnetic and ferrimagnetic materials are known as magnetic materials. These materials are strongly responsive to an external magnetic field. The different types of non-magnetic and magnetic materials are discussed briefly in the following sections.

20.4.1 Diamagnetic Materials

All materials exhibit diamagnetic property. In diamagnetic materials, magnetic dipoles are oriented in such a way that the resultant dipole moment is zero. When an external

The McGraw·Hill Companies

magnetic field is applied, the individual dipoles are rotated and hence, produce an induced dipole moment. The induced dipole moment opposes the external magnetic field. As a result, the magnetic fields are repelled from the materials. This effect is known as *diamagnetism*. The existence of diamagnetism is represented by means of simple representation as shown in Fig. 20.1. During the absence of an external magnetic field, the magnetic moment of atoms is zero. On the other hand, when an external field is applied, atoms acquire induced dipole moment and hence, move opposite to field directions. This results in repulsion of magnetic lines of force.



$H = 0$ and $M = 0$ $H = H_0$ and $M = -M_0$

(a) Absence of field *(b) Presence of field*

Fig. 20.1 *Diamagnetic properties*

Properties

The following are the properties of diamagnetic materials:

(i) They do not have a permanent dipole moment.
(ii) Magnetic effects are very weak and hence, often masked by other kind of magnetism.
(iii) They repel the magnetic lines of force. The existence of this behaviour in a diamagnetic material is shown in Fig. 20.2.

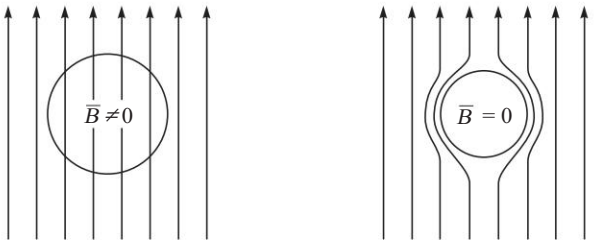


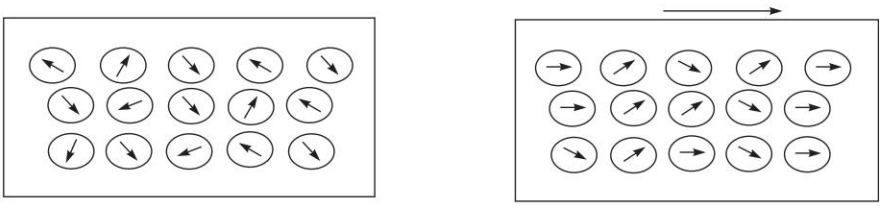
Fig. 20.2 *Diamagnetic material—Magnetic field*

(iv) The magnetisation becomes zero on removal of the external field.
(v) The susceptibility of a diamagnetic material is negative.
(vi) The susceptibility is independent of temperature and external field.

Examples for diamagnetic materials are copper, gold, mercury, silver and zinc.

20.4.2 Paramagnetic Materials

Paramagnetic materials possess a permanent magnetic moment. During the absence of an external magnetic field, the dipoles are randomly oriented. This is due to thermal agitation present in materials. Therefore, the net dipole moment and hence, magnetisation of the material is zero. When an external magnetic field is applied, the magnetic moments of individual atoms align themselves in the direction of the field and hence, give non-zero magnetisation. The behaviour of a paramagnetic material under the influence of an external field is shown in Fig. 20.3.



$H = 0$ and $M = 0$ $H = H_0$ and $M = +M_0$

(a) Absence of field *(b) Presence of field*

Fig. 20.3 *Paramagnetic properties*

Properties

The following are the properties of paramagnetic materials:

(i) Paramagnetic materials possess a permanent dipole moment.

(ii) The magnetic dipoles are aligned randomly. The magnetic spin alignment is shown in Fig. 20.4.

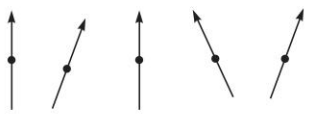


Fig. 20.4 *Paramagnetic material—spin arrangement*

(iii) They attract the magnetic lines of force.

(iv) The susceptibility is positive and depends on temperature:

$$\chi = \frac{C}{T} \tag{20.13}$$

where C is the curie constant and T, the temperature. Equation (20.13) is known as the Curie law of paramagnetism.

(v) Paramagnetic susceptibility is inversely proportional to temperature.

Examples of paramagnetic materials are aluminum, chromium, sodium, titanium, zirconium, etc.

The McGraw·Hill Companies

20.4.3 Ferromagnetic Materials

Ferromagnetic material possess permanent magnetic moment which is mainly due to the spin magnetic moment. The magnetic dipoles are aligned parallel to each other due to interaction between any two dipoles. The ferromagnetic materials exhibit spontaneous magnetisation, even in the absence of an external field. Due to the strong internal field which exists in material, the alignment of magnetic moments results. On the other hand, when a small magnetic field is applied, it produces a large value of magnetisation due to the parallel alignment of dipoles.

In a ferromagnetic material, the influences of internal and external fields make them different from other magnetic materials. The molecular field theory, or Weiss theory, is used to explain the ferromagnetic properties of materials.

Properties

The properties of ferromagnetic materials are as follows:

- The magnetic dipoles are arranged parallel to each other. The spin arrangement is shown in Fig. 20.5.

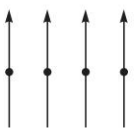


Fig. 20.5 *Ferromagnetic materials—spin arrangements*

- They possess permanent dipole moment.
- They attract the magnetic lines of force strongly.
- They have characteristic temperature, namely, *ferromagnetic Curie temperature* (θ_f). Materials below θ_f behave as ferromagnetic materials and obey hysteretic curve. A material behaves as a paramagnetic when it is above θ_f.
- During the absence of a magnetic field, it exhibits magnetisation which is due to the property of spontaneous magnetisation.
- The susceptibility of a ferromagnetic material is

$$\chi = \frac{C}{T - \theta} \tag{20.14}$$

where C is Curie constant and θ, the paramagnetic Curie temperature.

Equation (20.14) is shown as Curie–Weiss law. This law holds good when the temperature is greater than ($T > \theta$).

Examples for ferromagnetic materials are iron, cobalt and nickel.

20.4.4 Antiferromagnetic Materials

In ferromagnetic materials, we know that the magnetic dipoles are parallel to each other. This is due to the existence of a strong interaction known as *exchange interaction* between the spin magnetic moments. Similarly, in few materials, the exchange interaction leads to antiparallel alignment of dipoles. The magnitudes of all

dipoles are equal and hence, resultant magnetic moment and magnetisation is zero. The antiparallel alignment exists in material below a critical temperature known as *Neel temperature*.

Properties

The important properties of ferromagnetic materials are as follows:

- The dipoles are aligned anti-parallel as shown is Fig. 20.6.

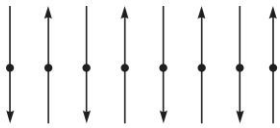


Fig. 20.6 *Antiferromagnetic materials—dipole arrangements*

- When the temperature increases, susceptibility increases and reaches a maximum at a temperature known as Neel temperature beyond which it decreases.
- The value of susceptibility is positive and is very small when T is greater than the Neel temperature, $T_{N.}$

$$\chi = \frac{C}{T + \theta} \tag{20.15}$$

- Anti-parallel alignment of dipole is due to exchange interactions.

Examples of antiferromagnetic materials are ferrous oxide, manganese oxide, manganese sulphite, chromium oxide, etc.

20.4.5 Ferrimagnetic Materials

Ferrimagnetic materials are a special case of antiferromagnetic materials. The magnetic dipoles are antiparallel, similar to an antiferromagnetic material. However, their magnitudes are not equal. This is due to the magnetic interactions existing between dipoles. Therefore, ferrimagnetic materials possess net magnetic moments due to the antiparallel dipoles with different magnitudes. Hence, they produce a large magnatisation even for a small applied external field.

Properties

The properties of ferromagnetic materials are as follows:

- The dipoles are antiparallel as shown in Fig. 20.7. However, the dipoles are not equal in magnitude.
- Net magnetisation is larger even for a small external field.
- The susceptibility is positive and very large when the temperature is higher than T_N,

$$\chi = \frac{C}{T \pm \theta} \tag{20.16}$$

- They behave as paramagnetic and ferromagnetic materials respectively above and below Curie temperature.
- Ferrimagnetic domains are used as magnetic bubbles in memory elements.

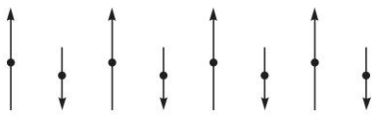


Fig. 20.7 *Ferimagnetic materials—dipole arrangements*

20.5 ORIGIN OF PERMANENT MAGNETIC MOMENT

The term permanent magnetic moment represents the presence of magnetic moment in an atom of a magnetic material even in the absence of a magnetic field. The magnetic moment arises in an element due to the presence of angular momentum. There are three different types of angular momentum, namely,

(i) orbital angular momentum of an electron,
(ii) spin angular momentum of an electron, and
(iii) nuclear spin angular momentum.

Let us discuss the contribution of magnetic moment due to the above different types of angular momenta.

20.5.1 Orbital Angular Momentum of an Electron

The orbital angular momentum of an electron arises due to its orbital motion. The relation between the magnetic moment and orbital angular momentum

$$\mu_m = \frac{e}{2m}\mu_a \tag{20.17}$$

where μ_a is the orbital angular momentum and is equal to

$$\mu_a = \frac{lh}{2\pi} \tag{20.18}$$

According to quantum theory, quantum numbers such as n, l, m_l and m_s are used to represent the state of an electron. The quantum number n is called the *principal quantum number*. It represents the energy possessed by an electron. It takes integer values such as 1, 2, 3… n. The quantum number ℓ is known as *orbital quantum number*. It takes values such as 0, 1, 2 …. (n–1). It represents the orbital motion of an electron. The quantum number m_l is known as *magnetic orbital quantum number*. It represents the orientation of electronic orbits based on the externally applied magnetic field. It takes discrete values ranging from $-\ell$ to $+\ell$ including zero.

The angular momentum of an electron due to its orbital motion is given by$l\,(h/2\pi)$. The quantum numbers for an electron in K shell are $n = 1$, $\ell = 0$, $m_l = 0$ and $m_s = \pm ½$. Substituting $\ell = 0$ in Eq. (20.18), we get the orbital angular momentum is zero. For an electron in the L shell, the quantum numbers are $n = 2$, $\ell = 0, 1$, $m_l = 0, \pm 1$, $m_s = 0$, $\pm ½$. Substituting $\ell = 1$ in Eq. (20.18), we get the angular momentum of an electron as $(h/2\pi)$. There are three possible directions ($m_l = 0, +1, -1$) for the orbital motion of an electron with $\ell = 1$. These three values for m_l contribute three different values of

the angular momentum of an electron. They are $\frac{h}{2\pi}$, 0, $-\frac{h}{2\pi}$. The magnetic moment contributed by electrons with $l = 1$ is written from Eq. (20.18) as,

$$\mu_m = \frac{e}{2m}\frac{h}{2\pi}, 0, -\frac{eh}{4\pi m} \tag{20.19}$$

For an element having completely filled shells, the total orbital magnetic moment is zero because the sum of the magnetic moment is zero. Therefore, if there is an unfilled electronic state in an atom for an element, then the total orbital magnetic moment is not equal to zero. Therefore, materials with unfilled electronic states are important for physicists and electrical engineers. The elements with atomic numbers 21 through 28 (the iron group), 39 through 45, 58 through 71 (the rare earths) and 89 through 92 have unfilled electronic states. For an electrical engineer, the iron group elements with atomic numbers 21 through 28 are the most important. The above information shows that elements with unfilled electronic states have permanent dipole moments.

20.5.2 Spin Angular Momentum of an Electron

The electrons possess angular momentum due to their spinning motion. The magnetic spin quantum number m_s gives the direction of spin vectors of an electron. There are two possible directions for spin vectors, namely, upward and downward spins. Therefore, the quantum number m_s takes only two values, namely, ±1/2.

The relation between magnetic moment and spin angular momentum is,

$$\mu_m = -\frac{e}{m}\mu_s \tag{20.20}$$

where μ_s is the angular momentum contributed by the spinning motion of an electron. The possible values of $\mu_s = \frac{1}{2}\frac{h}{2\pi}, -\frac{1}{2}\frac{h}{2\pi}$.

Therefore, from Eq. (20.20), the magnetic moment contributed by the spinning motion is,

$$\mu_m = -\frac{eh}{4\pi m}, +\frac{eh}{4\pi m}$$

$$= -1\ \beta \text{ or } 1\ \beta \tag{20.21}$$

For elements having completely filled electronic states, the magnetic moment contributed by the spin motion is zero. It confirms that a material with unfilled electronic states possesses permanent dipole moment due to spin motion.

20.5.3 Nuclear Magnetic Moment

The magnetic moment of a nucleus is known as *nuclear magneton* and it is given as,

$$\mu_m = \frac{eh}{4\pi M_N} \tag{20.22}$$

The McGraw·Hill Companies

where M_N is the mass of the nucleon. The mass of the nucleon is $M_N = 1.66 \times 10^{-27}$ kg. This mass is nearly 1824 times greater than that of an electron and hence, the nuclear magneton is nearly 10^{-3} times smaller than the electron magnetic moment. Therefore, this value is negligible compared to the magnetic moment of an electron.

20.6 Diamagnetism

Diamagnetic materials possess a weak induced magnetic field when an external magnetic field is applied. The induced magnetic field is produced due to electromagnetic induction. Lenz's law states that the magnetisation (M) will oppose the applied magnetic field (H). The susceptibility of a diamagnetic material is very small and negative. It is usually less than 10^{-5}. Diamagnetism is very weak and it is present in most materials. Examples of diamagnetics are copper, gold and germanium. Let us discuss the theory about diamagnetic materials in this section.

20.6.1 Langevin's Theory of Diamagnetism

The permanent magnetic moment arises if there is any unfilled electronic state in the atoms of an element. When all states are completely filled, then there is no permanent dipole moment. Inert gases such as He, Ne and Ar have completely filled electronic states. However, they possess a small value of induced magnetic moment, when there is an applied magnetic field. This property is known as diamagnetic property. This small value of induced magnetic moment is produced in all elements. Therefore, all elements possess diamagnetic property. The diamagnetic property arises due to Lenz's law operating in atomic scale.

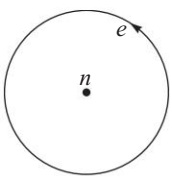


Fig. 20.8 *Atomic view—electron revolving around the nucleus*

In order to derive an expression for susceptibility of a diamagnetic material, consider an electron revolving around a nucleus. It acts like a current loop. The current produced by the revolving electron is proportional to ef, where e is the charge of an electron and f, its frequency. Let ω be the angular frequency of an electron, then $\omega = 2\pi f$. Therefore, the current produced by the revolving electron is,

$$I = -\frac{e\omega}{2\pi} \tag{20.23}$$

The magnetic moment induced by the revolving electron is, $\mu_m = IA$, where A is the area of cross-section. Substituting, the value of I and A ($\approx \pi r^2$), we get,

$$\mu_m = -\frac{1}{2}\frac{e\omega}{\pi}\pi r^2 = -\frac{1}{2}e\omega r^2 \tag{20.24}$$

The electron revolving around a nucleus looks like a current carrying coil of wire. Hence, it induces a magnetic field. Consider that the magnetic field is increased from zero to B. According to Lenz's law, the magnetic field associated with this current opposes the increase in flux. Hence, Lenz,s law is given as,

$$\int \bar{E}.d\bar{l} = -\frac{d\phi}{dt} \tag{20.25}$$

For an electron revolving in a circular orbit, $dl = 2\pi r$ and hence,

$$E = -\frac{1}{2\pi r}\frac{d\phi}{dt} \tag{20.26}$$

Therefore, the magnetic flux density,

$$B = \frac{\phi}{\pi r^2} \tag{20.27}$$

Substituting the value of dϕ from Eq. (20.27) in Eq. (20.26), we get,

$$E = -\frac{r}{2}\frac{dB}{dt} \tag{20.28}$$

The force experienced by an electron due to this electric field is $F = -eE$

i.e., $$F = \frac{er}{2}\frac{dB}{dt} \tag{20.29}$$

The force can be written as the rate of change of momentum,

i.e., $$F = \frac{dp}{dt} = \frac{d}{dt}(mv) = \frac{d}{dt}(mr\omega)$$

where v is linear velocity.

or $$F = mr\frac{d\omega}{dt} \tag{20.30}$$

Comparing Eq. (20.29) and Eq. (20.30), we get,

$$\frac{er}{2}\frac{dB}{dt} = mr\frac{d\omega}{dt}$$

or $$d\omega = \frac{e}{2m}dB \tag{20.31}$$

Integrating the above equation with limitation ω_0 and ω, we get,

$$\int_{\omega_0}^{\omega} d\omega = \frac{e}{2m}\int dB$$

or $$\omega - \omega_0 = \frac{e}{2m}B$$

i.e., $$\omega = \omega_0 + \frac{e}{2m} B \tag{20.32}$$

where ω_0 is the angular frequency of an electron when $B = 0$. When the field is applied, the angular velocity is given by,

$$\omega = \omega_0 + \frac{e}{2m} B = \omega_0 + \omega_L \tag{20.33}$$

where ω_L is called *Larmor angular frequency* and is equal to $eB/2m$. Equation (20.33) shows that the application of a magnetic field changes the angular frequency and hence, the magnetic moment. Substituting ω from Eq. (20.33) in Eq. (20.24), the magnetic moment can be written as,

$$\mu_m = \frac{1}{2} er^2 (\omega_0 + \omega_L)$$

$$= -\frac{1}{2} er^2 \omega_0 - \frac{e^2 r^2}{4m} B \tag{20.34}$$

where the term, $(-1/2)er^2\omega_0$ is the magnetic moment of an electron before the application of the field, and the second term, $-\frac{e^2 r^2}{4m} B$ represents the induced magnetic moment produced due to the application of field. The negative sign shows that the magnetic moment opposes the applied field.

The induced magnetic moment from Eq. (20.34) can be written as,

$$\mu_m \text{ (ind)} = -\frac{e^2 r^2}{4m} B \tag{20.35}$$

Substituting, $B = \mu H$, and for an element with N number of dipoles, Eq. (20.35) can be written as,

$$N\mu_m \text{ (ind)} = -\frac{Ne^2 r^2}{4m} \mu H \tag{20.36}$$

The term, $N\mu_m$(ind) gives the magnetisation induced and hence, Eq. (20.36) can be written as,

$$\chi = \frac{M}{H} = -\frac{Ne^2 r^2}{4m} \mu$$

We know that, $\mu = \mu_0 \mu_r$, hence the above equation can be written as,

$$\chi = -\frac{Ne^2 \mu_0 \mu_r r^2}{4m} \tag{20.37}$$

Equation (20.37) represents the susceptibility of a diamagnetic material. This equation is derived based on the assumption that an electron is revolving around a circular path. Generally, in atoms all orbits are not circular. Therefore, for a spherical symmetrical atom, let r_x, r_y and r_z be the average radii of all electronic orbits, then $r_0^2 = \overline{r_x^2} + \overline{r_y^2} + \overline{r_z^2}$

and $\overline{r_x^2} = \overline{r_y^2} = \overline{r_z^2} = \dfrac{\overline{r_0^2}}{3}$ due to spherical symmetry. Let $\bar{r}$ be the radius of an atom. The average radius of an atom can be written as,

$$\overline{r^2} = \overline{r_x^2} + \overline{r_y^2} = \frac{2}{3}\overline{r_0^2} \tag{20.38}$$

Substituting the value of $\overline{r^2}$ from Eq. (20.38) in Eq. (20.37), we get,

$$\chi = -\frac{Ne^2 \mu_0 \mu_r r_o^2}{6m} \tag{20.39}$$

The diamagnetic susceptibility of most materials vary from 10^{-5} to 10^{-6} and the experimental observations of diamagnetic susceptibility show that it is true. Substituting, $N = 5 \times 10^{28}$ atoms m^{-3}, $e = 1.6 \times 10^{-19}$ C, $m = 9.1 \times 10^{-31}$kg, $r = 1$Å and $\mu_r = 1$, we get,

$$\chi = -\frac{-5\times10^{28} \times \left(1.6\times10^{-19}\right)^2 \times (4\pi \times 10^{-7})(10^{-10})^2}{6\times9.1\times10^{-31}}$$

$$\chi = -2.946 \times 10^{-6} \tag{20.40}$$

In the above calculation, μ_r is taken as unity. The susceptibility is related to the dielectric constant as,

$$\chi_m = \mu_r - 1 \tag{20.41}$$

The value of χ_m varies from 10^{-5} to 10^{-6} and hence $\mu_r \approx 1$. The experimental value of χ for Cu is -9×10^{-6}. From Eq. (20.41), one can infer that the diamagnetic susceptibility is independent of temperature.

20.7 PARAMAGNETISM

We have known that in paramagnetic materials, permanent dipole moment exists even in absence of an external magnetic field. The dipoles try to align themselves on application of an external magnetic field. However, the thermal agitation present in the material disturbs the alignment of dipoles. Therefore, the resultant magnetisation is based on the intensity of applied magnetic field and thermal energy responsible for agitation. The classical theory of Langevin is used to explain the above mechanism, and it gives the necessary expression for susceptibility of paramagnetic materials. The principle behind paramagnetism is explained by the Langevin theory of paramagnetism.

20.7.1 Langevin Theory of Paramagnetism

Consider that a paramagnetic material consists of total n number of permanent magnetic dipoles which are oriented randomly. When an external magnetic field of strength H is applied, the dipoles caused to align themselves in the direction of the field.

The McGraw·Hill Companies

Under equilibrium conditions, the dipoles are inclined at an angle θ as shown in Fig. 20.9. The angle of inclination is proportional to the field.

i.e., $n \alpha \exp[-E/kT]$

$$n = c_o \exp[-E/kT] \tag{20.42}$$

where E is the energy of magnetic dipole, k the Boltzmann's constant, T the temperature and $c_{o,}$ a constant. The dipoles experience torque due to applied external field. The torque which is acting on the dipole is

$$\text{Torque} = m\, H \times QR \tag{20.43}$$

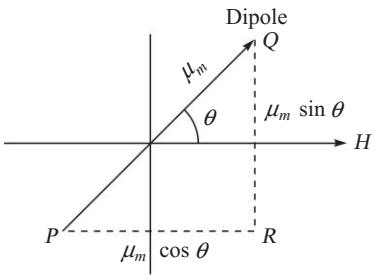


Fig. 20.9 *Alignment of dipole in a magnetic field*

From the Fig. 20.9, $QR = 2l \sin\theta$

Therefore, $\text{Torque} = m\, H\, 2l \sin\theta$

$$= -\mu_m H \sin\theta \tag{20.44}$$

where μ_m is the magnetic dipole moment.
Therefore, energy acquired by the dipole due the torque is

$$E = \int_{\pi/2}^{\theta} \tau\, d\theta \tag{20.45}$$

Substituting the torque value, we get,

$$E = -\int_{\pi/2}^{\theta} \mu_m H \sin\theta\, d\theta$$

Integrating the above equation, we get,

$$E = -\mu_m H \cos\theta \Big|_{\theta}^{\pi/2}$$

$$= \mu_m H \cos\theta \tag{20.46}$$

Substituting E from Eq. (20.46), in Eq. (20.42), we get,

$$n = c_o \exp\left[\frac{\mu_m H \cos\theta}{kT}\right] \tag{20.47}$$

Differentiating the above equation *w.r.t.* θ we get,

$$dn = -c_0 \exp\left[\frac{\mu_m H \cos\theta}{kT}\right]\left[\frac{\mu_m H}{kT}\right]\sin\theta\, d\theta$$

Rearranging the above equation, we get,

$$dn = -c_0 \exp\left[\frac{\mu_m H \sin\theta\, \cos\theta\, d\theta}{kT}\right]\left[\frac{\mu_m H}{kT}\right]$$

or

$$dn = c_0\left[\frac{\mu_m H}{kT}\right]\exp\left[\frac{\mu_m H \sin\theta\, \cos\theta\, d\theta}{kT}\right]$$

or

$$dn = C \exp\left[\frac{\mu_m H \sin\theta\, \cos\theta\, d\theta}{kT}\right] \tag{20.48}$$

where, $$C = \left(\frac{-c_o \mu_m H}{kT}\right)$$

The total dipole moment is obtained by integrating the magnetic dipole moment with limitation 0 and π.
Therefore, the average magnetic dipole moment is equal to

$$\mu_{av} = \frac{\text{Total dipole moment}}{\text{Number of dipoles}} \tag{20.49}$$

Substituting the value, we get,

$$\mu_m = \frac{\int_0^{\pi} \mu_m dn}{\int_0^{\pi} dn} \tag{20.50}$$

From Fig. (20.9), $\mu_m \cos\theta$ and $\mu_m \sin\theta$ are respectively horizontal and vertical components of the magnetic dipole moment. Therefore, the average magnetic dipole moment is

$$\mu_{av} = \frac{\int_0^{\pi} \mu_m \cos\theta\, dn}{\int_0^{\pi} dn} \tag{20.51}$$

Substituting the value of *dn* from Eq. (20.48) in the above equation, we get,

$$\mu_{av} = \frac{\int_0^{\pi} C\, \mu_m \cos\theta \exp\left[\frac{\mu_m H \cos\theta}{kT}\right]\sin\theta\, d\theta}{\int_0^{\pi} C \exp\left[\frac{\mu_m H \cos\theta}{kT}\right]\sin\theta\, d\theta} \tag{20.52}$$

Substituting $\alpha = \frac{\mu_m H}{kT}$, $x = \cos\theta$ and $dx = -\sin\theta\, d\theta$ in the above equation and rearranging the common terms, we get,

$$\mu_{av} = \frac{\mu_m \int_1^{-1} x e^{\alpha x} dx}{\int_1^{-1} e^{\alpha x} dx} = \frac{\mu_m \int_1^{-1} \left(\frac{x}{\alpha}\right) d\left(e^{\alpha x}\right)}{\int_1^{-1} \left(\frac{1}{\alpha}\right) d\left(e^{\alpha x}\right)}$$

$$= \frac{\frac{\mu_m}{\alpha} \int_1^{-1} x d\left(e^{\alpha x}\right)}{\frac{1}{\alpha} \int_1^{-1} d\left(e^{\alpha x}\right)} = \frac{\mu_m [(x e^{\alpha x})_1^{-1} + \int_1^{-1} e^{\alpha x} dx]}{(e^{\alpha x})_1^{-1}}$$

$$= \mu_m \left[\frac{e^{-\alpha} + e^{+\alpha}}{e^{\alpha} - e^{-\alpha}} - \frac{1}{\alpha}\right] = \mu_m \left[\coth \alpha - \frac{1}{\alpha}\right] \qquad (20.53)$$

where $L(\alpha)$ is known as the *Langevin function* and is equal to $\left(\coth \alpha - \frac{1}{\alpha}\right)$. When all N dipoles are aligned in the field direction, then it turns magnetisation M as saturation magnetisation M_s.

Therefore, $$M = M_s\, L(\alpha) \qquad (20.54)$$

or $$\frac{M}{M_s} = \left(\coth \alpha - \frac{1}{\alpha}\right) \qquad (20.55)$$

or $$\frac{M}{M_s} = L(\alpha) \qquad (20.56)$$

Langevin functions $L(\alpha)$ is shown graphically in Fig. 20.10. The curve is known as *Langevin curve*. At high magnetic field and low temperature, the value of $L(\alpha)$ approaches unity due to a larger value of α. The same is evident from Fig. 20.10. This means that all dipoles are oriented along the field direction. Thus, the magnetisation M approaches saturation magnetisation M_s $(= N\mu_m)$. However, in gases, the complete parallel alignment of dipoles is never reached.

Therefore, at high temperatures and moderate value of H, the value of α is small. We know that,

$$\coth \alpha = \frac{1}{\alpha} + \frac{\alpha}{3} + \frac{\alpha^2}{45} + \ldots\ldots \qquad (20.57)$$

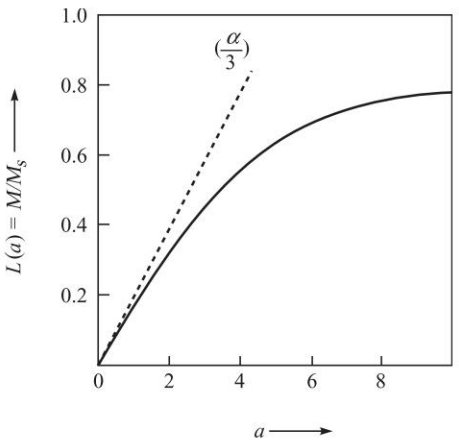


Fig. 20.10 *A plot of normalized magnetisation versus*

For high temperature and for small value of α, Eq. (20.57) can be written as,

$$\cot h\ \alpha = \frac{1}{\alpha} + \frac{\alpha}{3} \tag{20.58}$$

Substituting the above value in Eq. (20.55), we get,

$$L(\alpha) = \cot h\ \alpha - \frac{1}{\alpha} = \frac{\alpha}{3} \tag{20.59}$$

Therefore,
$$\frac{M}{M_s} = \frac{\alpha}{3}$$

or
$$\frac{\left(\frac{M}{M_s}\right)}{\alpha} = \frac{1}{3} \tag{20.60}$$

$$M = M_s \frac{\alpha}{3}$$

Equation (20.60) shows that the slope of Langevin curve is equal to (1/3), when α is small. This means that at the initial part, the curve is linear and coincides with the tangent to the curve at the origin.

We know that paramagnetic susceptibility

$$\chi_p = \frac{M}{H} \tag{20.61}$$

Substituting the value of M, we get,

$$\chi_p = \frac{M_s \alpha}{3H} \tag{20.62}$$

The McGraw·Hill Companies

Substituting the value of α and M_s in the above equation, we get,

$$\chi_p = \left(\frac{N\mu_m}{3H}\right)\left(\frac{\mu_m H}{kT}\right)$$

$$= \frac{N\mu_m^2}{3kT}$$

$$= \frac{N\mu_m^2}{3k}\frac{1}{T}$$

Therefore, $$\chi_p = \frac{C}{T} \tag{20.63}$$

where C is the known as *Curie constant* and is equal to $N\mu_m^2/3k$. The above results support the Curie experimental results on materials like Nd, $CuSO_4.K_2SO_4.6H_2O$ and O_2.

Drawbacks of Langevin's Theory

The following are the drawback of Langevin's theory:

- It fails to explain the relationship between para and ferromagnetism.
- It fails to explain the deviation exhibited in many substances like compounds and cooled gas, solid salts and crystals.

20.7.2 Weiss theory of Paramagnetism

Langevin's theory failed to explain the complicated temperature dependence of susceptibility of paramagnetic materials. In order to explain the same, the internal or molecular field concept of Weiss theory is used. According to this theory, in a real gas the internal or molecular field seen by a gaseous dipole is equal to the sum of applied field and the field due to the contribution from neighbouring gaseous dipoles. The resultant magnetic field (H_i) is

$$H_i = H_a + \gamma M \tag{20.64}$$

where H_a is the applied field; γ, the internal field constant and M, the intensity of magnetisation.

From Langevin's theory, the magnetisation experienced by a paramagnetic material is obtained from susceptibility value,

$$\chi_p = \frac{N\mu_m^2}{3kT}$$

or

$$\frac{M}{H} = \frac{N\mu_m^2}{3k}$$

Therefore, the magnetisation

$$M = \frac{N\mu_m^2 \mu_o H_i}{3kT} \tag{20.65}$$

Substituting the value of H_i from Eq. (20.64) in Eq. (20.65), we get,

$$M = \frac{N\mu_m^2 \mu_o}{3kT}(H_a + \gamma M)$$

or

$$M = \frac{N\mu_m^2 \mu_o}{3kT} H_a + \frac{N\mu_m^2 \mu_o}{3kT} \gamma M \tag{20.66}$$

Rearranging Eq. (20.66), we get,

$$M = \frac{\frac{N\mu_m^2 \mu_o}{3kT} H_a}{(1 - \frac{N\mu_m^2 \mu_o \gamma}{3kT})} \tag{20.67}$$

Substituting, $\frac{N\mu_m^2 \mu_o \gamma}{3k} = \theta$ and C = $\frac{N\mu_m^2 \mu_o}{3k}$ in Eq. (20.67), we get,

$$\chi = \frac{M}{H_a} \tag{20.68}$$

$$\chi = \frac{C}{(T - \theta)} \tag{20.69}$$

where C is known as *Curie constant* and θ, the *paramagnetic Curie point*.

Equation (20.69) is known as *Curie–Weiss law*. This equation shows that the susceptibility is negative below Curie temperature (when $T < \theta$). However, in most of paramagnetic materials, the Curie temperature is very low and hence, the occurrence of a situation such as $T < \theta$ is rare.

20.7.3 Theory of Paramagnetism—A two-level Quantum Mechanical Model

Consider a paramagnetic material consisting of N number of dipoles. Let N_a and N_p be respectively the number of dipoles aligned in antiparallel and parallel, when a magnetic field is applied.

In order to simplify the mathematical calculations, consider there are two dipoles in a system, one with spin up and another with spin down as shown in Fig. 20.11. Let B be the applied magnetic flux density. The energy acquired by the dipole that is aligned parallel to the field due to the application of the field is,

$$E_p = -\mu_m B = -\mu_0 \mu_m H \tag{20.70}$$

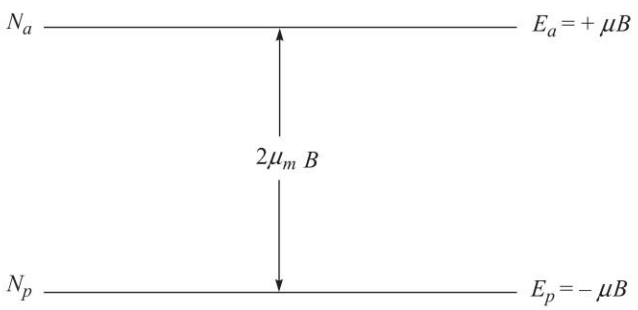


Fig. 20.11 *Two-level quantum mechanical models—Paramagnetic spin system*

The energy of the antiparallel dipole is equal to $E_a = \mu_m B = \mu_0 \mu_m H$. The concentration of dipoles N, N_a and N_p are given by Maxwell–Boltzmann distribution function as,

$$N = N_0 \exp\left[\frac{-E}{kT}\right],$$

$$N_a = N_0 \exp\left[\frac{-E_a}{kT}\right] \text{ and}$$

$$N_p = N_0 \exp\left[\frac{-E_p}{kT}\right] \tag{20.71}$$

The number of dipoles present in the system is

$$N = N_a + N_p \tag{20.72}$$

Substituting the values of N_a and N_p from Eq. (20.71) in Eq. (20.72), we get,

$$N = N_0 \left[\exp\left(\frac{-E_a}{kT}\right) + \exp\left(\frac{-E_p}{kT}\right)\right] \tag{20.73}$$

From Eq.(20.71), Eq. (20.72) and Eq. (20.73), we get

$$\frac{N_p}{N} = \frac{\exp\left(\frac{-E_p}{kT}\right)}{\exp\left(\frac{-E_p}{kT}\right) + \exp\left(\frac{-E_a}{kT}\right)} \tag{20.74}$$

and

$$\frac{N_a}{N} = \frac{\exp\left(\frac{-E_a}{kT}\right)}{\exp\left(\frac{-E_p}{kT}\right) + \exp\left(\frac{-E_a}{kT}\right)} \tag{20.75}$$

Substituting the value of E_p and E_a in Eq. (20.74) and Eq. (20.75), we get,

$$N_p = \frac{N \exp\left(\frac{\mu_m B}{kT}\right)}{\exp\left(\frac{\mu_m B}{kT}\right) + \exp\left(\frac{-\mu_m B}{kT}\right)} \tag{20.76}$$

and

$$N_a = \frac{N \exp\left(\frac{-\mu_m B}{kT}\right)}{\exp\left(\frac{\mu_m B}{kT}\right) + \exp\left(\frac{-\mu_m B}{kT}\right)} \tag{20.77}$$

The total dipole moment per unit volume is,

$$M = (N_p - N_a)\, \mu_m \tag{20.78}$$

Substituting the values of N_a and N_p from Eq. (20.76) and (20.77) in Eq. (20.78), we get,

$$M = N\mu_m \left[\frac{\exp\left(\frac{\mu_m B}{kT}\right) - \exp\left(\frac{-\mu_m B}{kT}\right)}{\exp\left(\frac{\mu_m B}{kT}\right) + \exp\left(\frac{-\mu_m B}{kT}\right)} \right] \tag{20.79}$$

Taking $\alpha = \frac{\mu_m B}{kT}$, Eq. (20.79) can be written as,

$$M = N\,\mu_m \left(\frac{e^{\alpha} - e^{-\alpha}}{e^{\alpha} + e^{-\alpha}} \right) = N\,\mu_m \tan ha$$

or

$$M = N\,\mu_m \tan h\left(\frac{\mu_m B}{kT}\right) \tag{20.80}$$

For smaller value of α, $\tan h\alpha = \alpha$. Therefore, for low value of magnetic field or for high value of temperature, Eq. (20.80) can be reduced to

$$M = \left(\frac{N\mu_m^2 B}{kT}\right) = \frac{N\mu_0\mu_m^2 H}{kT} \tag{20.81}$$

where μ_m is the permanent magnetic moment produced by a dipole and it is in the order of 1 Bohr magneton. Therefore, replacing μ_m by β, Eq. (20.81) can be written as,

$$M = \frac{N\mu_0\beta^2 H}{kT} = \frac{CH}{T}$$

i.e.,

$$\chi = \frac{M}{H} = \frac{C}{T} \tag{20.82}$$

where $C \left(= \frac{N\mu_o\beta^2}{k}\right)$ is the Curie constant and χ the susceptibility of material.

Equation (20.82) gives the susceptibility of a paramagnetic material. Equation (20.82) is obtained from a two-level quantum mechanical model and differs by a factor of 3 from Equation (20.63) obtained by classical theory.

20.7.4 Determination of Paramagnetic Susceptibility of a Solid

The susceptibility of a paramagnetic material is determined based on the principle of force acting on it when subjected to a magnetic field. Let μ_m be the magnetic moment and B the applied field. Then the relation for the force acting on a paramagnetic material is,

$$F = \left(\overline{\mu_m} . \nabla\right) \overline{B} \tag{20.83}$$

The x component of the force is,

$$F = \mu_m \frac{\partial B}{\partial x} \tag{20.84}$$

The magnetisation produced in a magnetic material,

$$M = \chi H \tag{20.85}$$

where H is the magnetic field strength and χ, the susceptibility of material. Let the volume of material is V, then the magnetisation is,

$$M = \frac{\mu_m}{V} \tag{20.86}$$

Substituting the values of M from Eq. (20.85) in the above equation, we get,

$$\frac{\mu_m}{V} = \chi . H$$

or

$$\mu_m = \chi H V \tag{20.87}$$

Using the relation, $B = \mu_o H$ and Eq. (20.87), Eq. (20.84) can be written as,

$$F = \chi H \mu_0 \frac{\partial H}{\partial x} V \tag{20.88}$$

Using the equation, $$\frac{\partial}{\partial x}\left(H^2\right) = 2H \frac{\partial H}{\partial x} \tag{20.89}$$

Equation (20.88) can be written as,

$$F = ½ \chi V \mu_0 \frac{\partial}{\partial x}\left(H^2\right) \tag{20.90}$$

Let χ_2 and χ_1 be the respective susceptibility of a material and surrounding medium (usually air) in which the measurement takes place. Therefore, Eq. (20.90) can be written as,

$$F = \frac{1}{2}(\chi_2 - \chi_1) V \mu_0 \frac{\partial}{\partial x}\left(H^2\right) \tag{20.91}$$

Equation (20.91) gives the force acquired by a paramagnetic material due to the application of field.

(i) Gouy's Method Gouy's method is one of the simple methods used to find the susceptibility of a paramagnetic material. The material under test is taken in the form of a powder or rod. The material in the powder form is taken in a glass tube or in a quartz tube of length equal to 10 to 15 cm. The sample is suspended in such a way that one of its ends lies in between the pole pieces of an electromagnet at the midpoint. The other end is kept away from the field and hence, the magnetic field acting at the other end is negligible or small. The schematic representation of the apparatus used for Guoy's method is shown in Fig. 20.12.

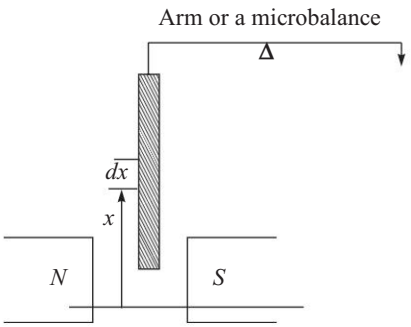


Fig. 20.12 *Gouy's method—susceptibility of a paramagnetic solid*

Let a be the area of cross-section of the sample and *dx* be a small portion in the sample. Then, the volume of sample is *adx*. The susceptibility of the material is measured by applying the magnetic field by an electromagnet. The force acting on the paramagnetic material is,

$$F = \frac{1}{2}(\chi_2 - \chi_1) V \mu_0 \frac{\partial}{\partial x}\left(H^2\right) \tag{20.92}$$

Substituting the value of *V*, we get,

$$F = \frac{1}{2}(\chi_2 - \chi_1) a dx \mu_0 \frac{\partial}{\partial x}\left(H^2\right)$$

or $$F = (\chi_2 - \chi_1) a \mu_0 H dH \tag{20.93}$$

The total force acting on material is determined by integrating the above equation,

i.e., $$F = (\chi_2 - \chi_1) a \mu_0 \int_0^H H dH \tag{20.94}$$

The McGraw·Hill Companies

The above equation is integrated from 0 to H, since the field is maximum (H) at one end and it is minimum ($H = 0$) at the other end of sample under test.

The total force acting on the sample is

$$F = \frac{1}{2}(\chi_2 - \chi_1) a \mu_0 H^2 \tag{20.95}$$

The force acting on the sample is determined by measuring the mass of sample before and after the field is switched on. Let m_1 and m_2 respectively be the mass of sample before and after the field is applied. Then, the downward force acting on material is

$$F = m_2 g - m_1 g \tag{20.96}$$

The susceptibility of the material is determined using the relation,

$$(m_2 - m_1)\, g = \frac{1}{2}(\chi_2 - \chi_1) a \mu_0 H^2 \tag{20.97}$$

The magnetic field H is measured using a search coil, and this experiment is also used to find the susceptibility of material at different temperatures.

(ii) Quincke's Method The susceptibility of a liquid paramagnetic material is determined using Quincke's method. The Quincke's method uses a U-tube as shown in Fig. 20.13. One end of the tube consists of a wide limb, while at the other end, the limb is narrow. The narrow limb portion of the tube is kept in between pole pieces of an electromagnet. The given paramagnetic liquid is taken up to the middle of the electromagnet. This method is based on the principle of the rise in the level of a paramagnetic liquid when it is subjected to a magnetic field.

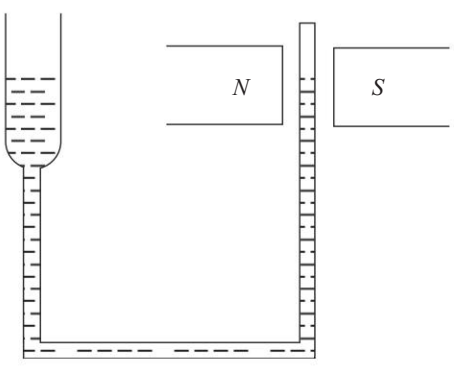


Fig. 20.13 *Quincke's method—paramagnetic susceptibility*

Consider that a magnetic field is applied by energising the electromagnet. The force acting on the liquid is

$$F = \frac{1}{2}(\chi_2 - \chi_1) V \mu_0 \frac{\partial H^2}{\partial x} \tag{20.98}$$

Let a be the area of cross-section of the bore and dx, the length of liquid. Then the volume of liquid is $V = a\,dx$.
Substituting the value of V, we get,

$$F = \frac{1}{2}(\chi_2 - \chi_1)\,a\,dx\,\mu_0\frac{\partial H^2}{\partial x} \tag{20.99}$$

The total vertical force acting on the paramagnetic liquid is

$$F = \frac{1}{2}(\chi_2 - \chi_1)\mu_0 a\int_0^H \frac{\partial}{\partial x}\left(H^2\right)dx$$

or

$$= \frac{1}{2}(\chi_2 - \chi_1)\mu_0 a H_m^2 \tag{20.100}$$

Let ρ and σ be the density of liquid and air respectively; and h, the rise in the level of liquid. The hydrostatic pressure acting on the liquid, when the level of liquid rises is $(\rho - \sigma)\,h\,g$. The corresponding force acting on the liquid is,

$$F = (\rho - \sigma)\,h\,g\,a \tag{20.101}$$

where a is the area of cross-section.

Equating Eq. (20.100) and Eq. (20.101), we get,

$$\frac{1}{2}(\chi_2 - \chi_1)\mu_0 a H_m^2 = (\rho - \sigma)\,h\,g\,a$$

or

$$\frac{1}{2}(\chi_2 - \chi_1)\mu_0 H_m^2 = (\rho - \sigma)\,h\,g \tag{20.102}$$

The rise in the level of liquid h is measured using a traveling microscope.

The diameter of the narrow limb should not be large, since the variation in the level of liquid is not appreciable. When the diameter of the limb is very low, there are some changes in the surface tension of the liquid. Therefore, the diameter of the limb should be of an optimum size.

20.8 FERROMAGNETIC MATERIALS

Materials like Fe, Co, Ni and certain alloys exhibit a high degree of magnetisation. When the magnetic field is absent, these materials exhibit finite value of magnetisation known as *ferromagnetism*. These materials are known as ferromagnetic materials. Ferromagnetic materials exhibit two different properties as a function of temperature. Above a particular temperature known as ferromagnetic Curie temperature (θ_f), they behave as paramagnetic materials. Below the ferromagnetic Curie temperature, the materials behave as ferromagnetic materials and show the hysteresis curve. Thus, ferromagnetic materials posses different properties above and below the ferromagnetic Curie temperature θ_f.

The McGraw·Hill Companies

Case i ($T > \theta_f$)

When the temperature is greater than θ_f, a ferromagnetic material behaves as a paramagnetic material. The material shows similar property as that of paramagnetic materials and an unique relationship between magnetisation and applied field. In this region, susceptibility depends on the Curie–Weiss law,

$$\chi = \frac{c}{T - \theta_f} \text{ (for T>> } \theta_f) \tag{20.103}$$

where C is the Curie constant and θ_f the paramagnetic Curie temperature. The validity of the above equation is analysed by drawing a plot between $(1/\chi)$ and T as shown in Fig. 20.14.

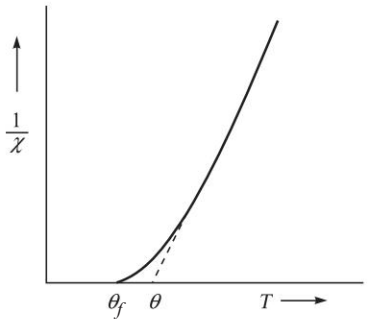


Fig. 20.14 *Reciprocal of susceptibility as a function of temperature—ferromagnetic material*

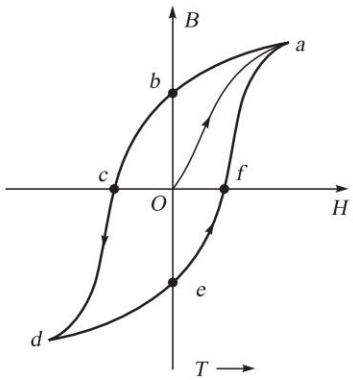
The linear portion of the curve is identified and extrapolated to determine the paramagnetic Curie temperature, θ. The intercept made on the x-axis gives the value of θ. Generally, the paramagnetic Curie temperature θ is greater than the ferromagnetic Curie temperature θ_f. The ferromagnetic and paramagnetic Curie temperatures of a few ferromagnetic materials are shown in Table 20.2 for comparison.

Table 20.2 *Ferromagnetic and paramagnetic Curie temperatures of few ferromagnetic materials*

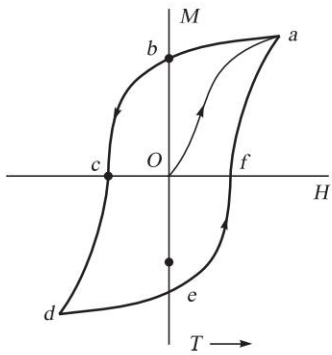
S. No.	*Material*	θ *K*	θ_f *K*
1	Fe	1093	1043
2	Co	1428	1393
3	Ni	650	631

Case ii ($T < \theta_f$)

When the temperature is less than θ_f, materials behaves as ferromagnetic materials and show the well-known hystersis curve. The hystersis curve drawn between B and H, and M and H is shown respectively in Fig. 20.15 a and b.



(a) Magnetic flux density as a function of applied field



(b) Magnetisation as a function of applied field

Fig. 20.15 *Hysteresis curve—ferromagnetic materials*

The magnetic flux density B increases with increase in applied field and it reaches saturation at a level known as B_{sat} beyond which it remains constant even for a further increase in applied field. On the other hand, when the field is decreased, the magnetic flux density decreases through a new path *ab* instead of *oa* as shown in Fig. 20.15a. When the field reaches zero value, the magnetic flux exists in the material and is known as remanent flux density B_r.

When a negative field is applied, the magnetic flux density decreases and it reaches zero value at a field strength of $-H_C$ by taking the path *bc*. The field required to bring the magnetic flux density to zero value, i.e., $-H_C$ is known as *coercive field*. When the field is increased in the negative direction, it results in magnetic flux saturation ($-B_{sat}$), by taking the path *cd*, beyond which the flux density remains constant for a further increase in field. When the field is increased in the positive direction, the flux density takes the path *de* and *ef*, and finally reaches the point *a*.

The path turned by the *B-H* curve forms a loop known as *hysteresis loop as* the shown in Fig. 20.15a. A similar curve is obtained by taking the magnetisation M and the applied field H as shown in Fig. 20.15b. This curve is known as *M-H* curve. From Fig. 20.15b, at zero magnetic field ($H = 0$), the magnetisation exists in materials known as *remanent magnetisation* (M_r). The field required to bring M_r to zero value is known as

coercive field or *coercivity*. The existence of remanent magnetisation at $H = 0$, is known as *spontaneous magnetisation*. The saturation flux density, magnetisation and Bohr magnetisation per atom of few ferromagnetic materials are given in Table 20.3.

Table 20.3 *Saturation values of flux density and magnetisation of ferromagnetic materials*

S. No.	Materials	Bohr magneton per atom	$B_{sat} = \mu_0 M_{sat}$ T	μ_{sat} $\times 10^{-6}$ A m^{-1}
1	Fe	2.22	2.2	1.75
2	Co	1.72	1.82	1.45
3	Ni	0.60	0.64	0.50

20.8.1 Weiss Theory of Ferromagnetism

In order to explain the ferromagnetic properties, Weiss proposed two postulates, namely, internal field and domain concept. Based on the Weiss theory, *internal field* is the field exerted between two interacting atoms in ferromagnetic materials. Therefore, internal field or molecular field produced in a material is,

$$H_i = \gamma M \tag{20.104}$$

where γ is the *internal field constant* or *Weiss molecular field constant* and M, the magnetisation.

The internal-field concept is used to explain most of the properties such as spontaneous magnetisation, Curie law and behaviours of ferromagnetic materials. However, the internal-field concept fails to explain the following points:

a. Based on Weiss theory, the Curie temperature for ferromagnetic and paramagnetic states is same. However, the experimental values show that ferromagnetic and paramagnetic Curie temperatures, i.e., θ_f and θ are different.

b. The classical theory fails to explain the magnetic interaction and interaction energy in a ferromagnetic material. The values of the internal-field constant γ and interaction energy (kT) obtained based on classical theory is about 1000 times less than the actual value. Therefore, the concept of wave nature based on quantum theory has been introduced. Hence, the magnetic interaction and interaction energy are explained in terms of exchange force in wave mechanics.

According to Weiss, the internal field expressed by dipoles is the sum of applied field and the field experienced by the contribution from neighbouring dipoles.

$$H_{ef} = H + \gamma M \tag{20.105}$$

where H_{ef} is the effective field seen by a dipole; H the applied field; γ the internal field constant and M, the magnetisation.

Case (i): High temperature ($T > \theta_f$) A ferromagnetic material behaves as a paramagnetic material when the temperature is greater than the Curie temperature.

In order to explain this property, consider the equation for magnetisation as given in Eq. (20.80),

$$M = N\,\mu_m \tan h\left[\frac{\mu_0 \mu_m H}{kT}\right]$$

where μ_m represents the permanent magnetic moment of a dipole. Taking $\mu_m = 1\ \beta$ (one Bohr magneton), the above equation can be written as,

$$M = N\,\beta \tan h\left[\frac{\mu_0 \beta H}{kT}\right] \tag{20.106}$$

Replacing H by $H + \gamma M$ in Eq. (20.106), we get,

$$M = N\,\beta \tan h\left[\frac{\mu_0 \beta (H + \gamma M)}{kT}\right] \tag{20.107}$$

when T is very high, the value within the square bracket in Eq. (20.107) will become small. For smaller value of x, tan hx is $\approx x$. Therefore, Eq. (20.107) can be written as,

$$M = \left[\frac{N\beta^2 \mu_0 (H + \gamma M)}{kT}\right] \tag{20.108}$$

Rearranging Eq. (20.108), we get,

$$M - \frac{N\beta^2 \mu_0 \gamma M}{kT} = \frac{N\beta^2 \mu_0 H}{kT} \tag{20.109}$$

The value of susceptibility can be written as,

$$\chi = \frac{M}{H} = \frac{(N\beta^2 \mu_o)/kT}{\left(1 - \dfrac{N\beta^2 \mu_o \gamma}{kT}\right)}$$

or

$$\chi = \frac{C}{T - \gamma C}$$

$$\chi = \frac{C}{T - \theta} \tag{20.110}$$

where $C = \dfrac{N\beta^2 \mu_o}{k}$ and $\theta = \gamma C$.

Equation (20.110) gives the susceptibility of a ferromagnetic material, when the temperature is very high. Equation (20.110) is similar to the susceptibility of a paramagnetic material. This shows that ferromagnetic materials behave as paramagnetic materials above a certain temperature known as Curie temperature. The internal field constant for a ferromagnetic material is equal to 10^3. The internal field is due to the wave nature of electrons. According to wave mechanics, the above fields are known as *exchange force*.

The McGraw·Hill Companies

Case (ii): Ferromagnetic Curie Temperature According to the Curie–Weiss law, the susceptibility becomes infinite, when $T = \theta_f$. This shows that the existence of a non-vanishing value of M, even though $H = 0$. The existence of magnetisation even in the absence of a magnetic field is known as spontaneous magnetisation. The spontaneous magnetisation is obtained by substituting, $H = 0$ in Eq. (20.107).

Therefore, $$M = N\beta \tan h \frac{\mu_0 \gamma \beta M}{kT} \tag{20.111}$$

Let $$x = \frac{\mu_0 \gamma \beta M}{kT} \tag{20.112}$$

Substituting the value of x in the above equation, we get,

$$M = N\beta \tan h\, x \tag{20.113}$$

where $N\beta$ represents the total magnetisation produced by all dipoles and it is known as *saturation magnetisation*, M_s. Therefore, Eq. (20.113) can be written as,

$$\frac{M}{N\beta} = \frac{M}{M_s} = \tan hx \tag{20.114}$$

From Eq. (20.112), we get,

$$M = \frac{kT}{\mu_0 \gamma \beta} x \tag{20.115}$$

Substituting the value of M from Eq.(20.115) in Eq. (20.114), we get,

$$\frac{M}{M_s} = \frac{kT}{N\mu_o \gamma \beta^2} x \tag{20.116}$$

where, $M_s = N\beta$

Therefore, $$\frac{M}{M_s} = \frac{T}{\theta} x \tag{20.117}$$

where $$\theta = \gamma C = \frac{N\beta^2 \mu_o \gamma}{k}.$$

A graphical solution for Eq. (20.117) is obtained by drawing tan hx curve between, M / M_s and x as shown in Fig. 20.6. From Eqs.(20.114) and Eq.(20.117), we infer that the value of M / M_s should satisfy these two equations.

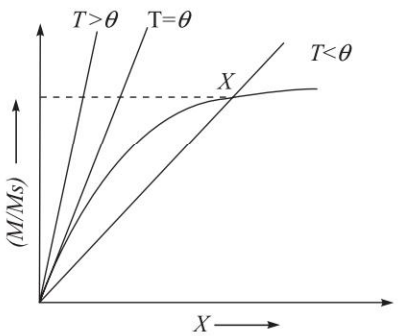


Fig. 20.16 *M / M_s as a function of x*

Using Eq. (20.117), three straight lines are drawn in Fig. 20.16, by taking $T > 0$, $T = 0$ and $T < 0$. Figure 20.16 shows that the line drawn for $T = \theta$ is tangential to the tan hx curve. Similarly, the line drawn for $T > \theta$, has no solution since it does not intersect the tan hx curve while the line drawn for $T < \theta$ intersects the tan hx curve. The value M/M_s should satisfy Eq. (20.114) and Eq. (20.117). The point of intersection of tan hx curve with the $T < \theta$ line gives the value of x. The value of x are obtained by drawing straight lines using Eq. (20.117), for $T < \theta$. Then another curve as shown in Fig. 20.17 is drawn between M / M_s and T / θ.

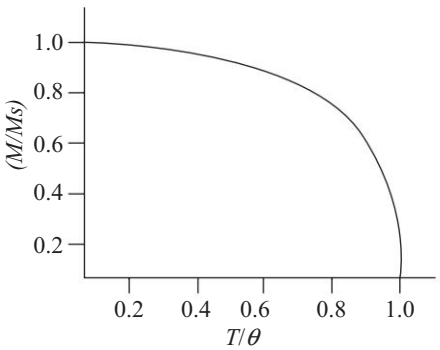


Fig. 20.17 *M / M_s as a function of T / θ*

This curve shows that the magnetisation is nearly equal to the saturation magnetisation when $T >> \theta$ and the magnetisation vanishes when $T < \theta$. The experimental values obtained for Fe, Co and Ni are found to obey Eq. (20.114) and Eq. (20.117). Similarly, the curves obtained for these three elements are similar to Fig. 20.17, even though they have widely different values of θ and M_s. Thus, the experimental evidence confirms the internal field concept of Weiss theory.

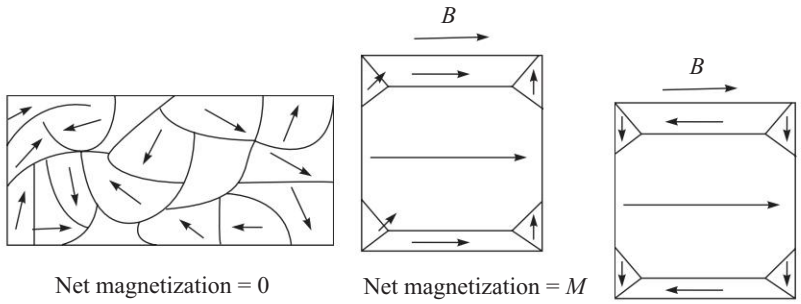
20.8.2 Domain Theory of Ferromagnetism

The magnetic properties of ferromagnetic materials are explained based on the Weiss domain concept. Hence, this theory is known as domain theory of ferromagnetism. The small region within which all spin magnetic moments are aligned in a specific direction is known as *magnetic domain*. The smallest region in which there is an alignment of spin in one direction is known as *ferromagnetic domain*. Thus, the ferromagnetic material consists of number of domains. Generally, the size of the domain will be of the order of 10^{-6}m or larger. Each domain acts as a single magnetic dipole and is oriented in random direction as shown in Fig. 20.18a. Therefore, the net magnetisation is zero, in the absence of a magnetic field.

Each domain is separated from other domains by a wall known as *bloch* or *domain wall*. When an external magnetic field is applied, the domains which are parallel or nearly parallel to the applied field, grow in size at the expense of other domains. The domains which are not parallel to the applied field have a decrease in size. During the absence of a magnetic field, even though the magnetic domains are ordered, the net magnetisation force in a cubic ferromagnetic crystal is zero. The same is shown

The McGraw·Hill Companies

in Fig.20.18b. When the material is placed in an external magnetic field, the central domain grows at the expense of other domains as shown in Fig. 20.18c.



(a) Iron—absence of field *(b) Ferromagnetic crystal —absence of field* *(c) Cubic ferro magnetic crystal— presence of field*

Fig. 20.18 *Ferromagnetic domains*

Domain theory—Experimental Verification

The domain concept was experimentally verified by the magnetic-powder-pattern technique, developed by Bitter. In this technique, the colloidal solution of a magnetic material is sprayed on the surface of the ferromagnetic material under study. We know that a strong magnetic field exists near the boundaries. The colloidal particles are attracted towards well-defined lines and show the boundaries between domains. Further, these domains move depending on the applied external field. A microscope with high resolving power helps to observe the domain walls and their movements.

During the growth of the domain, the internal energy is due to the net contribution of magnetostatic, anisotropy, domain wall and magnetostriction energies. Let us discuss the above energies and their contributions to internal energy briefly in the following sections.

Magnetostatic Energy Consider a single domain in a magnetic material. The orientation of the dipoles and the direction of magnetic lines of force in a single domain are shown in Fig. 20.19a. The potential energy stored inside the magnetic material is known as magnetostatic energy and is very high. This energy is reduced by splitting the domain into two domains as shown in Fig. 20.19b. The second domain consists of two dipoles with anti-spin alignment.

The magnetostatic energy of a domain is further reduced by creating another domain as shown in Fig. 20.19c. During this process, domain walls known as *bloch* are created. A further reduction in magnetostatic energy of the domain is obtained by closing the external field lines as shown in Fig. 20.19d. The domains which are used to close external field lines are called *closure domains*. The potential or magnetostatic energy of a domain is further reduced to a larger value by creating larger number of domains. The domains attain an optimum size at equilibrium state. Generally, domain size is of the order of 1 to 100 μm.

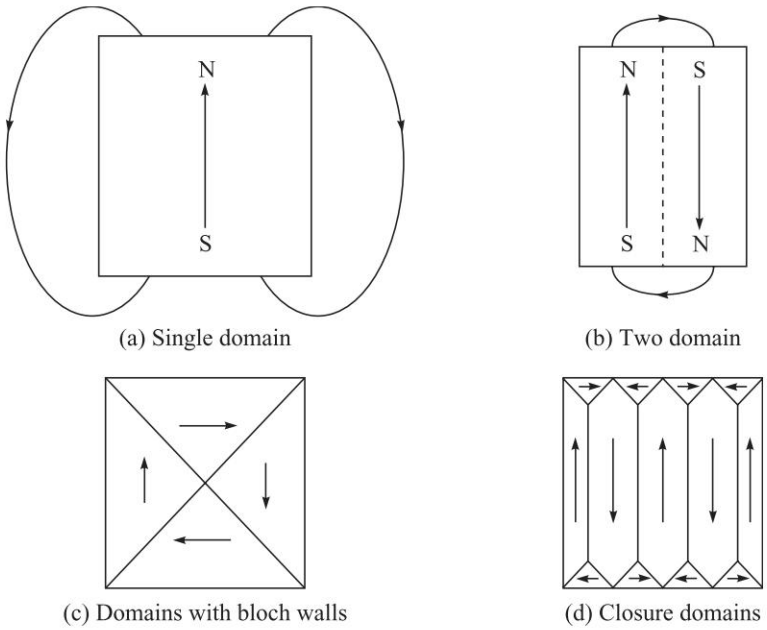


Fig. 20.19 *Domains of magnetic materials*

Antisotrophy Energy We know that crystals like iron consist of different directions, namely [100], [110] and [111]. The magnetic field energy required to magnetise the crystal is a function of crystal direction. This means that the crystal consists of hard and soft directions of magnetisation, and hence, it requires variable energy. For example, in a bcc iron crystal, direction [100] is easy, direction [110] is medium and direction [111] is hard. Therefore, the energy required to magnetise the iron varies depending on direction. The different magnetisation direction of an iron crystal is shown in Fig.20.20.

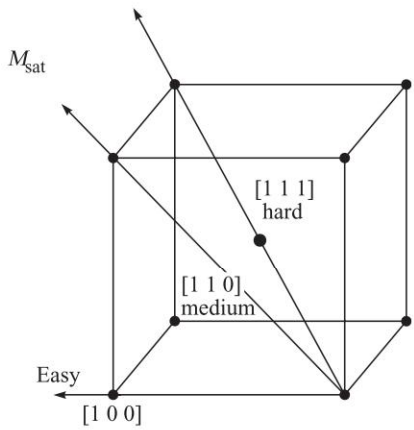


Fig. 20.20 *Anisotropy—iron crystal*

Bloch or Domain wall Consider two domains in magnetic materials. These two domains are separated by a bloch or domain wall. The domains are opposite in direction. The second domain is obtained by rotating the first domain through 180° as shown in Fig. 20.21. The rotation of the domain is carried out gradually due to the existence of exchange force and anisotropy energy. The role of exchange force needed to rotate the dipole which existing between the adjascent atomic spin is very less.

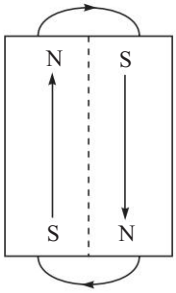


Fig. 20.21 *Magnetic materials—magnetic domains*

The gradual rotation of domains in an iron crystal is shown in Fig. 20.22. The energy required to rotate the domain in an easy direction is less when compared to the hard direction. In order to rotate the domain through an angle of 180^{o}, anisotropy energy requires a domain-wall thickness of nearly $1A^{0}$, while the exchange energy requires a larger domain wall thickness. However, a minimum wall thickness is required with minimum potential energy at equilibrium thickness condition. Therefore, the minimum potential energy of the domain wall is known as *domain-wall energy*. Generally, the thickness of a domain is of the order of hundreds of Å. For example, the domain-wall thickness of iron is equal to 0.1 μm.

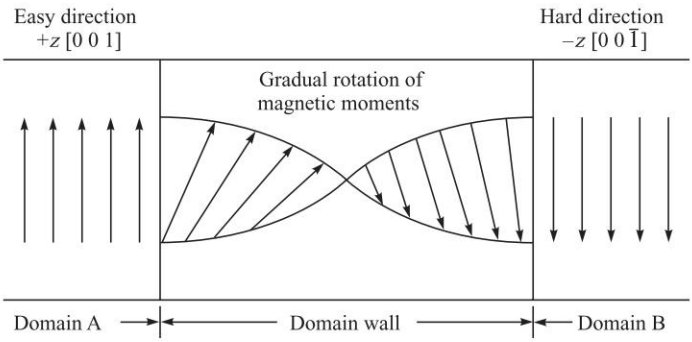


Fig. 20.22 *Gradual rotation of domain—iron crystal*

Magnetostriction Energy Consider a ferromagnetic material subjected to a magnetic field. There is an increase in domain size which is parallel to the field, while the antiparallel domains decrease in size. As a result, realignment of dipoles takes place. Thus, the material length increases in an easy direction due to reorientation of dipoles because of the application of an external magnetic field. The change in length of a ferromagnetic material due to the influence of a magnetic field is shown in Fig. 20. 23.

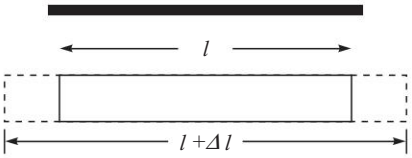


Fig. 20.23 *Magnetostriction—ferromagnetic materials*

When the magnetic field is applied in the hard direction, the material will experience a decrease in length. Let l be the original length of a ferromagnetic material and Δl be the change in length. Therefore, the ratio of change in length to original length is known as *magnetostriction constant* (λ).

$$\lambda = \frac{\Delta l}{l} \tag{20.118}$$

The magnetostriction constant reaches saturation (λ_{sat}) when the field reaches saturation. This condition is known as *saturation strain* and its value ranges from 10^{-6} to 10^{-5}. The energy associated with magnetostriction is called magnetostriction energy. The noise in transformers is produced by the magnetostriction energy.

Hysteresis Curve We know that ferromagnetic materials do not posses net magnetisation due to random orientation of magnetic domains. On the other hand, when an external field is applied, it acquires net magnetisation due to growth of magnetic dipoles in the field direction. The variation in magnetisation due to the application of an external field H is shown in Fig. 20.24. The magnetisation exhibits a curve known as hysteresis curve. When the material is at the starting point O, the domain structure is shown in Fig. 20.24. The variations in magnetisation is shown as points *a* to *b* on the curve due to increase in field strength. Beyond that, the magnetic domains try to rotate along the direction as shown in Fig. 20.24. At point *b*, magnetisation reaches a maximum and remains constant even for an increase in magnetic field strength H. When the domain walls are rotated, it gives out some fluctuations in domains due to irregular rotations. These fluctuations are known as *Barkhausen effect*. The increase in domain takes place up to *b*, beyond which there is no increase in size of domain. At b, the domain walls are in the direction of the field. The magnetisation at *b* is called *spontaneous magnetisation.* The magnetisation reaches its maximum value known as *saturation magnetisation*, M_s.

When the magnetic field is reduced gradually, the magnetisation decreases in the path *bc* and not in *ba* as shown in Fig. 20.24. It means that magnetisation M lags behind the applied field H. When the field is reduced to zero, the magnetisation has a non-zero magnetisation known as *remanent magnetisation.* The remanent magnetisation or *retentivity* is used to measure the permanent magnet strength of a magnet of ferromagnetic material. The magnetic domain-wall size reduces and allows the growth of other domains when the field is reduced.

In order to bring back the material to its original condition, the magnetic field is applied in the reverse direction. The magnetisation decreases from M_r and reaches zero value when the field reaches at $B = -B_C$. The field B_C is known as *coercive field* or *coercivity* which is the field required to bring the magnetisation to zero value. When the field is increased beyond, the material reaches reverse magnetic saturation M_s.

The McGraw·Hill Companies

The magnetic domain reduces in size and almost takes the original shape when the magnetic field reaches the point *i* as shown in Fig. 20.25. When the magnetic field is increased beyond, the magnetisation curve is closed. The closed curve is called *hysteresis loop*. When a ferromagnetic material undergoes hysteresis curve, there is an energy loss due to the rotation of domains. These losses are more for materials with more impurities and imperfections. The energy loss is the measure area of the hysteresis loop.

In a ferromagnetic material, the saturation magnetisation M_s depends on temperature. The temperature dependence of M_s is shown in Fig. 20.26. When the temperature is increased from 0 K, the saturation magnetisation decreases and it reaches zero value at a temperature known as Curie temperature, as shown in Fig. 20.26. The Curie temperature and saturation magnetisation temperature M_s of a few ferromagnetic materials are shown in Table 20.4.

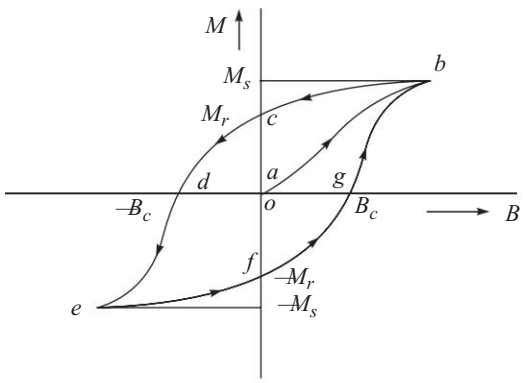


Fig. 20.24 *Ferromagnetic material—hysteresis curve*

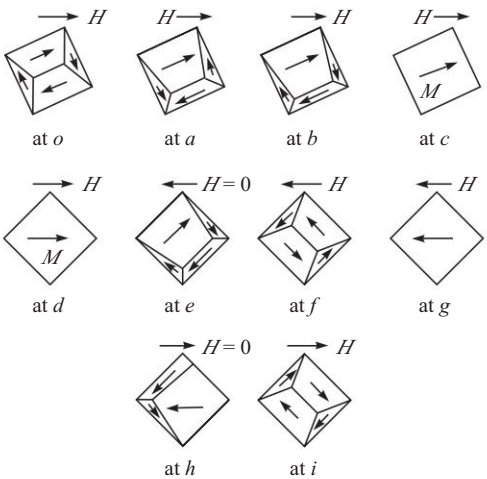


Fig. 20.25 *Ferromagnetic material—domain walls*

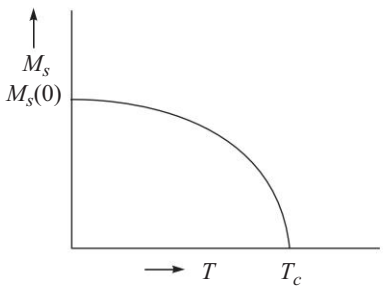


Fig. 20.26 *Magnetisation as a function of temperature*

Table 20.4 *Curie temperature and saturation magnetisation temperature of ferromagnetic materials*

S. No.	*Materials*	*Saturation magnetisation at 0 K* $\times 10^5$ A m^{-1}	*Curie Temperature* T_c *K*
1	Fe	1.38	1043
2	Co	1.15	1388
3	Ni	0.41	627

20.8.3 Heisenberg's Exchange Interaction

The parallel and antiparallel arrangements of magnetic spin respectively in ferromagnetic and antiferromagnetic materials are explained based on exchange interaction energy (E_{ex}). In order to explain the above concepts, consider the electronic configuration of an iron atom. The electric configuration of iron Fe^{26} is $1s^2$, $2s^2$, $2p^6$, $3s^2$, $3p^6$, $3d^6$, $4s^2$. The 4s subshell is filled with two electrons while the 3d subshell is unfilled as shown in Fig. 20.27. The 3d subshell consists of five orbitals out of which 1 orbital is completely filled. The available four unpaired electrons are aligned parallel in a separate orbital based on both Pauli's exclusion principle and electrostatic energy and not based on magnetic interaction. Thus, electrons will have the same spin moment M_s and a different orbital moment M_l for each orbital. The spin magnetic moment is mainly responsible for magnetisation in ferromagnetic materials. Applying quantum concept, an exchange interaction is constituted out of the combination of Pauli's exclusion principle and electrostatic interaction energy, to explain a large value of the internal field in ferromagnetic materials. The exchange interaction is known as *exchange field* or *molecular field* or *Weiss field.*

The McGraw·Hill Companies

Consider that S_i and S_j are the spin angular moments of i^{th} and j^{th} electrons. According to quantum mechanical concepts, exchange energy between any two electrons is

$$E_{ex} = -2 J_{ex} S_i S_j \tag{20.119}$$

where J_{ex} is known as the *exchange integral*. The negative sign indicates that exchange interaction energy depends on interatomic distance between two interacting atoms. Equation (20.119) is used to explain parallel and antiparallel alignment of spins in ferro and antiferromagnetic materials. For most solids, the value of J_{ex} and E_{ex}, are negative, i.e., antiparallel. Therefore, spin moment S_i and S_j are antiparallel as exist in antiferromagnetic materials. On the other hand, the value of J_{ex} is positive, while E_{ex} is negative in case of ferromagnetic materials such as Fe and Co. Hence, the value of S_i and S_j are parallel and it is confirmed that they are ferromagnetic materials. Thus, the sign of an exchange integral is used to know the status of ferro and antiferromagnetic materials.

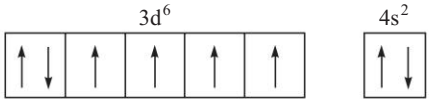


Fig. 20.27 *Iron atoms—electrons in 3d subshell*

The same can be explained by considering the exchange integral value as a function of ratio of r/r_d where r is interatomic distance and r_d, the radius of a 3d oribital. The ratio of r/r_d for different atoms is shown in Table 20.5. A graph is drawn between J_{ex} and r/rd as shown in Fig. 20.28, to explain the state of magnetic materials. For atoms like Fe, Co, Ni and Gd, the value of r/r_d is higher than 3, and hence, takes positive exchange integral values. Similarly, for atoms like Mn and Cr, the r/r_d value is less than 3 and hence, takes negative exchange integral values. Therefore, it is clear that when the value of r/r_d is higher than 3, materials are called ferromagnetic materials. When the value of r/r_d is less than 3, materials are known as antiferromagnetic materials.

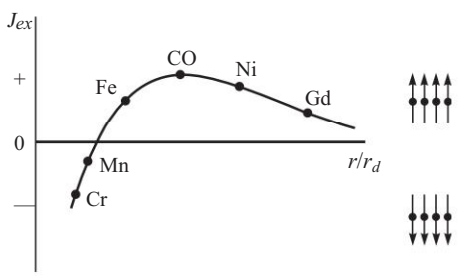


Fig. 20.28 *Exchange integral as a function of* r/r_d

Table 20.5 *Value of r/r_d and materials status*

S. No.	*Metals*	*r/r_d*	*Materials status*
1	Cr	2.60	Antiferromagnet
2	Mn	2.94	Antiferromagnet
3	Fe	3.26	Ferromagnet
4	Co	2.64	Antiferromagnet
5	Ni	2.94	Antiferromagnet
6	Cd	3.1	Ferromangnet

20.9 ANTIFERROMAGNETIC MATERIALS

In antiferromagnetic materials, adjacent dipoles are aligned antiparallel to each other. The antiparallel alignment of dipoles is produced due to the exchange interactions similar to that of a ferromagnetic material, when the distance between any two adjacent dipoles is very small. The antiparallel alignment can be explained considering a body centered cubic unit cell produced by two intersecting simple cubic unit cells of different kinds of atoms *A* and *B*. The antiparallel alignment of dipoles is contributed by the atom *A* having dipole orientation in one direction, and the atom *B* having orientation of dipoles in another direction. Antiferromagnetism was first observed in MnO and then a large number of compounds were found to have antiparallel alignment.

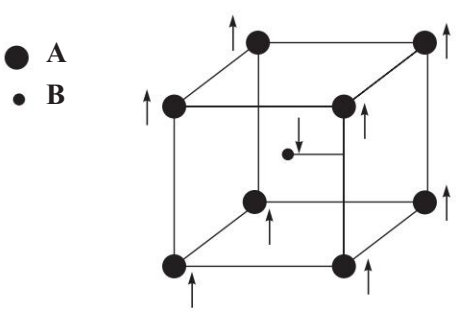


Fig. 20.29 *Antiparallel alignment of dipoles—two intersecting unit cells*

When a curve is drawn between the susceptibility and temperature, a curve as shown in Fig.20.29 is obtained for an antiferromagnetic material. The temperature corresponding to the maximum value of susceptibility of a material is known as *Neel temperature.*

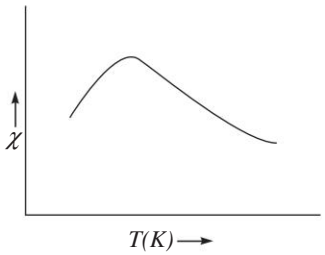


Fig. 20.30 *Susceptibility as a function of temperature—Antiferromagnetic materials*

The susceptibility of antiferromagnetic materials is found to obey the following equation above the Neel temperature,

$$\chi = \frac{C}{T+\theta} \quad \text{when } T > T_N \tag{20.120}$$

where C, T and θ are respectively the Curie constant, absolute temperature and paramagnetic Curie temperature.

The behaviour of an antiferromagnetic material above the Neel temperature is explained using the internal field concepts. The internal field produced is

$$H_a = H - \gamma M_b \tag{20.121}$$

$$H_b = H - \gamma M_a \tag{20.122}$$

where H_a and H_b are the internal fields produced respectively by the interactions of atoms A and B. M_a and M_b are the magnetisations respectively for atoms A and B. The negative sign shows that the interaction produces antiparallel alignment of dipoles.

Substituting the values of H_a and H_b in the Eq.(20.106), we get,

$$M_a = N\beta \tan h \frac{\mu_0 \beta}{kT}(H - \gamma M_b) \tag{20.123}$$

and

$$M_b = N\beta \tan h \frac{\mu_0 \beta}{kT}(H - \gamma M_a) \tag{20.124}$$

The total magnetisation is given by $M = M_a + M_b$. Therefore, the total magnetisation is

$$M = N\beta \tan h \left[\frac{\mu_0 \beta}{kT}(2H - \gamma(M_a + M_b))\right] \tag{20.125}$$

or

$$= N\beta \tan h \left[\frac{\mu_0 \beta}{kT}(2H - \gamma M)\right] \tag{20.126}$$

For smaller values of x, tan $hx = x$. Therefore, Eq. (20.126) can be written as,

$$M = \frac{N\beta^2 \mu_0}{kT}(2H - \gamma M) \tag{20.127}$$

Rearranging Eq. (20.127), we get,

$$\chi = \frac{M}{H} = \frac{2N\beta^2\mu_0 / kT}{1 + \frac{N\gamma\beta^2\mu_0}{kT}} \tag{20.128}$$

Substituting, $C = \frac{N\beta^2\mu_0}{k}$ and $\theta = \gamma C$ in Eq. (20.128), we get,

$$\chi = \frac{M}{H} = \frac{2C}{T + \theta} \tag{20.129}$$

Equation (20.129) gives the value of susceptibility of an antiferromagnetic material, when the temperature is greater than Neel temperature. Equation (20.129) is obtained by multiplying Eq. (20.120) by a factor of 2. A rigorous mathematical treatment gives the correct value of susceptibility. The above simple treatment explains the antiferromagnetic behaviour, and there is no need for any rigorous treatment.

The existence of Neel temperature is explained based on the following arguments. Consider that the dipoles of an antiferromagnetic material are nearly at 0 K. The dipoles are perfectly oriented in an antiparallel manner. Hence, the magnetisation produced by individual dipoles is very high. Due to the antiparallel alignments, the magnetisation produced by one dipole is cancelled by the other one. Consider that the temperature is slightly increased from 0 K, the magnetisation decreases gradually and it becomes zero at the Neel temperature. This shows that at the Neel temperature, the spontaneous magnetisation vanishes. Therefore, above the Neel temperature there is no spontaneous magnetisation.

Equations (20.123) and (20.124) are valid even in low temperature. By taking $\tan hx \approx x$, it can be written as,

$$M_a = N\beta\left[\frac{\mu_0\beta}{kT}(H - \gamma M_b)\right] \tag{20.130}$$

and

$$M_b = N\beta\left[\frac{\mu_0\beta}{kT}(H - \gamma M_a)\right] \tag{20.131}$$

Even though the spontaneous magnetisation is zero, the value of $H = 0$, is still valid. Substituting, $H = 0$ in Eqs. (20.130) and (20.131), we get,

$$M_a = \frac{-N\beta^2\mu_0\gamma M_b}{kT} \tag{20.132}$$

and

$$M_b = \frac{-N\beta^2\mu_0\gamma M_a}{kT} \tag{20.133}$$

The above equations can be written as,

$$M_a + \frac{\gamma C}{T} M_b = 0 \tag{20.134}$$

The McGraw·Hill Companies

and $$M_b + \frac{\gamma C}{T} M_a = 0 \qquad (20.135)$$

Above the Neel temperature, Eq. (20.134) and (20.135) has nontrivial solution, since there is no spontaneous magnetisation present above the Neel temperature. Let us consider that spontaneous magnetisation has set in at $T = T_N$. Then the equation has a trivial solution. Assuming that there is a spontaneous magnetisation at $T = T_N$, the values of M_a and M_b are equal to zero.

Therefore,

$$\left(\frac{\gamma C}{T_N}\right)^2 - 1 = 0 \text{ or } T_N = \gamma C = \theta \qquad (20.136)$$

Equation (20.136) shows that $T_N = \theta$, and the spontaneous magnetisation is observed at $T=T_N$. But most of antiferromagnetic materials have different values of T_N and θ. For example, FeO has $T_N = 200$ K and $\theta = 570$ K, and MnO_2 has $T_N = 84$ K and $\theta = 316$ K. This discrepancy in explaining the Neel temperature is due to our crude model used to explain the properties of antiferromagnetic materials. Another reason for the discrepancy in the present result is due to the assumption of the crystal structure. In this model, it is assumed that antiferromagnetic materials possess body centered cubic crystal structure. But certain ferromagnetic materials have different crystal structures.

20.10 FERRIMAGNETIC MATERIALS

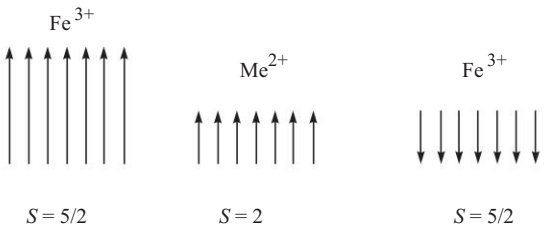
Ferrimagnetic materials, also known as *ferrites*, are a special class of ferromagnetic materials. Ferrites and ferromagnetic materials exhibit similar properties except for the alignments of spin magnetic moments. In a ferromagnetic material, spin magnetic moments are aligned in the same direction. However, in ferrites, spin magnetic moments are not in the same direction; rather, a few of them will be aligned in opposite directions. Therefore, ferrites have net spin magnetic moments. This special characteristic of ferrites makes them different from ferromagnetic materials and hence, they can be used for high frequency purposes and design of special magneticdevice applications.

20.10.1 Structure of Ferrites

The general chemical formula for ferrites is $Me^{2+}Fe_2^{3+}O_4^{2-}$, where Me^{2+} represents divalent metallic ions such as Zn^{2+}, Cd^{2+}, Fe^{2+}, $Mn^{2+}$and Mg^{2+}. Examples of ferrites are manganese ferrite, cadmium ferrite and zinc ferrite. The ferrites crystallise in a cubic structure. A ferrite unit cell consists of one ferrite molecule at each corner and hence, there is a total of 8 ferrous molecules in a ferrite unit cell. Therefore, a ferrous ferrite unit cell consists, of 32 oxygen (O^{2-}) ions, 16 Fe^{3+} ions and 8 Me^{2+} ions. A ferrite unit cell contains both octahedral sites and tetrahedral sites. The 8 Fe^{2+} ions and 8 Fe^{3+} ions occupy the octahedral sites, in which each ion is surrounded by six O^{2-} ions. The balance 8 Fe^{3+} ions occupy the tetrahedral sites and are surrounded by four O^{2-} ions. The total number spin in octahedral site *B* is higher than the tetrahedral

site *A*. The spin of 16 ions in the octahedral sites is aligned parallel while 8 ions in the tetrahedral sites are antiparallel to octahedral sites. If one considers the presence of oxygen ions, then the structure is a close packed face centered cubic structure. The arrangement of spin in a ferrite is shown in Fig. 20.31.

The electronic configuration, spin alignment in orbital, magnetic spin and magnetic moment of a few metal ions are given in Table 20.6.



(a) Octahedral sites (b) Tetrahedral sites

Fig. 20.31 *Spin arrangement—ferrite*

Ferrites takes two different types of structure, namely, inverse spinal and regular spinal. The brief details of these structures are given below.

(i) Inverse Spinal Consider a single ferrous ferrite cell which consist of Fe^{2+} and Fe^{3+} ions in the octahedral *B* site, and Fe^{3+} ions in the tetrahedral *A* sites as shown in Fig. 20.32. Half of the *B* site is occupied by ferrous ions, while the remaining half is occupied by ferric ions. The *A* site is occupied by balance ferric ions. The value of each Fe^{2+} ion corresponds to a magnetic dipole moment of 4β. Similarly, for Fe^{3+} ions, the magnetic dipole moment is 5β (the dipole moment produced by an unpaired electrons is equal to one Bohr magneton, i.e., $\mu_B = 1\beta$). The dipole moment of *B*-site Fe^{3+} ions, i.e., 5β cancels *A*-site Fe^{3+} ions (5β) and hence, the net dipole moment is 4β. This structural arrangement is known as inverse spinal. This leads to spontaneous magnetisation of ferrites.

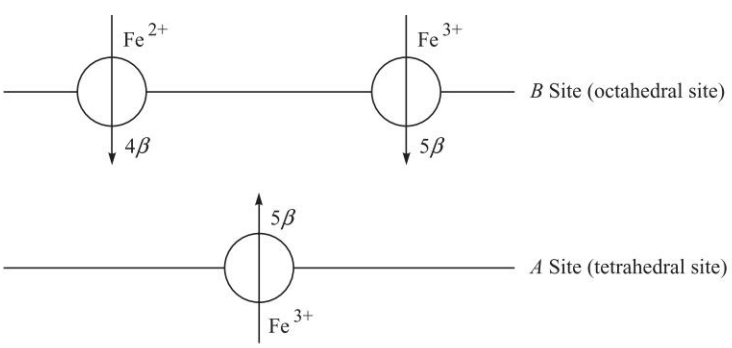
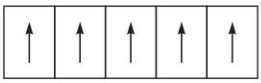
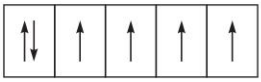
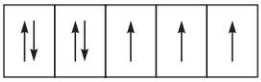
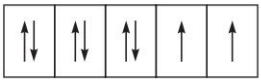
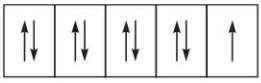


Fig. 20.32 *Ferrites—inverse spinal*

Table 20.6 *Magnetic moment and spin of metal ions*

S. No	Ions	Electric configuration inner core	Orbital spin alignment	S	Magnetic moment
1	Mn^{2+}	$2d^5$		5/2	$5\mu_B$
2	Fe^{2+}	$3d^6$		2	$4\mu_B$
3	Co^{2+}	$3d^7$		3/2	$3\mu_B$
4	Ni^{2+}	$3d^8$		1	$2\mu_B$
5	Cu^{2+}	$3d^9$		1/2	$1\mu_B$

(ii) Regular Spinal The general chemical formula for a regular spinal is Me^{2+} Fe_2^{3+} O_4^{2-} where Me^{2+} is a divalent non-magnetic ion like Zn^{2+}, Cd^{2+}, Mg^{2+}, etc. In the regular spinal structure, the divalent ions occupy tetrahedral *A*-site, while the trivalent ions occupy octahedral *B*-sites as shown in Fig. 20.33.

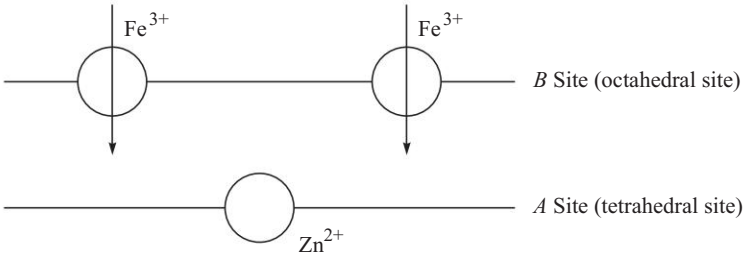


Fig. 20.33 *Ferrites—regular spinal*

Consider a zinc-ferrite molecule, i.e., Zn^{2+} $Fe_2^{3+}O_4^{2-}$. It consists of one ferrous ion (Zn^{2+}) and two ferric ions (Fe_2^{3+}). The two ferric ions occupy the octahedral *B* site, while ferrous ions occupy the tetrahedral *A*-site. The dipole moment of the *A*-site is 4β since a ferrous ion has four unpaired electrons. The dipole moment of the *B*-site is equal to $2 \times 5\ \beta = 10\beta$, since the ferric ion has five unpaired electrons. Therefore, the total dipole moment is 14β, when the ferrous and ferric ions are in parallel alignment. On the other hand, if one ferrous and ferric ion are in one direction, while the other ferric ions are in one direction, then the total dipole moment is 4β. The present results are in good agreement with experimental ones. Thus, in ferrous molecules, the dipoles are at antiparallel alignment.

20.10.2 Hystersis Curve

A ferrimagnetic material exhibits hysterisis curve and is in the form of a square as shown in Fig. 20.34.

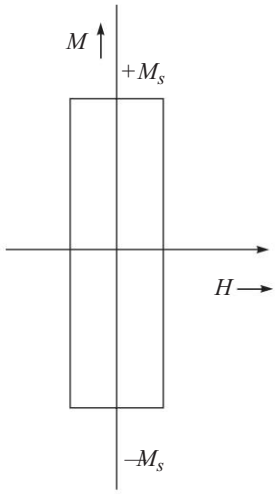


Fig. 20.34 *Hystersis curve—ferrite*

It is clear from the hysteretic curve that the ferrite exhibits in any one of the states, namely, $+M_s$ or $-M_s$. The above two states of ferrites are similar to that of binary digits 0 and 1. For a particular value of applied field, the reversal in magnetisation occurs and hence, it is more suitable for digital memory devices in computers.

20.10.3 Applications

Ferrites find versatile applications in industries due their characteristic features. Ferrites are classified into two types, namely, soft and hard ferrites. The following are some of the applications of ferrites.

- *Soft ferrites* are used in high-frequency transformer cores and computer memories. The devices which are based on soft ferrites are computer hard disks, floppy disks, credit cards, audio/video cassettes, recorder heads, etc.
- Microwave devices like isolators and circulator phase shifters are prepared employing ferrites.
- *Hard ferrites* are used to prepare permanent magnets for applications such as loudspeakers and wiper motors.
- Ferrites are used to generate lowfrequency ultrasonic waves using the magnetostriction principle.
- They are used to produce cores for lowpower transformers and high-flux transformers.
- In computer memory systems, non-volatile memories are made using ferrites. These memories retain the data stored in them even when the power is cut off.

The McGraw·Hill Companies

- Ferrite as a thin film is used in magnetic bubble memory. It is a non-volatile memory.
- Small antennas are made employing ferrite rods by winding with a coil of suitable material.

20.11 HARD AND SOFT MAGNETIC MATERIALS

Ferromagnetic materials are classified into two types, namely, soft and hard magnetic materials, based on the characteristic parameters such as hystersis and magnetisation.

20.11.1 Soft Magnetic Materials

The name implies that soft magnetic materials are easy to magnetise and demagnetise. When a small amount of magnetic field is applied, it results in a large magnetisation due to easy movements of magnetic domains. The degrees of magnetisation and demagnetisation of these materials are very high even for a small applied field. In order to prepare soft magnetic materials, the pure material is first heated to the required high temperature. When the temperature is high, the atoms are can move freely in the molten state and are allowed to settle in an ordered lattice during slow cooling. The resultant material is a soft magnetic material. The potential applications of soft magnetic materials are based on their characteristic properties.

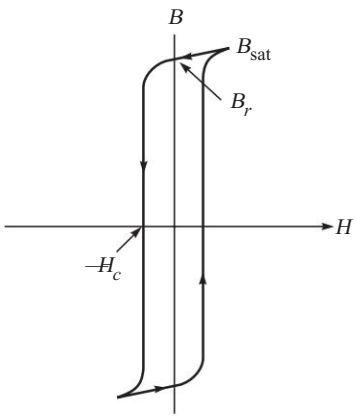


Fig. 20.35 *Hysteresis loop—soft ferrities*

Properties

The following are the properties of a soft magnetic material:

(i) It is easy to magnetise and demagnetise.
(ii) Its hysteresis loop is very thin and long as shown in Fig. 20.35.
(iii) The hysteresis loop area is very less and hence, the hysteresis loss is very small.
(iv) The coercive field is very low while the saturation magnetisation is very high.
(v) The susceptibility and permeability values are very high.

(vi) Its resistivity is very high and hence, eddy-current loss is very low.
(vii) The magnetostatic energy is very small since such materials are free from irregularities like strain.

Examples

(i) Iron–nickel–aluminum alloys with a small dopant cobalt known as FeAl-NiCo alloy
(ii) Iron–silicon alloys
(iv) Iron–cobalt–mangann alloys
(v) Copper–nickel–iron alloys

Applications

The following are the applications of soft magnetic materials.

- They are used in transformer cores, motors, relays and sensors.
- Iron–silicon alloy magnets are used in electrical equipment.
- Silicon steel magnets are used in alternators and high-frequency rotating materials.
- Soft magnetic materials are used in storage components and microwave isolators.

20.11.2 Hard Magnetic Materials

Hard magnetic materials are those which are difficult to magnetise and demagnetise. These materials require high magnetic field for both magnetisation and demagnetisation. It is difficult to rotate the magnetic domains due to the impurities and crystal defects existing in such materials. In order to make the materials hard magnetic, the molten mixture which is at a high temperature is rapidly quenched in cold liquid. In addition, impurities are added to the base materials to make them hard. The hard magnetic materials are used where a permanent and high magnetic field is required.

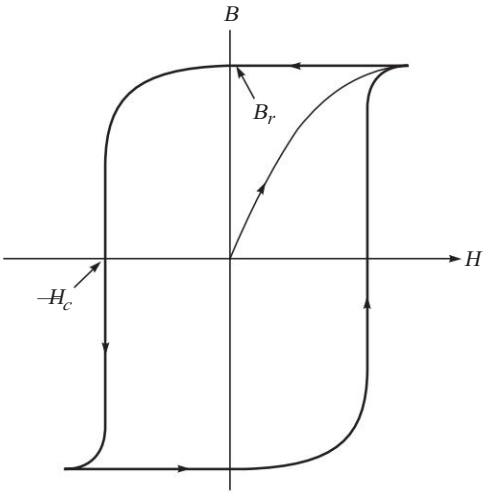


Fig. 20.36 *Hysteresis loop—hard ferities*

The McGraw·Hill Companies

Properties

The following are the properties of hard magnetic materials:

(i) They are hard to magnetise and demagnetise.
(ii) The hysteresis loop is very large and hence, hysteresis loss is heavy due to larger hysteresis area. The characteristics of the hysteresis loop is shown in Fig.20.36.
(iii) It is very difficult to rotate the domain walls due to impurities and crystal imperfections.
(iv) The susceptibility and permeability values are low.
(v) The values of coercivity and resistivity are larger.
(vi) The value of eddy current loss is very high.
(vii) The mechanical strain is more due to presence of impurities and crystal imperfections. Therefore, the magnetostatic energy is very large.

Examples

AlNiCo, rare earth metal alloys with Mn, Fe, Co, Ni, carbon steels and tungsten steels.

Applications

Applications of hard magnetic materials are as follows:

(i) They are used for production of permanent magnetism.
(ii) The magnetism in toys, compass needles, meters, etc., are made from carbon steel.
(iii) The magnets in dc meters and measuring devices are made from tungsten steel.
iv) Neodymium magnets are used in microphones instead of conventional magnets mainly to reduce the microphone size.
v) Cast AlNiCo magnets are used in speed meters, and sensors in automobiles, motors, etc.

20.12 ENERGY PRODUCT OF MAGNETIC MATERIALS

Permanent magnets require hard magnetic materials with large saturation magnetisation and large coercivity. The energy stored in hard magnetic materials is larger due to their large hysteresis loop. The amount of energy stored in hard magnetic materials is measured from the maximum area of the rectangle which fits in the second quadrant of the *B-H* curve. The area of the rectangle *CDEF* of the *B-H* curve shown in Fig. 20.37 gives the energy stored in a magnetic material.

Consider that B_m, H_m and V_m are the flux density, magnetic field and volume of magnetic materials respectively. From Fig. 20.37, energy stored in magnetic materials is

$$E_m = B_m H_m V_m$$

or

$$E_m = (BH)_{max} V_m \qquad (20.137)$$

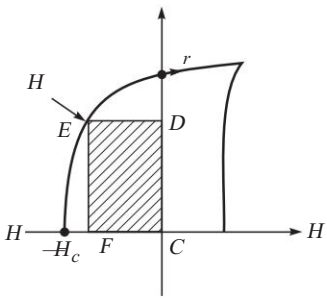


Fig. 20.37 *Hard magnetic materials—energy products*

where $(BH)_{max}$ is the maximum area of the quadrant in the *B-H* curve and is equal to maximum energy stored in magnetic materials. Therefore, the product of *B-H* is known as energy product of magnetic materials.

The energy stored in a magnetic material is determined from the energy density measurements. In order to determine the energy density, consider a permanent magnet as shown in Fig. 20.38.

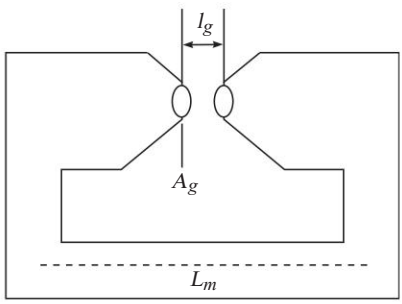


Fig. 20.38 *Permanent magnet—energy density*

Let B_g, H_g and V_g be respectively the flux, field and volume of the air gap. Therefore, the magnetic energy stored in the gap

$$= (B_g H_g)\, V_g \tag{20.138}$$

We know from Ampere's circuit theorem that, the total flux density is constant in a magnetic circuit.

i.e., $$(B_g H_g)\, V_g = -(B_m H_m)\, V_m \tag{20.139}$$

Equation (20.139) reveals that the product of *B* and *H* gives energy present per unit volume in the gap as well as in the magnet.

Generally, the magnetic circuit designer operates the magnet at a point known as the *optimum working condition point* of the magnetic material. The working point is obtained from the curve drawn between demagnetisation and energy product.

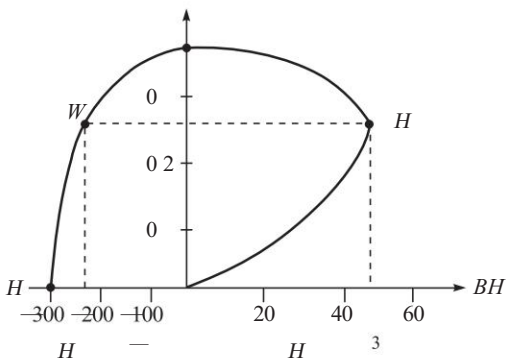


Fig. 20.39 *Demagnetisation and energy product*

The demagnetisation curve is drawn in the first and second quadrants as shown in Fig. 20.39. The parallel line to *x*-axis is drawn from the point $(BH)_{max}$ which meets the curve in the fourth quadrant at *W*. The point *W* is known as the *optimum working condition* of the magnetic material. The application of a few ferromagnetic materials along with the demagnetising field and energy product is given in Table 20.7. The energy product of a rare earth magnet is four to five times higher than AlNiCo magnet and hence, the applications of rare earth magnets are relatively higher than AlNiCo magnets.

Table 20.7 *Energy product and demagnetization field of ferromagnetic materials*

Magnetic Material	$\mu_0 H_c$ *(T)*	*Br (T)*	$(BH)_{max}$ *(KJ m^{-3})*	*Applications*
Alnico (Fe-Al-Ni-Co-Cu)	0.19	0.9	50	Wide range of permanent magnet applications
Alnico (Columnar)	0.075	1.35	60	
Strontium ferrite (Anisotropic)	0.3–0.4	0.36–0.43	24–34	Starter motors, dc motors, loud speakers, telephone receivers, various toys
Rare earth cobalt, e.g, Sm_2Co_{17} (sintered)	0.62–1.1	1.1	150–240	Servomotors, stepper motors, couplings, quality audio headphones
NdFeB magnets	0.9–1.0	1.0–1.2	200–275	Small motors (e.g., hand tools), walkman equipment, CD motors, MRI body scanners, computer applications
Hard particles $\Gamma - Fe_2O_3$	0.03	0.2		Audio and video tapes, floppy disks

20.13 FERRITE CORE MEMORY

Certain ferrite materials have square hysteresis curves as shown in Fig. 20.40. Ferrites such as (Mg, Zn, Mn) Fe_2O_4, (Cu, Mn) Fe_2O_4 and $Li_{½}$ $Fe_{½}$ Fe_2O_4 which have square hysteresis curves are used to prepare non-volatile computer memories. The square hysteresis property and the remanent state of these ferrite materials are altered by the application or the removal of a small field such as $\pm ½H_m$, where H_m is the maximum value of the magnetic field. When the applied field is greater than $\pm H_c$, then the remanent state is switched to opposite sense. The remanent states are used to represent the digits 0 and 1 respectively.

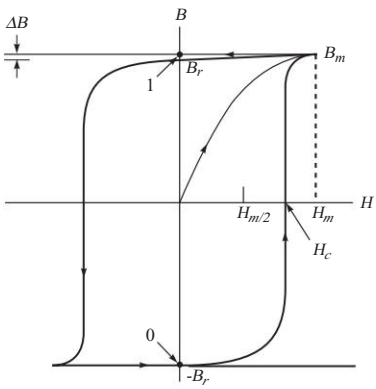


Fig. 20.40 *Square hysteresis—ferrite materials*

The magnetic cores are arranged in a matrix interlaced through fine metal wires, both horizontally and vertically, as shown in Fig. 20.41. Each core is threaded with 3 wires, namely *x* and *y* drive wires and a read wire. When an anticlockwise magnetisation is used to represent the 0 state, a large current *I* is passed through *y* drive wires from top to bottom portion. A large value of the current is needed to make all the cores into 0 states because the field produced at the centre is greater than H_m. This will clear all the memories stored in the core matrix and hence, all the cores are in the 0 state.

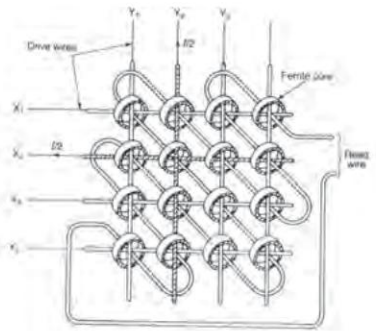


Fig. 20.41 *A part of memory core matrix*

The McGraw·Hill Companies

A change in magnetic state of the core is facilitated when the currents in the wires passing through the core reinforce each other. For example, to change the X_2 Y_2 core into the 1 state, a current of $I/2$ is applied from the bottom of Y_2 to the top portion and along X_2 from right to left. Otherwise, no change occurs. The core can be accessed randomly. Reading the information stored in a core requires a third wire threaded through the core. It picks up an induced current when test signals passing through the wires change the state of the core.

The departure of hysteresis curve from true squareness is represented by ΔB. It determines the magnitude of unwanted voltage pulse caused when a core does not switch on interrogation. A ferrite material with $\Delta B = 0$ is a suitable material for ferrite core memory. The speed of a computer is determined by the time taken to access information stored in the memory. This switching time is decreased by increasing H_C. The drive current is kept as minimum so as to minimise the power consumption. Reducing the drive current restricts increasing of H_C. By reducing the core size, for the same drive current, H_m is increased. Cores with 0.5 mm size are manufactured using dry pressing unit. When the core size is reduced, the switch speed and storage capacity are increased.

20.14 MAGNETIC RECORDING MATERIALS

Generally, magnetic materials play a dominant role in storage as well as recording/reading applications. A few examples for storage devices are magnetic tapes, floppy disks and hard disks. In addition, audio and video signals are recorded employing recording and reading heads which are made using magnetic materials. The principles, operations and applications of magnetic materials are discussed briefly in the following sections.

20.14.1 Recording-Head Materials

Generally, the recording head is made up of a soft magnetic material which has low coercivity and high saturation magnetisation. When a suitable input signal, either current or magnetic field, is applied it gives out a strong fringing magnetic field. The magnetic field produced by the recording head is stored on the magnetic film. The reading head receives the magnetic field and hence, converts the magnetic field into an electric field. Therefore, the reading head should also be of a soft magnetic material with low coercivity and high magnetisation. Magnetic materials which are used for recording heads are alloys, soft ferrites and amorphous metals. Recording-head materials are prepared employing thin-film techniques.

The following magnetic materials used for making recording heads:

(i) Alloys—Ni–Fe, Fe–Al–Si
(ii) Soft ferrites—MnZn, NiZn
(iii) Amorphous metals—Co–Zr–Nb

20.14.2 Magnetic Storage Media Materials

In magnetic storage devices like magnetic tapes, the information is stored by a process method known as *spatial magnetisation*. The information stored on magnetic devices is highly non-volatile and should not have any effect on the presence of a stray field.

Therefore, the material used for magnetic storage devices requires high coercivity as well as high remanent magnetisation. The stability of magnetic recording depends on coercivity and hence one has to maintain an optimum value for storage applications. Thus, magnetic materials with high remanent magnetisation and optimum coercivity are required for storage device applications. The different types of storage devices along with their applications are shown Table 20.8.

Table 20.8 *Magnetic storage devices and their applications*

Magnetic storage device	*Applications*
Magnetic tapes	• Store individual characters • Store information
Magnetic disks	• More storage • Continuous storage • Data/video library
Floppy disks	• Interchangeable storage for computers

20.15 MAGNETIC PRINCIPLE OF ANALOG RECORDING AND RECORDING

Generally, recording is done both by analog and digital devices. In digital recording, the information is stored both in the 1 and 0 states. On the other hand, in analog recording, the audio signal is combined with an ac bias signal. Digital recording requires a proper encoding procedure. A brief discussion on analog recording and reading are given in the following sections.

20.15.1 Recording Process

The schematic representation of recording of an audio signal on a magnetic tape is shown in Fig. 20.42. The recording head is made up of a toroidtype electromagnet with a small air gap. The electromagnet with this small air gap is known as a *recording head*. The recording head is made up of soft magnetic materials. The necessary current signal is applied to the electromagnet to produce the magnetic field in the material as well as in the air gap. The recording head is placed in contact with the top surface of the magnetic tape as shown in Fig. 20.42. The magnetic tape is made up of polymer-packing tape in which the top surface is coated with magnetic material. The magnetic tape is moves with a constant velocity.

The audio signal is first converted into an equivalent electrical current signal. The converted current signal is applied to the electromagnet which produces the equivalent magnetic field. At the same time, a fringing magnetic field is produced at the air gap which is used to magnetise the magnetic material present in the tape. The fringing magnetic field is used to magnetise the magnetic tape at a uniform speed. The strength of magnetisation on the tape depends on the strength of the fringing field which, in turn, depends on the applied current signal.

The McGraw·Hill Companies

Thus, the magnetisation on the tape depends on the strength of the current signal. Therefore, the current signal is stored on the tape as a spatial magnetic pattern permanently. In the above process, recording of the signal is carried in uni-direction and hence, this process of recording is called *longitudinal recording*.

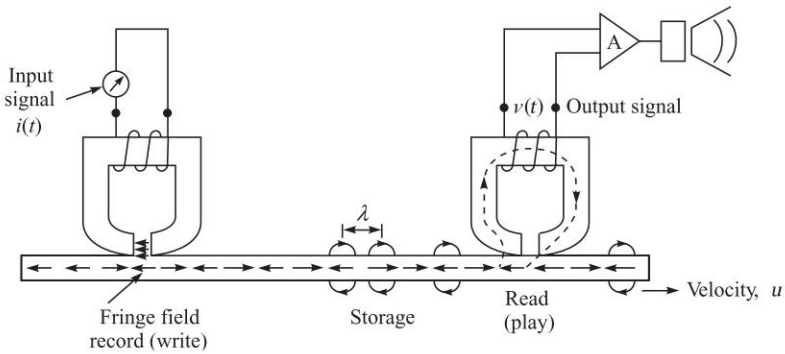


Fig. 20.42 *Magnetic recording and reading—audio cassette*

20.15.2 Reading Process

The principle of Faraday's law of induction is used for reading the audio cassettes. The recording head is used for retrieving the stored information on the magnetic tape. The magnetic tape which is moving at a constant speed is in contact with the reading head. Hence, the magnetic field which is present at the tape penetrates to the reading head. The penetrating magnetic field is strongly attracted by the reading-head magnetic material. Thus, the field is attracted towards the high permeability regions of magnetic materials. In view of the above reason, the toroid core of the electromagnet is made up of high-permeability magnetic materials. When the magnetic field from the tape enters the core, it forms a loop as shown in Fig. 20.42. This magnetic loop produces an emf at the output of the electromagnetic coil. The magnetic field available at the core of the electromagnet is converted into an equivalent voltage signal. The voltage signal at the output of the coil depends on the spatial magnetic pattern available on the magnetic tape. When the voltage signal is connected to a microphone, it converts the voltage signal into an audio signal. In a magnetic tape, the information is stored in the form of a spatial magnetic pattern. The number of data stored on the tape depends on the spatial wavelength, λ.

The spatial wavelength $\lambda = \Delta x = u/\nu$ (20.140)

where u is the velocity of the tape, ν is the frequency of the input signal and Δx is the distance moved by the tape for every $1/\nu$ seconds. A large amount of data is stored when the value of λ is very small. In most of the commercial applications, the recording and reading processes are done by the same head.

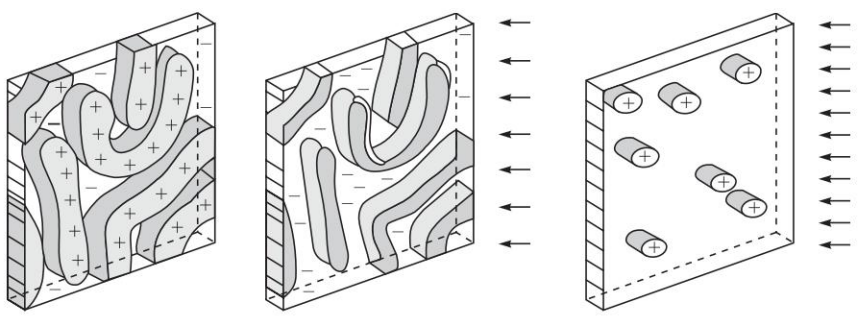
20.16 MAGNETIC BUBBLE MEMORY

A magnetic bubble memory is a non-volatile computer memory and it is in the form of thin films using soft magnetic materials. The magnetic domains with different orientations are embedded in a matrix. When a magnetic field is applied to the entire

The McGraw·Hill Companies

matrix, the magnetic domains shrink down into a tiny circle, known as *tiny magnetic domains*. The tiny magnetic domains are very small in size when compared to the magnetic domains in storage devices like tapes and hard disks. These tiny magnetic *domains* are known as *bubbles* and are used to store one bit of data. In bubble memories, the recording and reading of information are carried out employing the fast movement of bubbles by applying an electric field. As a result, the read and write time is very short in bubble memories than in those of magnetic disks and tape devices. The bubble memory was first demonstrated in rare earth orthoferrite ($RFeO_3$). A schematic representation of bubble memory is shown in Fig. 20.43.

Consider a thin magnetic film which is made up of ferrite or garnet. The magnetic domains are in the form of strips, which are arranged in two ways, namely, pointing up and downward. The behaviour of the strips totally depends on the applied field. For example, during the absence of a field, when a polarised light is applied, except for one bright strip, the others appear dark. The appearance of strips during the absence of the field is shown in Fig. 20.43a. When an external field is applied in a perpendicular direction of the thin film, the strips oppose the field and hence, the field starts to shrink as shown in Fig. 20.43b. On the other hand, when a high magnetic field is applied, the magnetic domains get concentrated into circular areas and form circular domains in the order of a few microns as shown in Fig. 20.43c. Such circular domains are called *magnetic bubbles*. The movement of these bubbles are controlled by the applied external field.



(a) Appearance of strips—absence of field *(b) Strips shrink—field is perpendicular to film* *(c) Circular domains high magnetic field*

Fig. 20.43 *Magnetic bubble memory*

The dependence of magnetic bubbles on the applied field is shown in Fig. 20.44. When the applied field crosses the value of H_s, the conversion of magnetic domains into strips takes place. A further increase in field decreases the bubble radius until the field reaches the value of H_{co}. Beyond H_{co}, the magnetic bubble collapses and becomes a magnetised material. Thus, the magnetic bubbles are stable only in certain regions of the field. These regions are used for bubble memories.

Structure of the Magnetic Bubble Memory

Magnetic bubble memory thin films are produced using the epitaxial coating method. A thin layer of magnetic garnet is coated on a non-magnetic substrate like gadolinium and gallium. The magnetic field required to convert the magnetic domains into bubbles is produced by the permanent magnet. The magnetic fields required to rotate the magnetic

The McGraw·Hill Companies

bubbles are produced employing a set of three coils. The structure of the magnetic bubble memory is shown in Fig. 20.45. The data in the bubble memories are stored similar to other storage devices using the logic 0 and logic 1 states. The rotating field is used to store the information on bubble memories. When the field is rotated once, it creates one bubble memory. The presence of a bubble memory indicates the logic 1 state, while the absence of a bubble memory indicates the logic 0 state. The existences of magnetic bubbles are studied using external magnetic fields. The data stored on the bubbles are independent of field and hence, the bubble memory is non-volatile. The materials used for the magnetic bubble memories are orthoferrites ($RFeO_3$), hexagonal ferrites ($PbFe_{12}O_{19}$), rare earth ferromagnetic and amorphous bubble materials like Gd–Co and Ge–Fe alloy films.

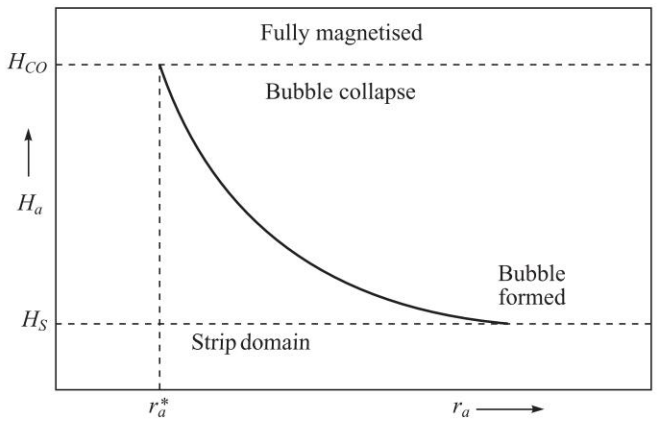


Fig. 20.44 *Dependence of magnetic bubble on applied field*

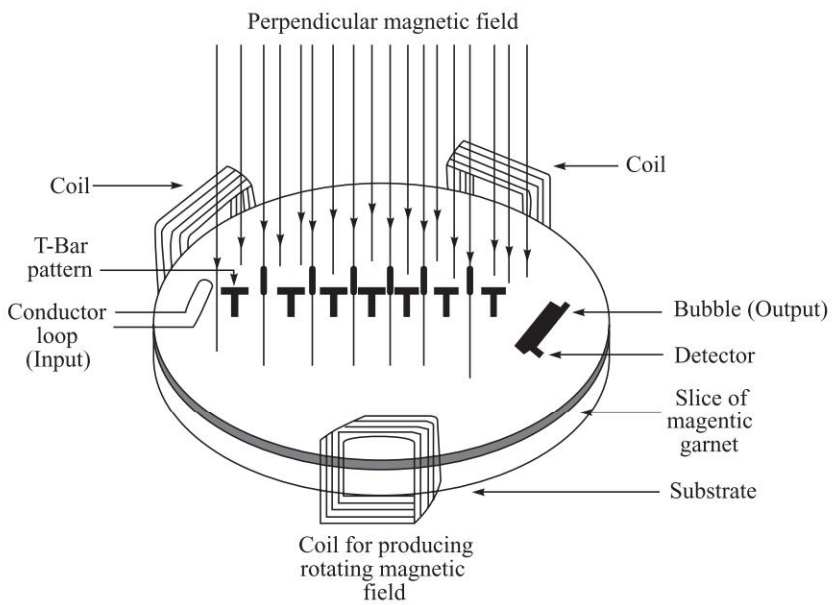


Fig. 20.45 *Structure of magnetic bubble memory*

The following are the main disadvantages of a magnetic bubble memory:

(i) It requires high recording time for storing and retrieving data.
(ii) It requires interface circuits.

20.17 MAGNETIC PRINCIPLE IN COMPUTER DATA STORAGE

The direction of magnetisation is used to represent the logic 0 and 1 states. The principle behind the data storage in storage devices is the ferromagnetic properties of magnetic materials. We know that the magnetic dipoles are arranged parallel to each other and hence, even for a small amount of external field, a large value of magnetisation is produced. The digital information stored in a magnetic device is explained based on the direction of magnetisation.

In ferromagnetic materials, the magnetic dipoles are arranged parallel to each other in the absence of an external field as shown in Fig. 20.46. Each dipole in the material will act as a tiny magnet. When an external magnetic field is applied, the north and south poles of the magnet are attracted towards the opposite poles of the external magnet. The magnet gets magnetised due to the application of an external field. When the external field is removed, the magnet posseses magnetisation. Therefore, it produces a magnetic field as shown in Fig. 20.46 due to its magnetic properties. The properties of the magnet are same as that of a permanent magnet.

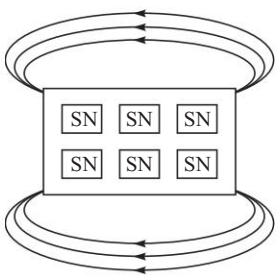


Fig. 20.46 *Magnetic material—after removing external magnetising force*

The direction of magnetising force orients the direction of south and north poles in the new magnet. Therefore, one can store data by changing the direction of magnetic domains. This means that when the field is applied in one direction, it stores logic one (i.e., 1) state, while when the field is reversed, it stores the logic zero (i.e., 0) state. The datas are stored in magnetic devices using the above principle. Generally, the following different types of magnetic storage devices are available in industries: magnetic tapes, floppy disks, hard disks and CD drives. The principles behind the above devices and the process of storing/retrieving information are discussed in detail in the following sections.

20.18 MAGNETIC TAPE

A magnetic tape is made up of a thin plastic (nylon) sheet with 0.5 inch width and large length depending on the requirements. The thin plastic sheet is coated with a

The McGraw·Hill Companies

very thin layer of ferrous (or) ferric oxide magnetic materials. Generally, a magnetic tape consists of 9 tracks and 18 tracks with a storage capacity of respectively 250 and 3800 characters per inch. A schematic representation of the cross–section of a magnetic tape is shown in Fig. 20.47.

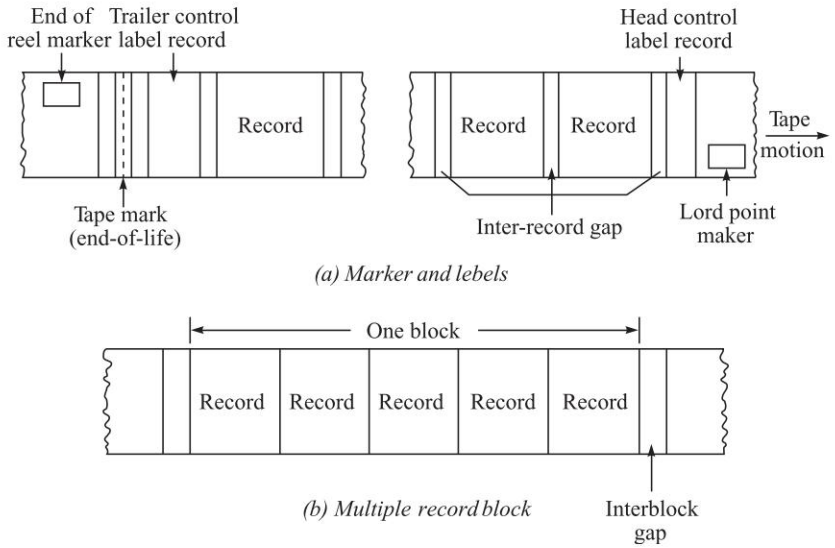


Fig. 20.47 *Magnetic tape*

The digital information are stored on the magnetic tape with the help of an electromagnet, which is known as *read/write head*. The principle of storing and retrieving information in a magnetic storage tape is similar to that of a magnetic tape used in an audio cassette. In the read/write electromagnet, the south and north poles are separated by a small gap known as the *head gap*. A schematic representation of magnetic tape and read/write head is shown in Fig. 20.48.

The information are stored on the magnetic tape using the write head, while the read head is used to retrieve the stored information on the magnetic tape. When an electric field is applied to the write head, it generates the magnetic field at the head gap. The magnetic field existing at the head gap magnetises the magnetic tape in one direction, say logic 1. This information is stored on the storage cell as the dipoles in one direction.

The next information is stored as logic 0 in the magnetic tape wherein the magnetic dipoles are in the opposite direction. This process is repeated to store the data on the magnetic tapes. The information stored on the tape is read by the read head. When the tape reaches the read head, it generates an induced magnetic field due to the existing magnetic dipoles in the storage cell. The induced magnetic field generates an electrical signal at the output of the read head. Similarly, when the data storage cell enters the read head, it will generate an induced electrical signal with different values. The difference in the electrical signals represents the logic 1 or logic 0 states which are stored in the magnetic tape.

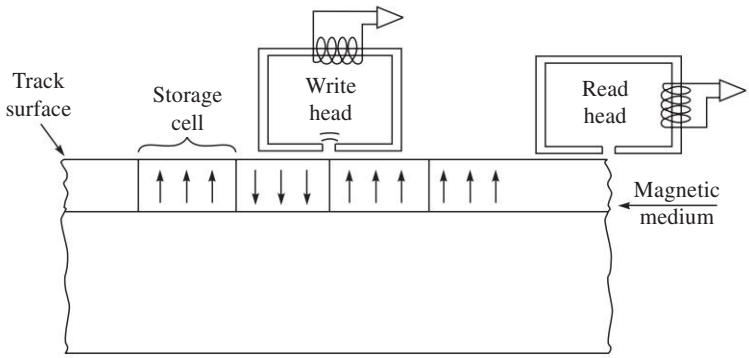


Fig. 20.48 *Schematic representation of magnetic tape and heads*

Even though the principle behind the magnetic storage is very simple, it fails to attain high density and high reliability magnetic storage devices. This is mainly due to the following reasons:

(i) The magnetic head gap must be very small.
(ii) A small magnetic field is available in a small head gap.
(iii) The small magnetic field head requires a close contact with the surface of the tape.
(iv) The data is stored on the magnetic tape with low-level field strength and hence, is easy to retrieve.
(v) A small external field may destroy the data stored on the magnetic tape.

20.19 FLOPPY DISK

Floppy disks or diskettes are used widely to store information and hence, are known as secondary and backup memory devices. A floppy disk is made up of very thin and flexible plastic (nylon) materials. Magnetic iron oxide material is coated on both sides of the disk. In order to store information on the disk, the surface of disk is divided into a number of concentric circles as shown in Fig. 20.49. These concentric circles are called *tracks*. Further, these tracks are divided into a number of sectors. The division of tracks and sectors on the disk are made only for storing and retrieving the information stored. However, these divisions are not done physically. The track is divided into a number of sectors and these are logical areas on disks. The information is stored in the disk in the form of logic 1 or logic 0 states employing magnetic domains. The logic 1 state is represented by one direction of the magnetic domain, while the logic 0 state is represented by the domain in the opposite direction. The reading and writing of data in the disk are carried out after inserting the disk in the floppy drive by the read/write head which is available in the floppy drive.

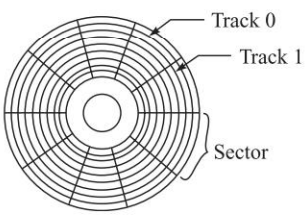


Fig. 20.49 *Magnetic floppy disk—tracks and sectors*

The McGraw·Hill Companies

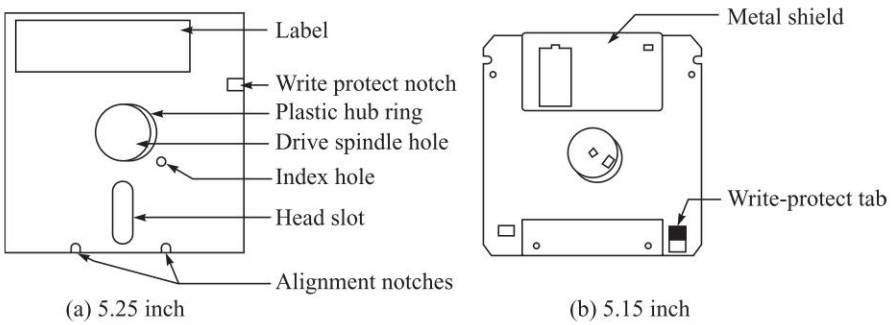


Fig. 20.50 *Floppy disk*

Generally, floppy disks are available in three different formats, namely, single, double and very high density. The techniques such as Frequency Modulation (FM) and Modified Frequency Modulation (MFM) are employed for reading/writing the single, double and very high density disks.

A simple floppy disk consists of a write protect notch, a head slot, a write protect tab and a hard plastic/thin vinyl covering cover. The read/write head is in contact with the disk through the head slot. The write protect notch/tabs is used to protect the stored information and also to control the reading and writing of data on the disk. The floppy disks are available in two different sizes, namely, 5.25 and 3.5 inches, both in double and high-density formats. The schematic representations of 5.25 and 5.15 inch floppy disks are shown in Fig. 20.50. The size and storage capacity of floppy disks are given in Table 20.9.

Table 20.9 *The size and storage capacity of floppy disks*

Size inch	*Format*	*Storage capacity*
5.25	DD	360 KB
	HD	1.2 MB
5.15	DD	7.20 KB
	HD	1.44 MB

20.20 MAGNETIC HARD DISK

A magnetic hard disk is made up of a very thin disk using metals and metal alloys. Soft magnetic materials are coated on both sides of the disk. The disk consists of thousands of concentric circles with increasing diameter from the centre. These concentric circles are known as *recording tracks*. The data are stored in the recording tracks using the read/write arm as shown in Fig. 20.51. Similar to the floppy disk, the hard disk is also divided into sectors.

Generally, the storage capacity of a single disk is very low. The hard disk capacity is increased by making a number of disks on a single driver unit as shown in

Fig. 20.52. This driver unit is called a *disk pack*. The disk pack is enclosed in a dustfree airtight container.

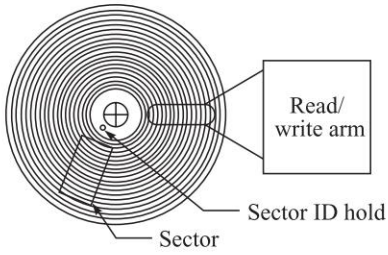


Fig. 20.51 *Magnetic hard disk*

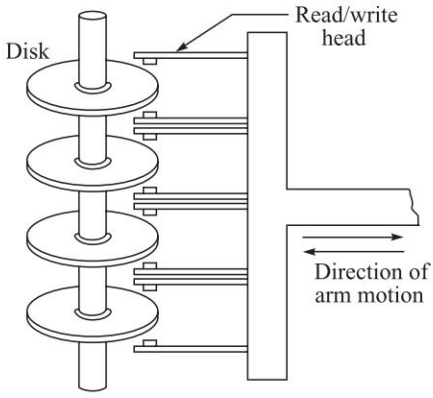


Fig. 20.52 *Magnetic hard disk pack*

The magnetic hard-disk pack consists of a number of disks, read/write head and moving arm. The disks are rotated using the disk-drive unit. Each disk has a read/write head which is attracted with a moving arm. The moving arm helps to select the particular disk and sector for recording and reading purposes. The amount of data stored in a hard disk depends on its capacity. Different types of hard disks and their storage capacities are given in Table 20.10.

Table 20.10 *Hard disk size and its storage capacity*

S. No.	*Size inches*	*Storage capacity*
1.	1.8	160 GB
2.	2.5	500 GB
3.	3.5	500–1000 GB

The McGraw·Hill Companies

20.21 COMPUTER AIDED TOMOGRAPHY

In recent years, computers have found enormous applications due to technological development. Computers are used for imaging, data acquisition, processing and display of pictures in medicine. This helps physicians solve medical problems in a simple way. Computer aided softwares help to obtain the three-dimensional image of the object.

Tomography is a technique used to slice (or) section an arbitary plane through a patient's body in order to view the net plane in detail. Computer aided tomography (CAT) helps to obtain 3D images and to get sections of images to better explore the inner fine details of tissues. Ultrasonography, X-rays, Computed Tomography (CT), Single Photon Emission Computed Tomography (SPECT), Pointer Emission Tomography (PET) and Nuclear Magnetic Resonance Imaging (NMRI) are some of the computer aided tomographic applications in imaging techniques.

20.21.1 Nuclear Magnetic Resonance Imaging

A Nuclear Magnetic Resonance (NMR) imaging technique is a standard spectroscopic technique used for characterisation of materials. The phenomenon of nuclear magnetic resonance was discovered by F Block and E. Purcell in 1946. In1972, the concept of NMR was used to study living organs like tumors by means of obtaining an image of the organs. The image obtained using NMR signals is known as *Magnetic Resonance Image (MRI)*. The magnetic signals are used to obtain the image by penetrating through medium like human tissues and air-filled structures with less attenuation.

20.21.2 Magnetic Resonance Imaging

We know that the human body is composed of tissues which contain water and fat. The major constituents of water and fat are hydrogen. The nucleus of an atom has a single proton which is a neutral choice for using the Magnetic Resonance (MR) to image the body. The response of a proton to the applied magnetic field is large in nature. Hence, the magnetic resonance image is based on the interaction between the applied magnetic field and the nucleus.

Consider that the nucleus of a hydrogen molecule is a small magnet. When the nucleus is in the direction of the applied field, it gains low energy and while it is in the opposite direction, it gains more energy. Thus, the direction of the nucleus changes based on the direction of the applied external field. The energy observed by the nucleus is emitted when it returns back to the equilibrium state. The process of absorption and emission takes place only at the resonant condition. The emitted energy from the nuclei is used to construct the image of the object.

The block diagram of an MRI system is shown in Fig. 20.53. It consists of a magnet, an rf power amplifier, transmitter/receiver coils, gradient amplifier/coil, computers and necessary softwares. The patient is subjected to a strong, uniform and stable magnetic field produced by the magnet. The magnet is placed very close to the patient. The three different gradient coils, namely, x, y and z are used to produce a controlled non-uniform magnetic field in all directions of the exposed

area of the body. The transmitter and receiver coils are placed on both sides of the patient so as to produce and receive the gradient field. When the frequency of a magnetic field which is present on the patient and a linear gradient field produced by the rf coils reaches the resonant condition, the superposition of the field takes place. When the resonance frequency of the processing nucleus is measured in the direction of the magnetic field, it gives a one-dimensional image of the object. The three/two-dimensional image of the object is obtained employing the x, y and z directions of the gradient coils. The applied magnetic fields interact with the nucleus which possesses the spin and gives out the magnetic resonance. The strength of the magnetic field and the motion of gradient coils are controlled employing computer controlled software.

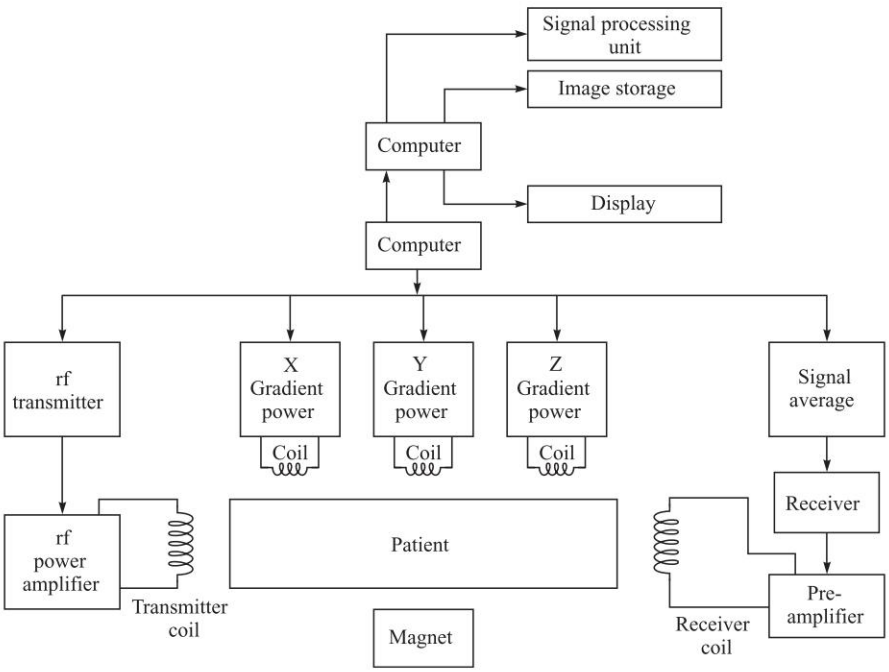


Fig. 20.53 *Block diagram of magnetic resonance imaging*

The receiver coil receives the magnetic resonance signals from the object. The received signals are amplified and then the image of the object is constructeded on the computer screen. During the construction of image, the Fourier transformation analysis is used to obtain MR images. The high quality MR images are obtained employing high temperature superconducting magnets and high resolution receiver coils. NMR has an excellent contrast behavior on soft tissues and hence, it is used to image objects such as the brain and spinal cord. Due to the development of instrumentations and image processing techniques, NMR is used to obtain the image of moving organs like the chest and abdomen.

The McGraw·Hill Companies

Keypoints to Remember

- Materials which are magnetised by the application of external field are known as magnetic materials.
- Materials which are not magnetised by the application of external field are known as non-magnetic materials.
- The north and south poles of a magnet are known as magnetic dipoles.
- The magnetic field strength is the force experienced by a unit north pole placed at that point.
- The magnetic moment per unit volume is known as magnetisation or intensity of magnetisation ($M = M_m/V$).
- The magnetic susceptibility is defined as the ratio of magnetisation to magnetic field ($\chi = M/H$).
- Magnetic permeability is defined as the ratio of amount of magnetic density to applied magnetic field intensity ($\mu = B/H$).
- The permeability and susceptibility are related as $\mu_r = 1 + \chi$.
- The magnetic materials are classified as diamagnetic, paramagnetic, ferromagnetic, antiferromagnetic and ferrimagnetic materials.
- A diamagnetic material repels magnetic lines of force.
- The Curie law for paramagnetism is $\chi = C/T$.
- Ferromagnetic materials exhibit spontaneous magnetisation even in the absence of an external field.
- Ferromagnetic material has two characteristic behaviours, namely, below the Curie temperature (θ_f), it behaves as a ferromagnetic material and above the Curie temperature, it behaves as paramagnetic material.
- Neel temperature is a critical temperature below which the material exists with antiparallel arrangements.
- Lenz's law states that the magnetisation (M) will oppose the applied magnetic field (H).
- The drawbacks of Langevin theories of paramagnetism are that it fails to explain the relationship between para and ferromagnetism and the deviation exhibited in many substances like compounds and cooled gases.
- Internal or field concept of Weiss theory is used to explain the complicated temperature dependence of susceptibility of paramagnetic materials.
- The paramagnetic Curie temperature is always greater than the ferromagnetic Curie temperature.
- The alignment of all spin magnetic moments in a small region is known as magnetic domain.
- The internal field which is present in a ferromagnetic material is known as exchange force.

- During the growth of domain, the internal energy is due to the net contribution of magnetostatic, anisotropy, domain wall and magnetostriction energies.
- During the rotation of domains walls, it gives out some fluctuations in domains due to irregular rotations which are termed as Barkhausen effect.
- When the applied field is reduced to zero, the existence of non-zero magnetisation is known as remanant magnetisation.
- Exchange interaction energy is used to explain the parallel and antiparallel arrangements of magnetic spin respectively in ferro and paramagnetic materials.
- The general chemical formula for ferrites is $Me^{2+}\ Fe_2^{3+}\ O_4^{2-}$.
- Soft magnetic materials are easy to magnetise and demagnetise.
- Hard magnetic materials are difficult to magnetise and demagnetise.
- The energy products of magnetic materials measure the amount of energy stored in a hard magnetic material.
- Magnetic materials are used in recording/reading head and magnetic devices.
- In an audio cassette, the reading process in based on the principle of Faraday's law of induction.
- Ferromagnetic properties of magnetic materials are used to store digital information in magnetic storage devices.

Solved Problems

Example 20.1 A magnetic material has a magnetisation of 2300 A m^{-1} and produces a flux density of 0.00314 Wb m^{-2}. Calculate the magnetising force and the relative permeability of the material.

Given Data

Magnetisation $M = 2300$ A m^{-1}
Flux density $B = 0.00314$ Wb m^{-2}

Solution

We know that, the magnetic flux density is

$$B = \mu_0 (M + H)$$

Rearranging the above equation, we get,

The magnetising force $$H = \frac{B}{\mu_0} - M$$

Substituting the values, we get,

$$= \frac{0.00314}{4\pi \times 10^{-7}} - 2300$$

$$= 198.7326\ A\ m^{-1}$$

The susceptibility $\chi = \frac{M}{H} = \mu_r - 1$

where μ_r is the relative permeability, i.e.

or $\mu_r = \frac{M}{H} + 1$

Substituting the values of M and H, we get,

$$= \frac{2300}{198.7236} + 1$$
$$= 12.57334$$

The magnetising force H is 198.7326 A m^{-1} and the relative permeability μ_r is 12.57334.

Example 20.2 A paramagnetic material has a magnetic field intensity of 10^4 A m^{-1}. If the susceptibility of the material at room temperature is 3.7×10^{-3}, calculate the magnetisation and flux density of the material.

Given Data

The magnetic field intensity $H = 10^4$ A m^{-1}
The susceptibility $\chi = 3.7 \times 10^{-3}$

Solution

We know that the susceptibility

$$\chi = \frac{M}{H}$$

The magnetisation $M = \chi H$

$$= 3.7 \times 10^{-3} \times 10^4$$
$$= 37 \text{ A m}^{-1}$$

We know that the flux density

$$B = \mu_0 (M + H)$$

Substituting the values of M and H, we get,

$$= 4\pi \times 10^{-7} \times (37 + 10^4)$$

$$= 1.26 \times 10^{-2} \text{ Wb m}^{-2}$$

The magnetisation in the material is 37 A m^{-1} and flux density in the material is 1.26×10^{-2} Wb m^{-2}.

Example 20.3 The magnetic field strength of copper is 10^6 ampere/metre. If the magnetic susceptibility of copper is -0.8×10^{-5}, calculate the flux density and magnetisation in copper

Given Data

The magnetic field intensity $H = 10^6 \text{ A m}^{-1}$

The susceptibility $\chi = -0.8 \times 10^{-5}$

Solution

The susceptibility $\chi = \dfrac{M}{H}$

The magnetisation $M = \chi H$

$= -0.8 \times 10^{-5} \times 10^6$

$= -8 \text{ A m}^{-1}$

The flux density B $= \mu_0 (M + H)$

Substituting the values of M and H, we get,

$= 4\pi \times 10^{-7} \times (-8 + 10^4)$

$= 1.26 \times 10^{-2}$

The flux density in the material is $1.26 \times 10^{-2} \text{ Wb m}^{-2}$

Example 20.4 A magnetic filed of 1800 ampere/metre produces a magnetic flux of 3×10^{-5} weber in an iron bar of cross sectional area 0.2 cm^2. Calculate permeability.

Given Data

Magnetising field intensity H $= 1800 \text{ A m}^{-1}$
Magnetic flux ϕ $= 3 \times 10^{-5} \text{ Wb}$
Area of cross-section A $= 0.2 \times 10^{-4} \text{ m}^2$

Solution

We know that the magnetic flux density $B = \dfrac{\phi}{A}$

Substituting the values, we get,

$= \dfrac{3 \times 10^{-5}}{0.2 \times 10^{-4}}$

$B = 1.5 \text{ Wb m}^{-2}$

We know that the permeability $\mu = \dfrac{B}{H}$

Substituting the values of B and H, we get,

$\mu = \dfrac{1.5}{1800}$

The permeability is $8.333 \times 10^{-4} \text{ H m}^{-1}$.

The McGraw·Hill Companies

Example 20.5 The saturation magnetic induction of nickel is 0.65 weber/metre^2. If the density of nickel is 8906 kg/m^3 and atomic weight is 58.7, calculate the magnetic moment of the nickel atom in Bohr magnetron.

Given Data

Magnetic induction of nickel $B = 0.65$ Wb m^{-2}

Density of nickel $\rho = 8906$ kg m^{-3}

Atomic weight $M = 58.7$

Solution

We know that the magnetic flux is

$$B = N \mu_0 \mu_m$$

where N is the number of atoms per unit volume (atoms/m^3) and it is given by

$$N = \frac{\text{density} \times \text{Avagadra's number}}{\text{Atomic number}}$$

$$= \frac{\rho N_A}{M}$$

Substituting the values, we get

$$N = \frac{8906 \times 6.023 \times 10^{26}}{58.7}$$

$$= 9.14 \times 10^{28} \text{ atoms m}^{-3}$$

The magnetisation produced per atom

$$\mu_m = \frac{B}{N \mu_0}$$

Substituting the values of B, N and μ_0, we get,

$$\mu_m = \frac{6.5}{9.14 \times 10^{28} \times 4\pi r \times 10^{-7}}$$

$$= 5.66 \times 10^{-24} \text{ A m}^2$$

We know that 1 Bohr magnetron

$$= 9.27 \times 10^{-24} \text{ A m}^2$$

$$\mu_m = \frac{5.66 \times 10^{-24}}{9.27 \times 10^{-24}}$$

$$\mu_m = 0.61 \text{ Bohr magneton}$$

The magnetic moment of the nickel atom is 0.61 Bohr magneton.

Example 20.6 A paramagnetic material has bcc structure with a cube edge of 2.5Å. If the saturation value of magnetisation is 1.8×10^6 A m^{-1}, calculate the average magnetisation contributed per atom in a Bohr magneton.

Given Data

The interatomic spacing a $= 2.5 \times 10^{-10}$ m
The magnetisation M $= 1.8 \times 10^6$ A m^{-1}

Solution

The number of atoms present per unit volume

$$= \frac{Number.of\ atoms\ present\ in\ an\ unit\ cell}{Volume\ of\ the\ unit\ cell}$$

$$= \frac{2}{\left(2.5 \times 10^{-10}\right)^3} = 1.28 \times 10^{29}\ m^{-3}$$

The magnetisation produced per atom

$$= \frac{1.8 \times 10^6}{1.28 \times 10^{29}}$$

$$= 1.40625 \times 10^{-23}\ \text{A m}^2$$

Bohr magneton, β $= \dfrac{e\,h}{4\pi m}$

$$= \frac{1.6 \times 10^{-19} \times 6.625 \times 10^{-34}}{4\pi \times 9.1 \times 10^{-31}}$$

$$= 9.27 \times 10^{-24}\ \text{A m}^2$$

The magnetisation produced per atom

$$= \frac{1.40625 \times 10^{-23}}{9.27 \times 10^{-24}}$$

$= 1.517$ Bohr magneton

The average magnetisation contributed per atom = 1.517 Bohr magneton.

Example 20.7 A system of electron spins is placed in a magnetic field of 2 wb m^{-2} at a temperature T. The number of spins parallel to the magnetic field is twice as large as the number of antiparallel spins. Determine T using classical statistics.

Given Data

A system of electron spins is placed in a magnetic field = 2 Wb m^{-2}

Solution

The number of electron spins orienting in any direction making an angle θ with the applied field is

$$n = c \times n_o \times \exp\left(-Eg / kT\right)$$

The McGraw·Hill Companies

where $E_g = -\mu \cdot H$ with the orientation direction perpendicular to the field taken as reference.

For parallel orientation potential energy $= -\mu H$

For antiparallel orientation potential energy $= \mu H$

i.e., $$n_p = C \times n_0 \times \exp(\mu H / kT)$$

$$n_p = C \times n_0 \times \exp(-\mu H / kT)$$

Hence $$\frac{n_p}{n_a} = \frac{\exp(\mu H / kT)}{\exp(-\mu H / kT)}$$

$$= \exp(2\mu H / kT)$$

Therefore, the electron spins is placed in a magnetic field of 2 Wb m^{-2}

$$\frac{n_p}{n_a} = 2$$

Hence $$\exp(2\mu H / kT) = 2$$

Taking $\log_e$ on both sides, we get,

$$\log_e 2 = 2\mu H / kT$$

or $$T = \frac{2 \times \mu \times H}{\log 2 \times k}$$

Substituting the value of μ, H and k, we get,

$$= \frac{2 \times 9.4 \times 10^{-24} \times 2}{0.6931 \times 1.38 \times 10^{-23}}$$

The temperature of the system T is 3.9 K.

Example 20.8 The rare earth element gadolinium is ferromagnetic below 16^0C with 7.1 Bohr magnetron per atom. Calculate the magnetic moment per gram. What is the value of saturation magnetisation, given that the atomic weight of gadolinium is 157.26 and its density is 7.8×10^3 kg m^{-3}?

Given Data

Bohr magnetron per atom $= 7.1\ \beta$
Atomic weight of Gd $= 1.8 \times 10^6$ A m^{-1}
The density of Gd $= 7.8 \times 10^3$ kg m^{-3}

Solution

We know that the number of atoms N is the number of atoms per unit volume (atoms/m^3) in 1 kg is

$$N = \frac{\text{density} \times \text{Avagadra's number}}{\text{Atomic number}}$$

The McGraw·Hill Companies

$$= \frac{\rho N_A}{M}$$

Substituting the values, we get

$$N = \frac{7800 \times 6.023 \times 10^{2b}}{157.26}$$
$$= 298.836 \times 10^{26} \text{ atoms m}^{-3}$$

The magnetic moment per 1 gram atom

$$= 298.836 \times 10^{2b} \times 7.1 \text{ Bohr magnetron}$$
$$= 298.836 \times 10^{2b} \times 7.1 \times 9.27 \times 10^{-24}$$
$$= 1966.849 \text{ A m}^2$$

We know that the saturation magnetisation

$$B_s = N\mu_0\mu_m$$

Substituting the values, we get,

$$= 298.836 \times 10^{26} \times 4\pi \times 10^{-7} \times 9.27 \times 10^{-24} \text{ Wb m}^{-2}$$
$$= 2.4726 \text{ Wb m}^{-2}$$

Therefore, the saturation magnetic field of Gd atom is 2.4726 Wb m^{-2}.

Example 20.9 Calculate the saturation magnetisation for Fe_3O_4 given that each cubic unit cell contains 8 Fe^{2+} and 16Fe^{3+} ions, and that the unit cell edge length is 0.839 nm.

Given data

Interatomic distance $a = 0.839 \times 10^{-9}$ m

Solution

Fe_3O_4 is ferrous ferrite ($FeO.Fe_2O_3$). It contains 16 Fe^{3+} and 8 Fe^{2+} ions per unit cell. If a single Fe_3O_4 molecule is considered, one Fe^{2+} and one Fe^{3+} ions occupies octahedral site and another Fe^{3+} ion occupies tetrahedral site. The magnetic moments of Fe^{2+} and Fe^{3+} ions are 4β and 5β respectively. The magnetic moment per Fe_3O_4 molecule is 4β. That is, the magnetic moment is contributed by Fe^{2+} ions. There are 8 Fe^{2+} ions in a unit cell. Therefore, the magnetic moment per unit cell is $8 \times 4\beta = 32\beta$.

The magnetisation is defined as the magnetic moment per volume. Therefore, saturation magnetisation, M_s = (magnetic moment / unit cell)/volume of the unit cell $M_s = \frac{32\beta}{V}$, where V is the volume of the unit cell. The unit cell is cubic and its volume is a^3.

Therefore, $$M_s = \frac{32\beta}{V} = \frac{32 \times 9.27 \times 10^{-24}}{\left(0.839 \times 10^{-9}\right)^3} = 5.022776 \times 10^{-5} \text{ A m}^{-1}.$$

The saturation magnetisation is 5.023×10^{-5} A m^{-1}.

The McGraw·Hill Companies

Example 20.10 Nickel is fee and its magnetic moment per atom is 0.6β. Its density and atomic weight are 8.9 g cm^{-3} and 58.71. Calculate (a) the saturation magnetisation and (b) the saturation flux density.

Given data

Density of Ni	= 8900 kg m^{-3}
Atomic weight	= 58.71
The magnetic moment / atom	= 0.6β

Solution

Saturation magnetisation, M_s = magnetic moment per atom × number of atoms / volume

$$= 0.6\beta N$$

Number of atoms per volume, $$N = \frac{\text{density} \times \text{Avogadro constant}}{\text{atomic weight}}$$

$$= \frac{8900 \times 6.022 \times 10^{26}}{58.71} = 9.1289 \times 10^{28}$$

Saturation magnetisation, $$M_s = 0.6 \times 9.27 \times 10^{-24} \times 9.1289 \times 10^{28}$$

$$= 5.077 \times 10^5 \text{ A m}^{-1}.$$

Saturation flux density, $$B_s = \mu_o M_s = 4\pi \times 10^{-7} \times 5.077 \times 10^5$$

$$= 0.638 \text{ Wb m}^{-2}.$$

The saturation magnetisation is 5.077×10^5 A m^{-1}.

The saturation flux density is 0.638 Wb m^{-2}.

Example 20.11 Calculate the saturation magnetisation of gadolinium. Its atomic weight, atomic number and density are respectively 157.25, 64 and 7860 kg m^{-3}.

Given data

Atomic weight	= 157.25
Atomic number	= 64
Density	= 7860 kg m^{-3}

Solution

There are eight unfilled electrons and hence the magnetic moment per atom of gadolinium is 8β, since the dipole moment contributed by one electron is 1β. The total spin is $8 \times ½ = 4$. The number of unfilled electrons is determined by writing the electronic configurations. For gadolinium, the electronic configurations are written as, $1s^2$, $2s^2$, $2p^6$, $3s^2$, $3p^6$, $4s^2$, $3d^{10}$, $4p^6$, $5s^2$, $4d^{10}$, $5p^6$, $6s^2$, $4f^7$ and $5d^1$.

We know that the saturation magnetisation,

$$M_s = 8\beta N$$

Number of atoms per volume, $N = \dfrac{\text{density} \times \text{Avogadro constant}}{\text{atomic weight}}$

Substituting the values, we get,

$$= \frac{8 \times 9.27 \times 10^{-24} \times 7.86 \times 10^{3} \times 6.022 \times 10^{26}}{157.25}$$

$$= 2.23 \times 10^{6} \text{ A m}^{-1}$$

The saturation magnetisation of gadolinium is 2.23×10^{6} A m^{-1}.

Objectives Questions

20.1 The materials which are magnetised are known as _____ materials.
20.2 _____ materials are not possible to magnetise.
20.3 The magnetic dipoles are not separate poles unlike an electric pole. (True/False)
20.4 The magnetic induction is equal to $B =$
a) ϕA b) ϕ/A
c) $\phi A/H$ d) $\phi H/A$
20.5 The magnetic dipole moment $\mu_m =$ _____
20.6 The intensity of magnetisation $M =$ _____
20.7 The magnetic susceptibility is equal to _____
a) $\chi = mH$ b) $\chi = mH/B$
c) $\chi = m/H$ d) $\chi = mB/H$
20.8 The magnetic permeability is equal to _____
a) $\mu = B/H$ b) $\mu = BH$
c) $\mu = B$ d) none of these
20.9 The permeability of free space is equal to _____ H m^{-1}.
20.10 The relation between permeability and susceptibility is _____
a) $\mu_r = 1/\chi$ b) $\mu_r = 1\chi$
c) $\mu_r = \chi$ d) $\mu_r = 1 + \chi$
20.11 The value of one Bohr magneton is $1\beta =$ _____ A m^{-2}.
20.12 All materials exhibit diamagnetic property. (True/False)
20.13 Examples for diamagnetic materials are _____, _____, _____ and _____.
20.14 Diamagnetic materials do not repel magnetic lines of force. (True/False)
20.15 The Weiss law for paramagnetism is _____
a) $\chi = T$ b) $\chi = CT$
c) $\chi = C/T$ d) $\chi = C$
20.16 _____ materials exhibit spontaneous magnetisation.

20.17 Curien–Weiss law is _____

a) $\chi = \dfrac{C}{T-\theta}$ b) $\chi = \dfrac{C}{2T-\theta}$

c) $\chi = \dfrac{C}{T+\theta}$ d) none of these

20.18 The dipoles are aligned antiparallel with equal magnitude in _____ materials.

20.19 Examples for antiferromagnetic materials are _____, _____and _____.

20.20 The orbital angular magnetisation of an electron _____

a) $\mu_a = \dfrac{lh}{\pi}$ b) $\mu_a = \dfrac{lh}{4\pi}$

c) $\mu_a = \dfrac{lh}{2\pi}$ d) $\mu_a = \dfrac{lh}{8\pi}$

20.21 The spin angular moment and magnetic moment are related as

a) $\mu_m = \dfrac{e}{2m}\mu_s$ b) $\mu_m = -\dfrac{e}{2m}\mu_s$

c) $\mu_m = \dfrac{e}{m}\mu_s$ d) $\mu_m = -\dfrac{e}{m}\mu_s$

20.22 The nuclear magneton is equal to _____

a) $\mu_m = \dfrac{eh}{2\pi M_N}$ b) $\mu_m = \dfrac{eh}{4\pi M_N}$

c) $\mu_m = \dfrac{eh}{6\pi M_N}$ d) $\mu_m = \dfrac{eh}{8\pi M_N}$

20.23 Examples for inert gases are _____, _____ and _____.

20.24 The paramagnetism principle is explained by _____ theory.

20.25 The complicated temperature dependence of susceptibility of paramagnetic materials is explained by _____ theory.

20.26 The force acquired by a paramagnetic material due to application of field _____

a) $F = \dfrac{1}{4}(\chi_2 + \chi_1)V\mu_0 \dfrac{\partial}{\partial x}(H^2)$ b) $F = \dfrac{1}{2}(\chi_2 + \chi_1)V\mu_0 \dfrac{\partial}{\partial x}(H^2)$

c) $F = \dfrac{1}{8}(\chi_2 - \chi_1)V\mu_0 \dfrac{\partial}{\partial x}(H^2)$ d) $F = \dfrac{1}{4}(\chi_2 - \chi_1)V\mu_0 \dfrac{\partial}{\partial x}(H^2)$

20.27 _____ method is used to determine the susceptibility of a paramagnetic material.

20.28 The susceptibility of a liquid paramagnetic material is determined using _____ method.

20.29 Examples for ferromagnetic methods are _____, _____ and _____.

The McGraw·Hill Companies

20.30 The paramagnetic Curie temperature is greater than ferromagnetic Curie temperature. (True/False)

20.31 The paramagnetic Curie temperature of Fe, Co and Ni are respectively _____, _____ and _____.

20.32 According to Weiss theory, the internal field produced in a material is equal to _____.

a) $H_i = \gamma_M$ b) $H_i = 2\gamma_M$

c) $H_i = 3\gamma_M$ d) $H_i = 4\gamma_M$

20.33 The internal field present in a ferromagnetic material is known as _____ force.

20.34 The irregular fluctuation in domains due to irregular rotation during the rotation of domain wall is known as _____ effect.

20.35 The magnetic spin arrangements are explained using _____ energy.

20.36 Exchange interaction energy between any two electrons is _____.

a) $E_{ex} = +2 J_{ex} S_i S_j$ b) $E_{ex} = +4 J_{ex} S_i S_j$

c) $E_{ex} = -2 J_{ex} S_i S_j$ d) $E_{ex} = -4 J_{ex} S_i S_j$

20.37 When the value of r/r_d is higher than 3 then the material is known as _____ material.

20.38 _____ materials have the value of r/r_d less than 3.

20.39 _____ is the general chemical formula for ferrites.

20.40 The hysteresis curve exhibited by ferrites is in the form of _____.

20.41 The soft magnetic materials can be magnetised easily. (True/False)

20.42 _____ magnetic materials are difficult to magnetise and demagnetise.

20.43 The energy stored in hard magnetic materials is measured using the _____.

20.44 _____ magnetic materials are used to make recording head.

20.45 High remanant magnetisation materials are used for storage devices. (True/False)

20.46 Examples for storage magnetic materials _____, _____ and _____.

20.47 _____ magnetic materials are used as magnetic bubbles.

20.48 Examples for magnetic bubbles devices _____, _____ and _____.

20.49 The principle behind magnetic storage device to store digital data is _____ properties.

20.50 The number of characters stored in an 18-track tap is _____ characters per inch.

20.51 During the read/write process, the floppy disk is rotated at speed of _____ rpm.

20.52 _____ modulation is used to write/read data in single density format disk.

20.53 The read/write process is done by using _____ modulation in case of double density disks.

20.54 The different sizes of storage disks are _____, _____ and _____ inch.

20.55 Expand CET.

20.56 Acronym for NMR.

20.57 Expansion for MRI.

20.58 Expand PET and SPECT.

The McGraw·Hill Companies

KEY

20.1 Magnetic
20.2 Non-magnetic
20.3 True
20.4 b)
20.5 $m^2\ell$
20.6 M_m/V
20.7 c)
20.8 a)
20.9 $4\pi \times 10^{-7}$
20.10 d)
20.11 9.27×10^{-24}
20.12 False
20.13 Copper, gold, mercury and silver
20.14 False
20.15 c)
20.16 Ferromagnetic
20.17 d)
20.18 Antiferromagnetic materials
20.19) Ferrous oxide, magnese oxide and chromium oxide
20.20 c)
20.21 d)
20.22 b)
20.23 He, Ne, and Ar
20.24 Langevin's
20.25 Weiss theory
20.26 d)
20.27 Gouy's
20.28 Quincke's
20.29 Fe, Co, Ni
20.30 True
20.31 1043 K, 1393 K, 631 K
20.32 a)
20.33 Exchange
20.34 Barkhausen
20.35 Exchange interaction
20.36 c)
20.37 Ferromagnetic materials
20.38 Antiferromagnetic
20.39 $Me^{2+}\ Fe_2^{3+}\ O_4^{2-}$
20.40 Square
20.41 False
20.42 Hard
20.43 Energy product
20.44 Soft
20.45 True
20.46 $\gamma - Fe_2O_3$, Co modified $\gamma - Fe_2O_3$, metallic particles
20.47 Soft
20.48 $RFeO_3$, $R_3Fe_5O_{12}$, $_6Fe_{12}O_{19}$
20.49 Ferromagnetic
20.50 3800
20.51 300
20.52 Frequency
20.53 Modified frequency
20.54 3.5, 5.25, 8
20.55 Computer aided tomography
20.56 Nuclear magnetic resonance
20.57 Magnetic resonance imaging
20.58 Positron emission tomography and emission corrupted tomography with single photon

Exercises

20.1 The magnetic field strength in a piece of copper is 10^6 A m^{-1}. Given that the magnetic susceptibility of Cu is -0.5×10^{-5}, find the flux density and the magnetisation in Cu.

The McGraw·Hill Companies

20.2 The saturation value of the magnetisation of iron is 1.75×10^6 A m^{-1}. Given that iron has bcc structure with an elementary cube edge of 2.86 Å, calculate the average number of Bohr magneton contributed to the magnetisation per atom.

20.3 The magnetic field in a piece of Fe_2O_3 is A m^{-1}. Given that the susceptibility of Fe_2O_3 at room temperature is 1.4×10^{-3}, find the flux density and the magnetisation in the material.

20.4 Calculate (a) saturation magnetisation, and (ii) the saturation flux density of nickel ferrite $(NiFe_2O_4)_8$ which has a unit cell edge length of 0.8337 nm.

20.5 The magnetic flux density within a bar of some material is 0.630 Wb m^{-2} at a field H of 5×10^5 A m^{-1}. Compute the following for this material, (a) the magnetic permeability, and (b) magnetic susceptibility (c) What type(s) of magnetism would you suggest by this material? Why?

20.6 An iron bar of coercivity of 4000 A m^{-1} is to be demagnetized. If the bar is inserted within a cylindrical wire coil 0.15 m long and having 100 turns, what electric current is required to generate the necessary magnetic field? [Hint: $B = \mu_o \times$ turns/length $\times$ I]

Short–Answer Questions

20.1 Define magnetic materials.
20.2 What is meant by non-magnetic materials?
20.3 Define magnetic dipole.
20.4 What is magnetic moment?
20.5 Explain magnetic induction.
20.6 Define the term magnetic flux density.
20.7 Explain magnetic field strength.
20.8 What is meant by magnetisation?
20.9 Define the term magnetic susceptibility.
20.10 Define magnetic permeability.
20.11 Define Bohr magneton. What is its unit?
20.12 What is a diamagnetic material?
20.13 What are the different types of magnetic materials?
20.14 Mention the magnetic materials having permenant magnetic moment.
20.15 Mention the properties of a diamagnetic material.
20.16 What is a paramagnetic material?
20.17 Mention the properties of a paramagnetic material.
20.18 Distinguish diamagnetic and paramagnetic materials.
20.19 What is a ferromagnetic material?
20.20 Explain the properties of a ferromagnetic material.
20.21 Define the terms: (i) retentivity (ii) coercivity.

The McGraw·Hill Companies

20.22 List out the properties of ferromagnetic materials.
20.23 Explain an antiferromagnetic material.
20.24 What are the properties of an antiferromagnetic material?
20.25 What are ferromagnetic materials?
20.26 Distinguish between ferromagnetic and antiferromagnetic materials.
20.27 Explain the properties of ferromagnetic materials.
20.28 What is meant by internal field of a paramagnetic material?
20.29 What is a domain?
20.30 Write down the applications of ferrites.
20.31 What is a hard magnetic material?
20.32 What are the properties of a hard magnetic material?
20.33 What are soft magnetic materials?
20.34 Distinguish between soft and hard magnetic materials.
20.35 List out the properties of a soft magnetic material.
20.36 Mention the applications of a soft magnetic material.
20.37 Mention the applications of a hard magnetic material.
20.38 Define the term energy product of a magnetic material.
20.39 What are ferities?
20.40 What is Bohr magnetism?
20.41 Mention the applications of ferrites.
20.42 Define orbital angular momentum of an electron.
20.43 Explain spin angular momentum of an electron.
20.44 State Lenz's law of magnetisation.
20.45 What is meant by Langevin's function?
20.46 Define saturation magnetisation.
20.47 Mention the drawbacks of Langevin theory.
20.48 Explain Weiss theory of paramagnetism.
20.49 What is Curie–Weiss law?
20.50 Explain the principle behind Gouy's method.
20.51 Give four examples of ferromagnetic materials.
20.52 Define paramagnetic Curie temperature.
20.53 What is ferromagnetic Curie temperature?
20.54 What is meant by hysteresis curve?
20.55 What is *B-H* curve?
20.56 What is *M-H* curve?
20.57 Explain the domain concept.
20.58 What is meant by exchange force?
20.59 Define spontaneous magnetisation.
20.60 What is meant by magnetic domain?
20.61 Define magnetostatic energy.
20.62 What is magnetostriction energy?
20.63 Define molecular field.
20.64 What is meant by Neel temperature?
20.65 Gibe any two examples for inverse spinal.
20.66 Write the chemical formula for regular spinal.
20.67 Differentiate $+M_s$ and $-M_s$.
20.68 Mention any four applications of ferrites.
20.69 Define the term 'energy product' of a magnetic material.

20.70 What is a floppy disk?
20.71 What is a hard disk?
20.72 What is magnetic tape?
20.73 What are magnetic bubbles?
20.74 What is the magnetic principle used in computer data storage?
20.75 Explain the process of magnetic recording in an audio tape.
20.76 Explain the process of reading in an audio tape.
20.77 Mention some materials used for magnetic recording.
20.78 What is meant by Bloch wall?
20.79 What is saturation strain?
20.80 What is Barkhausen effect?
20.81 What is meant by exchange interaction?
20.82 What is hard direction in a magnetic material?
20.83 What is easy direction in a magnetic material?
20.84 What do you mean by anisotropy energy?
20.85 What is ferrite core memory?

Descriptive Questions

20.1 What is meant by permanent dipole moment? Explain the origin of permanent dipole moment in magnetic materials.
20.2 What is a magnetic material? Distinguish between a hard and a soft magnetic material.
20.3 What is diamagnetic material? Explain the Langevin's classical theory of diamagnetism and hence derive an expression for the susceptibility of a diamagnetic material.
20.4 What is paramagnetic material? Explain the Langevin's classical theory of paramagnetism and hence derive an expression for the susceptibility of a paramagnetic material.
20.5 What are the drawbacks of the Langevin's classical theory of paramagnetism? Explain how the drawback of Langevin's theory is overcome in Weiss theory and hence derive an expression for the susceptibility of a paramagnetic material using Weiss theory.
20.6 Using a two-level quantum mechanical model, derive an expression for the susceptibility of a paramagnetic material.
20.7 Derive an expression for the susceptibility of a paramagnetic material usingquantum theory.
20.8 Explain the experimental determination of the paramagnetic susceptibility of a (i) solid using Gouy's method, and (ii) a liquid using Quincke's method.
20.9 Explain the Weiss theory of ferromagnetic materials. Explain how the saturation magnetisation is explained using Weiss internal field concept.
20.10 What are magnetic domains? Explain, with suitable sketches, the hysteresis property of a ferromagnetic material.

The McGraw·Hill Companies

20.11 What is an antiferromagnetic material? Explain the antiparallel alignment of dipoles in antiferromagnetic material with suitable sketch and hence derive an expression for the susceptibility of an antiferromagnetic material.

20.12 (a) Explain the Heisenberg theory ferromagnetism.
(b) Explain the hysteresis curve using domain concept.

20.13 What are ferrites? Explain the different types of structures of ferrites with suitable diagrams and hence explain the antiparallel alignment of dipoles in ferrites using magnetic moment calculations.

20.14 Explain the magnetic principle involved in computer data storage and hence explain how the data are stored in (i) magnetic tape, (ii) floppy disk, and (iii) hard disk.

20.15 What do you understand by magnetic exchange interaction of dipoles? Explain the Heisenberg theory of ferromagnetism.

20.16 Write short notes on the following: (i) anisotropy energy, (ii) magnetostrictive energy, (iii) magnetostatic energy, and (iv) domain wall energy.

20.17 Explain in detail the energy product of magnetic material. What is its Importance?

20.10 Explain in detail the process of recording and reading an audio tape. Mention the magnetic materials used for storage media and recording head.

Chapter

Dielectric Materials

21

OBJECTIVES

- To explain the dielectric nature of the material with necessary definitions
- To discuss the different types of polarization according to its dipoles
- To derive the local field of dielectric material
- To discuss the various types of dielectric materials and its application
- To derive the Claussius–Mosotti equation for the dielectric material and discuss its importance
- To discuss the dielectric loss and explain the various factors to destroy the dielectric properties
- To discuss the change in the polarization of selected dielectric materials as a function of applied temperature and applied frequency
- To discuss the active and passive dielectric materials
- To find some applications of dielectric materials

21.1 INTRODUCTION

All dielectric materials are insulators. The distinction between a dielectric material and an insulator lies in the application to which one is employed. The insulating materials are used to resist the flow of current through it when a difference of potential is applied across its ends. On the other hand, the dielectric materials are used to store electrical energy.

A dielectric material is one which stores electrical energy with a minimum dissipation of power, since the electrons are bound to their parent molecules and hence, there is no free charges.

Let us consider some important terms related to this section.

The McGraw·Hill Companies

21.2 DEFINITIONS

(i) Electric Field Intensity or Electric Field Strength Consider a point charge dq in the region of an electric field. Let F be the force acting on the point charge dq.

The force per unit test charge dq is known as electric field strength (E), given by

$$E = \frac{F}{dq} = \frac{Q}{4\pi\varepsilon r^2} \tag{21.1}$$

From Coulomb's law, when two point charges Q_1 and Q_2 are separated by a distance r, the force of attraction or repulsion between the two charges is

$$F = \frac{Q_1 Q_2}{4\pi\varepsilon r^2}\hat{n} \tag{21.2}$$

where ε is the permittivity or dielectric constant of the medium in which the charge is placed. For vacuum, $\varepsilon = \varepsilon_0 = 8.854 \times 10^{-12}$ F m^{-1}.

(ii) Electric Flux Density or Electric Displacement Vector The electric flux density or electric displacement vector D is the number of flux lines crossing a surface normal to the lines, divided by the surface area.

The electric flux density at a distance r from the point charge Q can be written as,

$$D = \frac{Q}{4\pi r^2} \tag{21.3}$$

where $4\pi r^2$ is the surface area of a sphere of radius r.
From Eqs (21.1) and (21.2), we get,

$$D = \varepsilon E = \varepsilon_0 E + P \tag{21.4}$$

where P is the polarisation and it has the same unit as D, i.e., coulomb per square metre (c m^{-2}).

(iii) Dielectric Parameters The dielectric parameters are

(a) Dielectric constant (ε_r)
(b) Electric dipole moment (μ)
(c) Polarisation (P), and
(d) Polarizability (α).

(a) Dielectric Constant The dielectric constant (ε_r) of a material is defined as the ratio of the permittivity of the medium (ε) to the permittivity of free space (ε_0),

$$\varepsilon_r = \frac{\varepsilon}{\varepsilon_0} \tag{21.5}$$

where ε_r, is the dielectric constant, which is a dimensionless quantity. The measure of dielectric constant or relative permittivity gives the properties of a dielectric material. The dielectric constant of air is one.

(b) Electric Dipole Moment Consider two charges of equal magnitudes but of opposite polarities separated by a distance r, as shown in Fig. 21.1.

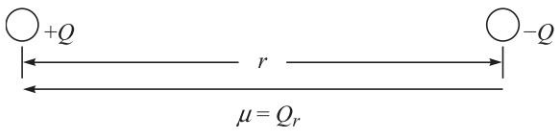


Fig. 21.1 *Electric dipole*

The arrangement of two equal and opposite charges + Q and – Q, separated by a distance r is known as electric dipole. The product of magnitude of the charge and the distance of separation is known as electric dipole moment (m).

$$\mu = \text{charge} \times \text{distance } Q_r \tag{21.6}$$

where μ is the dipole moment and its unit is coulomb metre (c m). Since the dipole moment is a vector, it points from the negative to the positive charges as shown in Fig. 21.1.

The total dipole moment of a system constituting of point charges $Q_1, Q_2, Q_3, .., Q_n$ and the distances of separation $r_1, r_2, .., r_n$ is

$$m_{\text{total}} = \sum_{i=1}^{n} Q_i \, r_i \tag{21.7}$$

(c) Polarisation When an electric field is applied to a solid material consisting of positive and negative charges, the positive charges are displaced opposite to the direction of the field, while the negative charges are displaced in the direction of the field. The displacement of these two charges create a local dipole in the solid. This type of displacement of positive and negative charges by the application of electric field leads to polarisation.

$$P = \frac{\mu}{\text{volume}} \tag{21.8}$$

where P is the polarisation in c m^{-2}. The term, 'Polarisation' is defined as the induced dipole moment per unit volume.

(d) Polarisability When a dielectric material is placed in an electric field, the displacement of electric charges gives rise to the creation of dipole in the material. The polarisation P of an elementary particle is directly proportional to the electric field strength E

i.e., $$P \propto E$$

$$P = \alpha E \tag{21.9}$$

where α is a proportionality constant known as polarisability. The unit of α is F m^2.

If the solid material contains N number of particles per unit volume, then the polarisation can be written as,

$$P = N \alpha E \tag{21.10}$$

where $\alpha = \alpha_e + \alpha_i + \alpha_0$. Here α_e, α_i and α_0 are respectively the electronic, ionic and orientation polarisability. In the following sections, the above three polarisations are discussed in detail:

21.3 DIFFERENT TYPES OF POLARISATIONS

The application of an electric field to a dielectric material creates or realigns the dipoles resulting in polarisation. There are four different types of polarisations. They are listed below.

(i) Electronic or induced polarisation (P_e)
(ii) Atomic or ionic polarisation (P_i)
(iii) Orientation polarisation (P_0), and
(iv) Interfacial or space charge polarisation

21.3.1 Electronic or Induced Polarisation

A dielectric material consists of a large number of atoms. Let us, consider a dielectric material consisting of only one atom, as shown in Fig. 21.2a. The nucleus is at its centre while the electrons are revolving around the nucleus. When an electric field is applied to this atom, the nucleus moves away from the field while the electrons move towards the field. Therefore, there is a displacement between the nucleus and the electrons. This displacement produces an induced dipole moment, and hence, polarisation. The polarisation produced due to the displacement of electrons is known as electronic polarisation.

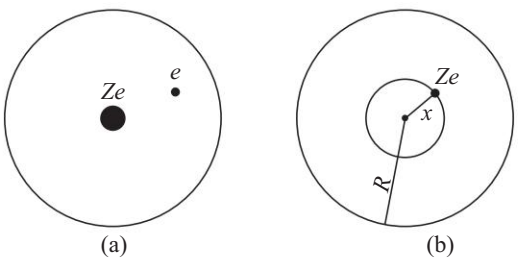


Fig. 21.2 *Electronic polarisation*

Let Z and R be respectively the atomic number and radius of an atom. Similarly, let e be the charge of an electron. If E is the electric field intensity applied to the atom, then the force acquired by the nucleus is ZeE. Due to the application of the electric field, there is a displacement x of the nucleus from its position. If a sphere is drawn by taking x as the radius, then the entire portion of the atom is splitted into two. From Gauss's theorem it can be proved that the electron cloud present outside the small sphere of radius x has no impact on the nucleus. Therefore, the nucleus exerts force only due to the electron cloud present in the inner portion of the sphere of radius x.

Let the nucleus be present at the centre of the sphere of radius x and the electron clouds are present throughout this sphere.

$$\text{The charge inside the sphere} = \frac{\frac{4}{3}\pi x^3}{\frac{4}{3}\pi R^3} Ze$$

$$\Rightarrow \qquad = \frac{x^3}{R^3} Ze \qquad (21.11)$$

The Coulomb force acting between the nucleus and the electron clouds inside the sphere of radius x is given by,

$$F = \frac{Ze\left(\frac{-x^3}{R^3} Ze\right)}{4\pi\varepsilon_0 x^2} \tag{21.12}$$

At equilibrium, the nucleus is balanced and hence, the total force on the nucleus is zero. Therefore,

$$ZeE = \frac{Ze\left(\frac{x^3}{R^3} Ze\right)}{4\pi\varepsilon_0 x^2}$$

Simplifying the above equation, we get,

$$x = \frac{4\pi\varepsilon_0 R^3 E}{Ze} \tag{21.13}$$

The induced dipole moment is given by

$$\mu_{\text{ind}} = Ze\, x$$

Substituting the value of x from Eq. (21.13) in the above equation, we get,

$$\mu_{\text{ind}} = 4\pi\, \varepsilon_0\, R^3 E$$

or

$$\mu_{\text{ind}} = \alpha_e\, E$$

where α_e is the electronic polarisability and is equal to

$$\alpha_e = 4\pi\, \varepsilon_0\, R^3 \tag{21.14}$$

It is clear from Eq. (21.14) that the electronic polarisability is directly proportional to R^3. From Eq. (21.14), the electronic polarisation can be written as,

$$P_e = N\, \alpha_e\, E \tag{21.15}$$

Substituting the value of α_e from Eq. (21.14) in the above equation, we get,

$$P_e = N\, 4\pi\, \varepsilon_0\, R^3\, E$$

i.e.,

$$\varepsilon_0(\varepsilon_r - 1)\, E = 4\, N\, \pi\, \varepsilon_0\, R^3\, E$$

$$\varepsilon_r - 1 = 4\, N\, \pi\, R^3 \tag{21.16}$$

where ε_r is the dielectric constant of material.

The dielectric constant of Al at 1 atmospheric pressure is found to be 1.0000192129 and the number of dipoles per unit volume present is 2.7×10^{25} atoms m^{-3}. The value of R obtained from Eq. (21.16) is 0.384 Å, which is of the order of the radius of an atom. The displacement of electrons and nucleus due to the application of electric field is a crude assumption which is not accepted by the atomic model. Anyhow, the results obtained by using this assumption give the correct values of the atomic parameters. Therefore, even though it is a crude model, it cannot be neglected.

21.3.2 Atomic or Ionic Polarisation

When an electric field is applied to an ionic crystal, the polarisation that arises due to the displacement of the positive ions away from the field and the displacement of the negative ions towards the field, is known as atomic or ionic polarisation. This type of polarisation is produced in ionic molecules such as NaCl, KBr, KCl and LiBr.

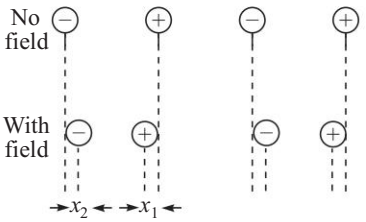


Fig. 21.3 *Ionic polarisation*

Consider the arrangement of ionic molecules as shown in Fig. 21.3. In the absence of electric field, there is no displacement of ions. When an electric field is applied, there is a displacement of the positive and negative ions which produces an induced dipole moment μ_i. This induced dipole moment in turn produces an induced polarisation. The induced average ionic polarisation produced per ionic dipole is given by

$$P_i = \alpha_i E_i$$

The polarisation produced for a crystal having N number of dipoles per unit volume is given by,

$$P_i = N \alpha_i E_i \tag{21.17}$$

Let x_1 and x_2 be the displacements of the positive and negative ions, produced due to the application of an electric field. The dipole moment induced is given by,

$$\mu_i = e\,(x_1 + x_2)$$

Let F be the force experienced by the ions due to the application of the electric field. The restoring force acting on the ions is directly proportional to their displacements.

i.e.,
$$F \propto x_1$$
$$\propto x_2$$
$$= \beta_1 x_1 = \beta_2 x_2 \tag{21.18}$$

where β_1 and β_2 are the proportionality constants and they are directly proportional to the mass and angular frequency of the respective ions.

i.e., $\beta_1 \propto m$ and $\beta_2 \propto M$

$$\propto \omega_0^{\,2}$$

Form Eq. (21.18), the force experienced by the positive and negative ions is given by,

$$F = eE = m\omega_0^{\,2} x_1 = M\omega_0^{\,2} x_2$$

Therefore, $$x_1 = \frac{eE}{m\omega_0^2} \quad \text{and} \quad x_2 = \frac{eE}{M\omega_0^2}$$

The McGraw·Hill Companies

Substituting the values of x_1 and x_2, the induced dipole moment can be written as,

$$\mu_i = \frac{e^2}{\omega_0^2}\left[\frac{1}{m} + \frac{1}{M}\right]E \tag{21.19}$$

The ionic polarisability $\alpha_i = \frac{e^2}{\omega_0^2}\left(\frac{1}{M} + \frac{1}{m}\right)$ (21.20)

Using Clausius–Mosotti equation, the dielectric constant for an ionic crystal can be written as,

$$\frac{\varepsilon_r - 1}{\varepsilon_r + 2} = \frac{N\alpha_i}{3\varepsilon_0} \tag{21.21}$$

In addition to ionic polarisation, an ionic molecule also possesses electronic polarisation due to the displacement of electron clouds. The electronic polarisation of an ionic molecule will be in the order of $1/10^{th}$ of ionic polarisation. Hence, its magnitude is much smaller than the ionic polarisation.

21.3.3 Dipolar or Orientation Polarisation

The dipolar or orientation polarisation is produced only in case of polar molecules such as H_2O, HCl and nitrobenzene. When an electric field is applied to a polar molecule, the dipoles experience a torque and try to align parallel to the applied field, which results in a rotation of the dipoles. The mechanism of dipolar polarisation is shown in Fig. 21.4.

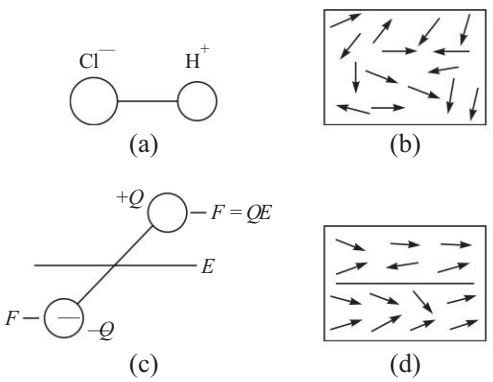


Fig. 21.4 *Dipolar polarisation: (a) The dipole in a HCl molecule, (b) The random orientation of dipoles in the absence of electric field, (c) The rotation of a dipole due to the applied electric field, and (d) The dipoles try to align parallel to the applied field*

Consider a polar molecule is subjected to an electric field of strength E. Let θ be the angle of rotation of the dipoles.
Then, the torque produced by the field on the dipole is

$$\tau = \mu_p E \sin\theta \tag{21.22}$$

where μ_p is the permanent dipole moment.

The maximum work is done when the dipole is rotated through an angle of $\theta = 180°$. When the dipole is already parallel to the applied field, i.e., when $\theta = 0°$, the work done is minimum. Therefore, the maximum energy is given by

$$E_{max} = \int_0^{\pi} \mu_p E \sin\theta \, d\theta$$

$$= 2\mu_p E \tag{21.23}$$

The average dipole energy E_{dip}

$$E_{dip} = \mu_p E \tag{21.24}$$

The ratio of the average dipole energy to average thermal energy is, given by

$$\frac{\text{Average dipole energy}}{\text{Average thermal energy}} = \frac{\mu_p E}{\frac{5}{2} kT} \tag{21.25}$$

If this ratio is greater than unity, then the orientation polarisation is said to be effective. The average orientation polarisation is

$$P_o \propto \frac{\text{Permanent dipole moment} \times \text{average dipole energy}}{\text{Average thermal energy}}$$

or,

$$P_o \propto \mu_p \frac{\mu_p E}{\frac{5}{2} kT} \tag{21.26}$$

If the calculation for average dipole energy is properly done using Boltzmann's statistics, then the average orientation polarisation is

$$P_o = \frac{1}{3} \frac{\mu_p^2 E}{kT} \tag{21.27}$$

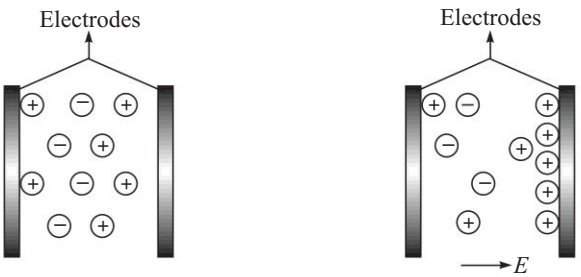
Therefore, the orientation polarisability can be written as,

$$\alpha_e = \frac{\mu_p^2}{3kT} \tag{21.28}$$

It is clear from Eq. (21.28) that the orientation polarisation is temperature dependent.

21.3.4 Interfacial or Space Charge Polarisation

Consider a dielectric medium is placed between any two electrodes, as shown in Fig. 21.5. When no field is applied to the electrode, the positive and negative charges are not separated and there are fixed number of charges. On the other hand, when an electric field is applied, the charges are separated. The positive charges are accumulated near the negative electrode. Therefore, a dipole moment is induced due to displacement of the ions. Then, the induced dipole moment per unit volume gives the induced polarisation. This polarisation is known as interfacial polarisation.



(a) *absence of an electric field* (b) *presence of an electric field*

Fig. 21.5 *Space-charge polarisation*

21.3.5 Total Polarisation

In the calculation of total polarisation, the space charge polarisation is not taken into account, since it occurs at interfaces and it is very small and hence negligible. In addition to this, the fields are not well defined at interfaces.

Therefore, the total polarisation is the sum of the electronic, ionic and orientation polarisation. Therefore, total polarisation is given by

$$P = P_e + P_i + P_o \tag{21.29}$$

Substituting the values of P_e, P_i and P_o respectively from Eqs (21.15), (21.18) and (21.27) in Eq. (21.29), we get,

$$P = N\,\alpha_e E_i + N\,\alpha_i\,E_i + N\,\alpha_o\,E_i \tag{21.30}$$

Substituting the values of α_e, α_i and α_o respectively from Eqs (21.14), (21.20) and (21.28) in Eq. (21.30), we get,

$$P = N\,E_i\left(4\pi\varepsilon_0 R^3 + \frac{e^2}{\omega_0^2}\left(\frac{1}{M}+\frac{1}{m}\right)+\frac{\mu_p^2}{3kT}\right) \tag{21.31}$$

When a plot is drawn between P and $\frac{1}{T}$, a straight line is obtained, as shown in Fig. 21.6. The straight line makes an intercept at the y-axis, when $\frac{1}{T} = 0$. The value of the temperature-independent portion in the plot is $N(\alpha_e + \alpha_i)$. Therefore, by drawing a plot between P and $\frac{1}{T}$, one can determine the orientation polarisation and the sum of electronic and ionic polarisation.

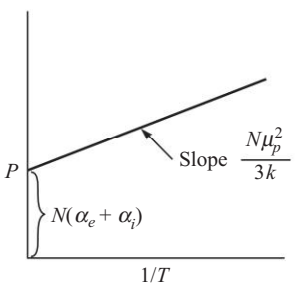


Fig. 21.6 *Plot of P versus 1/T*

21.4 LOCAL OR INTERNAL FIELD

In a dielectric material, the field acting at the location of an atom is known as local or internal field E_i. The internal field seen by an atom is not equal to the externally applied electric field E.

The internal field E_i must be equal to the sum of applied field plus the field produced at the location of the atom by the dipoles of all other atoms.

$$E_i = E + \text{the field due to all other dipoles} \tag{21.32}$$

Calculations of Internal Field for Solid Dielectrics To calculate the internal field seen by an atom in a solid dielectric, let us consider a simple example which consists of an array of atoms represented as A, B_1, B_2, C_1, C_2, etc. B_1, B_2, C_1, C_2, etc. are atoms equidistant from A, arranged in the form of a string, as shown in Fig. 21.7. Let α be the polarisability of these atoms.

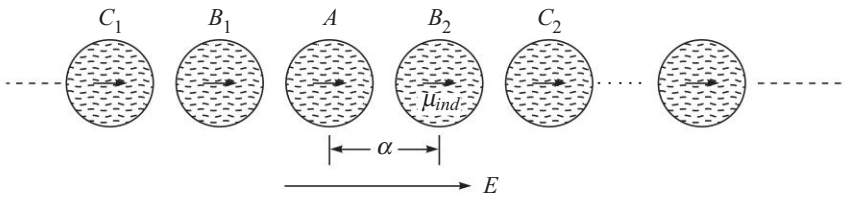


Fig. 21.7 *Local field*

From the symmetry of the system, it is clear that the internal field E_i acting on the site of atom A will be parallel to the applied field E, and will be same in all other atoms B_1, B_2, C_1, C_2, etc.

The induced dipole moment on each atom of the string

$$\mu_{ind} = \alpha E_i \tag{21.33}$$

where E_i is the local field acting at the atom A due to the external field E and the field by the dipoles on atoms B_1, B_2, C_1, C_2, etc.

$$E_i = E + \text{the field due to all other dipoles at the atom } A. \tag{21.34}$$

Consider the point dipole on atom B_1, whose dipole moment is μ. According to the field theory, the potential around a point dipole in vacuum is

$$V(r, \theta) = \frac{1}{4\pi\varepsilon_0} \frac{\mu \cos\theta}{r^2} \tag{21.35}$$

where r is the distance from the dipole to the point charge and q the angle between r and μ, as shown in Fig. 21.8.

The field around the dipole has two components, one due to r and the other due to θ.

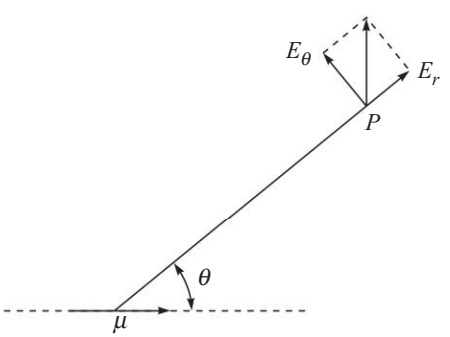


Fig. 21.8 *Potential at a point P due to a dipole*

The McGraw·Hill Companies

The r component field

$$E_r = \frac{-\partial V}{\partial r}$$

$$E_r = \frac{1}{2\pi\varepsilon_0}\frac{\mu\cos\theta}{r^3} \tag{21.36}$$

The θ component field

$$E_\theta = -\frac{1}{r}\times\frac{\partial V}{\partial\theta}$$

$$= \frac{\mu}{2\pi\varepsilon_0}\times\frac{\sin\theta}{r^3} \tag{21.37}$$

[The negative gradient of potential is due to the electric field at A]. The field produced by B_1 atom at the location of the atom A can be obtained by taking $r = a$ and $\theta = 0$ in Eqs (21.36) and (21.37), since the point dipoles lie on the axis of the dipole.

The field acting on the atom A by the atom B_1

$$= E_r + E_\theta = \frac{\mu_{ind}}{2\pi\varepsilon_0 a^3} + 0 \tag{21.38}$$

Substituting the values of E_r and E_θ in the above equations, we get,

$$= \frac{\mu_{ind}}{2\pi\varepsilon_0 a^3}$$

where $\mu = \mu_{ind}$, since the dipole is an induced one.

The field produced at A by dipole on B_1 is equal to that produced by B_2. The atoms are separated by a distance a.

Therefore, the internal field at A

$$E_i = E + 2\times\frac{\mu_{ind}}{2\pi\varepsilon_0 a^3} + 2\times\frac{\mu_{ind}}{2\pi\varepsilon_0(2a)^3} + \ldots$$

$$= E + 2\times\frac{\mu_{ind}}{2\pi\varepsilon_0 a^3}\left[1 + \frac{1}{2^3} + \frac{1}{3^3} + \ldots\right]$$

$$= E + \frac{\mu_{ind}}{\pi\varepsilon_0 a^3}\times\sum\frac{1}{n^3} \tag{21.39}$$

where $n = 1, 2, 3,$ etc.

We know that from Rice man's zeta function $\sum\frac{1}{n^3} = 1.2$. Substitutly this values is Eq. (21.39) can be written as

$$E_i = E + \frac{1.2\mu_{ind}}{\pi\varepsilon_0 a^3} \tag{21.40}$$

From Eq. (21.40), it is clear that the internal field is greater than the external field for the model chosen, suggesting that the dipoles cooperate with each other. The

The McGraw·Hill Companies

cooperation becomes stronger as the polarisability of the atom increases and the distance between them decreases.

The internal field E_i in a three dimensional case is very complicated and depends upon the crystal structures as defined by

$$E_i = E + \frac{\gamma P}{\varepsilon_0} \tag{21.41}$$

where P is the dipole moment per unit volume (polarisation) and γ proportionality constant known as internal field constant. When the atoms in a solid are surrounded cubically by other atoms, the value of $\gamma = \frac{1.2}{\pi} \approx 1/3$.

Equation (21.41) can be written as,

$$E_i = E + \frac{P}{3\varepsilon_0} \tag{21.42}$$

This field is known as Lorentz internal field.

It is clear from the expression of Lorentz field that the internal field at the atoms is larger than the applied field by an amount directly proportional to the induced dipole moment.

21.5 TYPES OF DIELECTRIC MATERIALS

The dielectric materials are classified into three types as given below.

(i) Solid dielectric
(ii) Liquid dielectric, and
(iii) Geseous dielectric

21.5.1 Solid Dielectric

There are three types of solid dielectric materials. They are

(i) Elemental solid dielectrics
(ii) Ionic non-polar solid dielectrics
(iii) Polar solid dielectrics

Elemental Solid Dielectrics Solid materials consisting of single type of atoms, such as diamond,sulphur, germanium, etc., are said to be elemental solid dielectric materials. These materials exhibit only the electronic polarisability.

Ionic Non-Poplar Solid Dielectrics In ionic crystals such as alkali halides, the total polarisation is ionic and electronic in nature. These solids contain more than one type of atoms, but no permanent dipoles. This type of dielectric materials are said to be ionic non-polar dielectric materials.

Polar Solid Dielectrics In solids, whose molecules possess permanent dipole moments, the total polarisation has all the three components, i.e., it is the sum of electronic, ionic and orientation polarisations. This type of materials are known as polar solid dielectric, e.g. solid $C_6H_5NO_2$.

The McGraw·Hill Companies

21.5.2 Liquid Dielectric Materials

Liquid dielectric materials are also classified into three types as given below.

(i) Mineral insulating oils
(ii) Synthetic insulating oils, and
(iii) Miscellaneous insulating oils

Mineral Insulating Oils Transformer oil, capacitor oil and cable oil are said to be mineral insulating oils. They are directly obtained from crude petroleum by distillation. They are used as coolant in transformers and capacitors.

Synthetic Insulating Oils Synthetic insulating oils are also used for the purpose of cooling in high-tension transformers. Compared to mineral insulating oils, synthetic oils are inferior. Also, they are very cheap, e.g., as karels, sovol, etc.

Miscellaneous Insulating Oils Silicone liquid and vegetable oil are said to be miscellaneous insulating oils. The silicone liquid is costly compared to synthetic insulating oils. It is used in high tension transformers.

Applications Petroleum oils are used for electrical insulation. Nowadays, mineral insulating oils are being replaced by synthetic hydrocarbon oils. Polychlorinated biphenyls (askarels) are widely used, where non-inflammable insulation is required for transformers and capacitors. But, due to environmental pollution, askarels are being withdrawn and silicone liquids are being used for non-inflammable insulation in transformers and capacitors. Castor oil is used as an insulating liquid in the developing countries.

Some of the uses of liquid dielectrics are given as follows.

(i) As a filling and cooling medium in transformers and other electronic equipment
(ii) As a filling medium in capacitors and bushing, etc.
(iii) As an insulating and arc-quenching medium in switch gears
(iv) As an impregnate of absorbent insulation, e.g., paper, press board and porous polymers. They are used in transformers, switchgears, capacitors and cables

21.5.3 Gaseous Dielectrics

Air, nitrogen, hydrogen, sulphur hexafluoride are some of the gaseous dielectric materials used.

Air Air is one of the most important and naturally available gas used for electrical insulation. Its resistivity is infinity under normal conditions, when there is no ionization. The relative permittivity of air is 1. It is used as a dielectric in long-distance electrical transmissions, and in air-capacitors. Compressed air is used as an arc-extinguishing medium which provides dielectric insulation in air-blast circuit breakers.

Nitrogen Nitrogen is chemically inert and is used as a dielectric medium to prevent oxidation. It is used in gas-filled high-voltage cables as an inert medium to replace air in the space above the oil. It is also used in transformers and in low-loss capacitors for high-voltage testing, and so on.

The McGraw·Hill Companies

Sulphur Hexafluoride It is an electronegative gas and is used as a dielectric in X-ray equipment, waveguides, coaxial cables, transformers and as an arc-quenching medium in circuit breakers. Its dielectric strength is nearly 2.3 times higher than that of air or nitrogen. It sublimes at about 209 K and may be used up to a temperature of 423 K. It is non-toxic, non-inflammable and chemically inert.

Hydrogen The dielectric strength of hydrogen is about 65% higher than that of air. It is used as a cooling medium in large turbo-generators and synchronous motors. The injurious effects of hydrogen gas are considered to be negligibly small, since during the discharge it does not produce ozone or oxides of nitrogen. Further, the high voltage discharge of hydrogen is not so severe.

21.6 CLASSIFICATION OF ELECTRICAL INSULATING MATERIALS

Electrical insulating materials are generally in the form of solids, liquids or gases. These materials may be either organic or inorganic and can be either natural or synthetic. The electrical insulating materials, used in generators, motors, transformers and switchgears, etc. are classified into seven temperature classes, according to their temperature limits based on their thermal stability (Table 21.1).

Table 21.1 *Classification of insulating materials*

Temperature K	*Class*
363	*Y*
378	*A*
393	*E*
403	*B*
428	*F*
453	*H*
>453	*C*

Class Y Materials such as un-impregnated paper, cotton or silk, vulcanized natural rubber, aniline, urea and various thermoplastics fall under this class. They can be used up to a temperature of 363 K due to the softening points.

Class A The following are some of the materials that fall under this class: (i) cotton, silk, paper or similar organic materials impregnated with oil or varnish, (ii) laminated materials with cellulose filler phenolic resins, and (iii) a variety of organic varnishes which can be used for wire coating and bonding.

Class E Polyurethane, epoxy resins and varnishes, cellulose triacetate, phenol formaldehyde, polyethylene teraphthalate, etc., are the materials which come under the class *E*.

Class B Mica, glass and asbestos fibres and fabrics bonded and impregnated with suitable organic materials such as shellac bitumen, epoxy, phenol or melamine formaldehide are the materials that come under class *B*.

The McGraw·Hill Companies

Class F Same as class *B*, but with resins which are approved for class *F* operations. For example, alkyd, epoxy and silicone alkyl.

Class H Same as class *B*, but with silicone resins or other resins suitable for class *H* operations. For example, silicone rubber.

Class C Mica, asbestos, ceramics, glass, quartz and similar inorganic materials such as polytetrafluoro ethylene and this provide ethylene are the materials under class *C*. Let us discuss some of these solid insulating materials.

- ***Mica*** Mica is an inorganic compound of silicates of aluminium, soda potash and magnesia. It is crystalline in nature and can be splitted into thin sheets. Muscovite and phlogopite are the two important types of mica. Muscovite mica is one of the best insulating materials. Its dielectric constant varies from 5 to 7.5 and its loss tangent varies from 0.0003 to 0.015. The dielectric strength of the mica lies between 700 and 1000 kV mm^{-1}. It is used as an insulating material in electrical machines, switch gears, armature windings, and electrical heating devices like electric iron and hot plates.
- ***Glass*** Glass is a brittle and amorphous material. The essential structural constituent of most of the commercial glasses is SiO_2. The dielectric constant of glass varies from 3.7 to 10 and its dielectric strength lies between 3000 and 5000 V m^{-1}. The dielectric loss of glass is more complicated. It is used as an internal support in articles like electric light bulbs, electronic valves, mercury switches, X-ray tubes, etc. It is also used as a dielectric material in capacitors.
- ***Ceramics*** Ceramics are brittle and can be either amorphous or crystalline. The dielectric and mechanical properties of ceramics are excellent. Their dielectric constant varies from 4 to 10. Ceramics are used as insulators in switches, plug-holders and cathode heaters. They can also be used as dielectric materials in capacitors. Porcelein, alumina and titanates are some of the examples of ceramics.
- ***Asbestos*** It is an inorganic material made of magnesium silicate. It has good dielectric and mechanical properties. It is used as an insulator in the form of paper tape, cloth and board.
- ***Rubber*** Rubber is an organic polymer and is available both in natural and synthetic forms. The dielectric constant of rubber varies from 2.5 to 5 and its loss tangent lies between 0.01 and 0.03. It is widely used in electric wires, cables, tapes, coatings, transformers, motor windings, etc.

21.7 CLAUSSIUS-MOSOTTI EQUATION

Elemental solid dielectrics like germanium, silicon, diamond, sulphur, etc. have no permanent dipoles or ions and will have only cubic structure, thereby exhibiting only electronic polparisation. The total polarisiation is written as,

$$P = N \alpha E_i$$

where $$\alpha = \alpha_e + \alpha_i + \alpha_0$$

Elemental solid dielectrics have only electronic polarisation and the remaining polarisation, namely ionic and orientation, are equal to zero, since there is no ionic character or permanent dipoles.

Therefore, the total polarisation

$$P = N \alpha_e E_i \tag{21.43}$$

where $\alpha_i = \alpha_0 = 0$, and N is the number of atoms per unit volume.

Substituting the value of internal field, Eq. (21.43) can be written as

$$P = N \alpha_e \left[E + \frac{P}{3\varepsilon_0} \right] \tag{21.44}$$

We know that

$$D = \varepsilon_0 E + P$$

$$\frac{P}{E} = \frac{D}{E} - \varepsilon_0 \tag{21.45}$$

From the definition of electric displacement vector,

$$D = \varepsilon E$$

Therefore,

$$\frac{P}{E} = \varepsilon - \varepsilon_0 = \varepsilon_r \varepsilon_0 - \varepsilon_0$$

where $\varepsilon_r = \varepsilon/\varepsilon_0$

$$\frac{P}{E} = \varepsilon_0(\varepsilon_r - 1)$$

or $$P = E \varepsilon_0 (\varepsilon_r - 1) \tag{21.46}$$

Substituting P value in Eq. (21.44), we get,

$$E \varepsilon_0 (\varepsilon_r - 1) = N \alpha_e \left[E + \frac{E \varepsilon_0 (\varepsilon_r - 1)}{3\varepsilon_0} \right]$$

$$\varepsilon_0 (\varepsilon_r - 1) = N \alpha_e \left[1 + \frac{\varepsilon_0 (\varepsilon_r - 1)}{3\varepsilon_0} \right]$$

$$N \alpha_e = \frac{\varepsilon_0 (\varepsilon_r - 1)}{\left[1 + \frac{\varepsilon_0 (\varepsilon_r - 1)}{3\varepsilon_0} \right]}$$

$$\frac{N \alpha_e}{3\varepsilon_0} = \frac{\varepsilon_0 (\varepsilon_r - 1)}{3\varepsilon_0 + \varepsilon_0 \varepsilon_r - \varepsilon_0}$$

Simplifying the above equation, we get,

$$\frac{N \alpha_e}{3\varepsilon_0} = \frac{\varepsilon_r - 1}{\varepsilon_r + 2} \tag{21.47}$$

where N is the number of dipoles per unit volume. Equation (21.47) is known as *Claussius–Mosotti equation*. In this equation, by substituting the values of ε_r, ε_0 and N, one can determine the electronic polarisability α_e.

The McGraw·Hill Companies

For a dielectric material consisting of N number of dipoles

$$\frac{\varepsilon_r - 1}{\varepsilon_r + 2} = \frac{1}{3\varepsilon_0} \sum_i N_i \alpha_e \tag{21.48}$$

where N_i and a_{ei} are the appropriate quantities for the types of atoms or molecules.

21.8 EXPERIMENTAL DETERMINATION OF DIELECTRIC CONSTANT

The Schering bridge is used to determine the dielectric constant of any solid material experimentally. It is one of modified forms of Wheatstone's bridge. The experimental arrangement of Schering bridge is shown in Fig.21.9. Consider that *P, Q, R* and *S* are the four arms of the Schering bridge. The arm *P* consists of a fixed resistance R_1 and a variable capacitor C_1, while the arm *S* is made by connecting a resistor R_4 and variable capacitor C_4 in parallel. The arms *Q* and *R* consist of respectively the variable resistor R_3 and the fixed capacitor C_2. The high frequency oscillator *AC* is used to give necessary potential to the bridge circuit. The balanced condition of the bridge is obtained by adjusting the values of R_3 and C_4 in order to get a null deflection in the galvanometer *G*.

The dielectric material whose dielectric constant is to be determined is placed in between the parallel plate capacitor C_1. The dielectric constant of the material is determined by measuring the change in capacitive value of capacitor C_1 before and after introducing the material in between the parallel plate of $C_{1.}^{1}$

Therefore, the capacitance of a capacitor $C_{1.}^{1} = \dfrac{\varepsilon A}{d}$ (21.49)

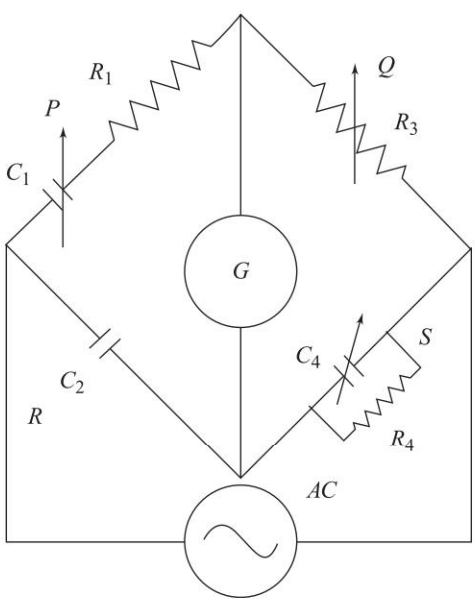


Fig. 21.9 Scherring bridge—Dielectric Constant

where ε is the permittivity of medium present between the parallel plates, d the thickness of the sample and A the area of cross-section. After measuring the capacitance of the capacitor, the dielectric material is removed. Therefore, the capacitance of the capacitor C_1 is again determined, and is equal to

$$C_1 = \frac{\varepsilon_0 A}{d} \tag{21.50}$$

where ε_0 is the permitively of free space.
The dielectric constant of a material is equal to the ratio of the capacitance of the capacitor with material and free space.

$$\text{Dielectric constant} = \frac{C_1^1}{C_1} \tag{21.51}$$

Substitute the values of C_1 and C^1_1, we get,

$$\text{Dielectric constant} = \frac{\varepsilon A}{d} \times \frac{d}{\varepsilon_o A}$$

$$= \frac{\varepsilon}{\varepsilon_o}$$

$$\text{Therefore, the dielectric constant } \varepsilon_r = \frac{\varepsilon}{\varepsilon_0} \tag{21.52}$$

The value of C_1 and C^1_1 of the capacitor are determined by considering the balanced condition of Wheatstone bridge network, i.e.,

$$\frac{P}{Q} = \frac{R}{S} \tag{21.53}$$

We know that from Fig. 21.9,

$$P = r_1 + \frac{1}{j\omega C_1} \tag{21.54}$$

$$Q = r_3 \tag{21.55}$$

$$R = \frac{1}{j\omega C_1} \tag{21.56}$$

$$S = C_4 \,||\, r_4 = \frac{C_4 r_4}{C_4 + r_4} \tag{21.57}$$

Substituting the value C_4 in the above equation, we get,

$$S = \frac{\frac{r_4}{j\omega C_4}}{\frac{r_4}{j\omega C_4} + r_4}$$

$$= \frac{r_4}{1 + r_4 j\omega C_4} \tag{21.58}$$

The McGraw·Hill Companies

Substituting the values of *P, Q, R* and *S* in Eq. (21.53), we get,

$$\frac{r_1 + \dfrac{1}{j\omega C_1}}{r_3} = \frac{\dfrac{1}{j\omega C_2}}{\dfrac{r_4}{1 + j\omega C r_4}}$$

Rearranging the above equation, we get,

$$\frac{r_3}{j\omega C_2} = \frac{r_1 r_4}{1 + j\omega C_4 r_4} + \frac{r_4}{j\omega C_1}\left[\frac{1}{1 + j\omega C_4 r_4}\right]$$

or $$\frac{r_3(1 + j\omega C_4 r_4)}{j\omega C_2} = r_1 r_4 + \frac{r_4 C_2}{C_1}$$

or $$r_3 + j\omega\, C_4 r_4 r_3 = r_1\, r_4 j\omega\, C_2 + \frac{r_4 C_2}{C_1} \tag{21.59}$$

Equating the real parts in the above equation, we get,

$$r_3 = \frac{r_4 C_2}{C_1} \tag{21.60}$$

Therefore, $$C_1 = \frac{r_4}{r_3} C_2 \tag{21.61}$$

The resistance r_4 and capacitor C_2 are fixed. Therefore, the value of C_1 is determined by varying the variable resistance r_3.
Therefore, dielectric constant of material,

$$\varepsilon_r = \frac{C^1_1}{C_1} \tag{21.62}$$

where C^1_1 is the capacitance of the capacitor with dielectric material and C_1 the capacitance of capacitor with air as the dielectric.

Equation (21.62) is used to determine the dielectric constant of any material experimentally using Schering bridge method.

21.9 DIELECTRIC LOSS

When an ac field is applied to a dielectric material, some amount of electrical energy is absorbed by the dielectric material and is wasted in the form of heat. This loss of energy is known as dielectric loss.

The dielectric loss is a major engineering problem. In an ideal dielectric, the current leads the voltage by an angle of 90° is shown in Fig. 21.10. But in case of a

commercial dielectric, the current does not exactly lead the voltage by 90°. It leads by some other angle q that is less than 90°. The angle $f = 90 - q$ is known as the dielectric loss angle.

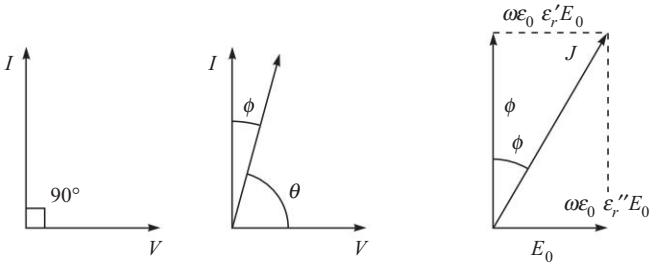


Fig. 21.10 *Relation between current and voltage in dielectrics*

For a dielectric having capacitance C and voltage V applied to it at a frequency f Hz, the dielectric power loss is given by

$$P = VI \cos \theta$$

Since $I = V/X_c$ where X_c is the capacitive reactance and is equal to $1/j\omega C$.

Therefore,
$$P = \frac{V^2}{X_c} \cos (90 - \Phi) = V^2 j\omega C \sin \Phi \qquad (21.63)$$

Since θ is very small, $\sin \Phi = \tan \Phi$ and $P = j\ V^2\ \omega C \tan \Phi$, where $\tan \Phi$ is said to be the power factor of the dielectric. The power loss depends only on the power factor of the dielectric as long as the applied voltage, frequency and capacitance are kept constant.

The dielectric loss is increased by the following factors:

(i) High frequency of the applied voltage
(ii) High value of the applied voltage
(iii) High temperature, and
(iv) Humidity

The frequency dependence of the dielectric loss is shown in Fig. 21.11.

The dielectric losses in the radio frequency region are usually due to dipole rotation. The dielectric losses at lower frequencies are mainly due to dc resistivity. The dielectric losses in the optical region are associated with electrons and they are known as optical obsorption.

21.10 DIELECTRIC BREAKDOWN

When a voltage is appplied to a dielectric material and thereby the electric field is increased, it can withstand up to a certain maximum voltage before it permits large currents to pass through it. This phenomenon in which the dielectric material fails to

The McGraw·Hill Companies

offer insulation resistance for large applied voltage is known as dielectric breakdown and the corresponding voltage is known as breakdown voltage.

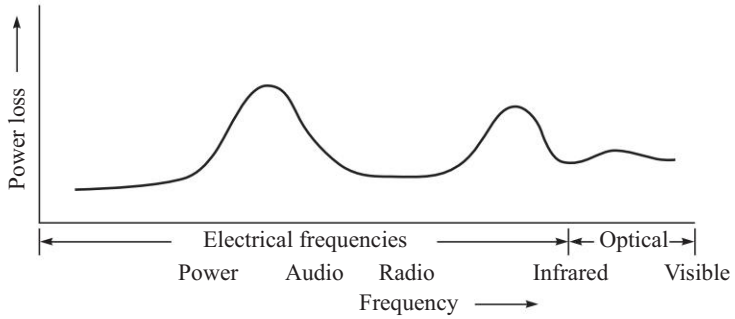


Fig. 21.11 *Frequency versus dielectric loss*

The different types of dielectric breakdown are given below.

(i) Intrinsic breakdown
(ii) Thermal breakdown
(iii) Electrochemical breakdown
(iv) Defect breakdown, and
(v) Discharge breakdown

21.10.1 Intrinsic Breakdown

If a dielectric material is subjected to a large amount of electric field, then the electrons in the valence band acquire sufficient amount of energy so that they easily cross the energy gap. These electrons while flowing through the conduction band produce a large current. This large current causes the dielectric material to undergo a breakdown. The large number of conduction electrons in the conduction band may collide with the covalent band and consequently, they dislodge more number of electrons. These dislodged electrons, in turn, dislodge some other electrons. Thus, this process goes on as a chain reaction. The breakdown produced in this way is said to be intrinsic breakdown or avalanche breakdown.

Characteristics

(i) It can occur even at low temperature.
(ii) It requires relatively large electric field.
(iii) This kind of breakdown occurs mostly in thin samples.
(iv) It does not depend on the electrode configuration and shape of the material.

21.10.2 Thermal Breakdown

In a dielectric material when an electric field is applied, some amount of heat is produced. The heat produced in the material should be dissipated from it. In some cases, the amount of heat generated is very high compared to the heat dissipated and

The McGraw·Hill Companies

this excess of heat may produce breakdown of the dielectric material. This type of breakdown is known as thermal breakdown.

Characteristics

(i) It can occur only at high temperatures.
(ii) The strength of the electric field to create dielectric breakdown depends upon the material's size and shape.
(iii) The breakdown time is in the order of few milliseconds.

21.10.3 Electrochemical Breakdown

The chemical and electrochemical breakdowns have close relationship with thermal breakdown. If the temperature of a dielectric material increases, it will increase the mobility of the ions and hence, electrochemical reaction will take place. When ionic mobility increases, leakage current will increase, thereby decreasing the insulation resistance, and this will result in dielectric breakdown.

In case of rubber, oxides are produced in air and hence, it gradually loses its dielectric property. If one can bring rubber into ozone atmosphere, some cracks are produced.

Characteristics

(i) Electrochemical breakdown is determined by the leakage current, density of ions, temperature and permanent dipoles in the material.
(ii) To avoid electrochemical breakdown, impurities should not be mixed with the pure dielectric materials.
(iii) Electrochemical breakdowns are accelerated by temperature. To avoid electrochemical breakdown, the dielectric material should not be operated at high temperatures.

21.10.4 Discharge Breakdown

Some dielectric materials may have occluded gas bubbles. If these dielectric materials are subjected to high voltages, the gaseous substances are easily ionised and they produce a large ionisation current. This large ionisation current may produce dielectric breakdown. This type of breakdown is known as discharge breakdown.

Characteristics This type of breakdown can occur at low voltages where there are large number of occluded gas bubbles.

21.10.5 Defect Breakdown

Some dielectric materials may have surface defects like cracks, pores, etc. Moisture and other impurities can get filled up at these places leading to breakdown. This is known as defect breakdown.

21.11 FERROELECTRIC MATERIAL

In certain dielectric materials, polarisation is not a linear function of applied electric field. Such materials exhibit hysteresis curve similar to that of ferromagnetic materials and are known as *ferroelectric materials*. The hysteresis curve exhibited by a ferroelectric material is shown in Fig. 21.12.

The McGraw·Hill Companies

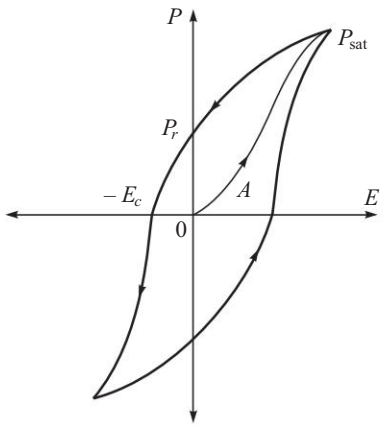


Fig. 21.12 *Hysteresis curve for ferroelectric material*

Hysteresis Properties When an increasing electric field is applied to a ferroelectric material, it results in an increase in polarisation and it reaches a maximum value for a particular field strength. On the other hand, if we decrease the electric field, the polarisation decreases. When the field strength is zero, i.e., $E = 0$, a small amount of polarisation exists in the material. This polarisation is known as *remanent polarisation*. Now the ferroelectric material is said to be spontaneously polarised. In order to reduce the value of polarisation to zero, an electric field strength ($-E_c$) should be applied. This field is known as *coercive field.*

The hysteresis loop of a ferroelectric material is explained on the basis of the domain concept. Domain is experimentally observed by applying polarised light, which makes the domain visible. Figure 21.13 explains the domain structure of barium titanate crystal.

Domains which are opposite in direction to the applied field decrease in size and domains which are parallel to the field direction increase in size in the form of needles of approximately 10^{-6} m width. Substances like Rochelle salt (sodium-potassium salt of tartaric acid; $NaKC_4H_4O_2 1.4H_2O$), potassium dihydrogen phosphate (KH_2PO_4), barium titanate ($BaTiO_3$), etc., are typical examples. Ferroelectric crystals lack centre of symmetry.

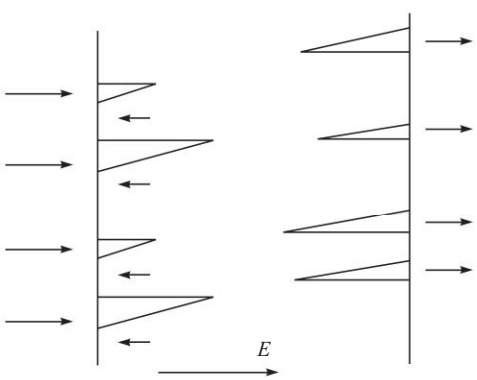


Fig. 21.13 *Domain structure of barium titanate*

Classification of Ferroelectric Crystals Ferroelectric crystals are classified into three groups namely

(i) Rochelle salt
(ii) Potassium dihydrogen phosphate, and
(iii) Barium titanate

(i) Rochelle Salt It behaves as a ferroelectric material in the temperature range from 255 to 296 K, i.e., it possesses two transition temperatures (Curie temperature). The crystal structure is orthorhombic above upper curie temperature. Below this transition temperature, the crystal structure is monoclinic in nature. A plot of polarisation versus temperature is shown in Fig. 21.14.

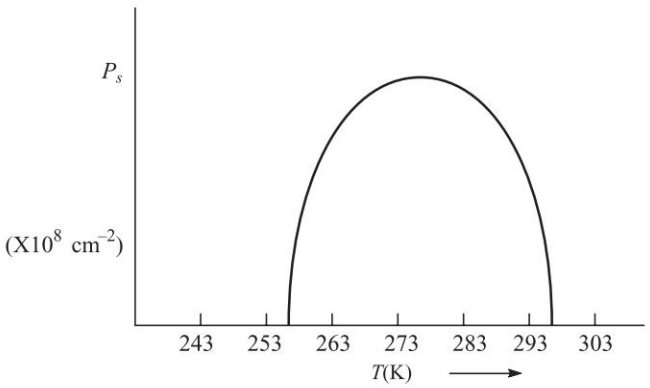


Fig. 21.14 *Temperature versus polarisation of Rochelle salt*

(ii) Potassium Dihydrogen Phosphate It has only one Curie temperature and its value is 123 K. It possesses ferroelectric property below this temperature. Above Curie temperature, its crystal structure is tetragonal. Below Curie temperature, the crystal structure is orthorhombic. Other ferroelectric crystals belonging to this family include KD_2PO_4, RbH_2,PO_4, CsH_2AsO_4, CsD_2AsO_4. The variation of polarisation with respect to temperature is shown in Fig. 21.15.

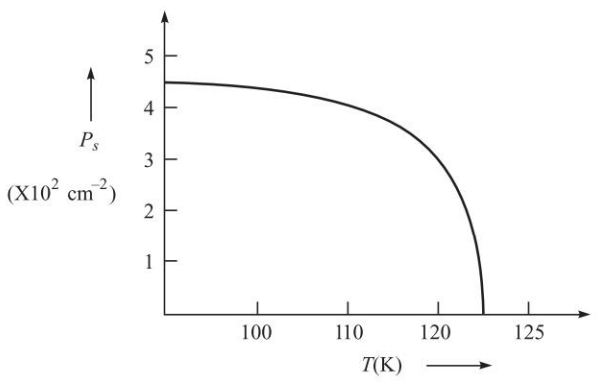


Fig. 21.15 *Polarisation versus temperature in potassium dihydrogen phosphate*

The McGraw·Hill Companies

(iii) Barium Titanate This crystal exhibits three different ferroelectric phases. The transition temperatures of barium titanate are 278 K, 193 K and 393 K. The structure of barium titanate is cubic above 393 K. It has orthorhombic structure when the temperature lies between 278 K and 193 K. Below 193 K, it has rhombohedron structure. The structure of $BaTiO_3$ above 393 K and the variation of polarisation with temperature are given, respectively, in Figs 21.16 and 21.17.

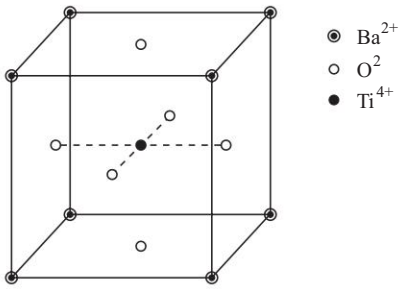


Fig. 21.16 *Crystal structure of barium titanate (above 393 K)*

21.12 DIELECTRIC PROPERTIES

The following are some of the key properties of dielectric materials that determine their suitability for specific applications.

(i) Relative Permittivity Relativie permittivity of a material may be expressed by ε_r relative to that of vacuum by $\varepsilon_r = \varepsilon/\varepsilon_0$. Since the dielectric constant is frequency dependent, it should be taken into account while selecting a dielectric material for a particular use.

(ii) Dielectric Strength The dielectric strength is the maximum electric field that can be maintained in the dielectric. The dielectric strength depends on the material, its thickness, the shape and size of the conductors between which the dielectric is placed, moisture content of the material, temperature and pressure.

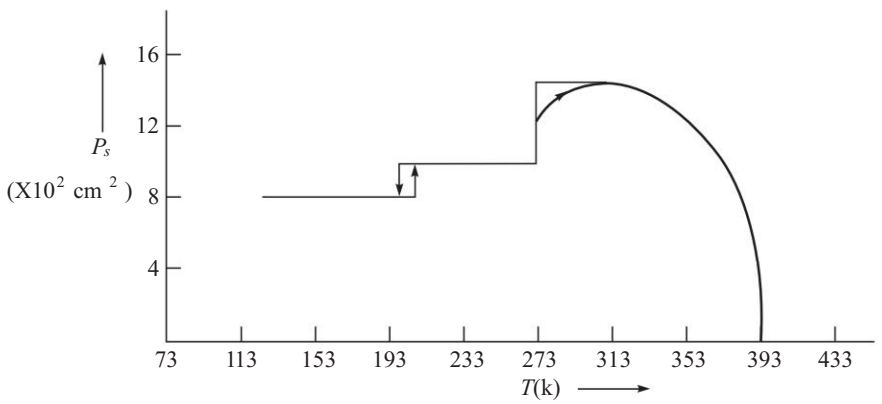


Fig. 21.17 *Polarisation versus temperature in barium titanate*

The McGraw·Hill Companies

(iii) Dielectric Losses The extent to which a dielectric in an alternating electric field loses energy is indicated by the loss factor, tan ϕ. A very good dielectric would have a loss factor of the order of 10^{-5} with a poor one of about 0.1.

(iv) Mechanical Properties In selecting a dielectric there may be a need to consider the mechanical properties of the material. Many of the dielectrics are very brittle. Properties which may need to be considered are the effects of bending, impact strength, tearing strength, tensile strength, compressive strength, shear strength, and machinability.

(v) Insulation Resistance Dielectrics need to have a high insulation resistance in order to prevent leakage of charge. The resistance of these materials should be very high.

(vi) Temperature Effects One of the main effects of an increase in temperature is a decrease in resistivity and higher leakage currents through the dielectrics. The leakage current increases with the increase in temperature and it will produce breakdown in the dielectric material.

Other thermal properties that need to be considered in selecting a dielectric are melting points, freezing points for liquid dielectrics, thermal expansivity, specific heat capacity and ageing effect.

21.13 ACTIVE AND PASSIVE DIELECTRICS

A dielectric material which exhibits gain (current or voltage or both) and has directional electronic properties is said to be active dielectrics. Piezoelectric, ferroelectric and several other voltage induced or voltage inducing materials are said to be active dielectrics. For example, quartz is an active dielectric material even though it is a piezoelectric material. Similarly, a dielectric material which does not exhibit any gain or directional property is said to passive dielectrics. Most of the insulating materials like ceramics, mica, etc., belongs to this category.

21.14 FREQUENCY AND TEMPERATURE DEPENDENCE OF DIELECTRIC PROPERTIES

The dielectric properties such as polarisation, dielectric constants and dielectric loss depend on the frequency of the applied field and temperature. Let us discuss the frequency and temperature dependence of polarisation and dielectric constant.

21.14.1 Frequency Dependence of Dielectric Properties

Polarisation When an ac field is applied to a dielectric material, the polarisation of the dielectric material is a function of time. It follows the equation

$$P(t) = P\left[1 - \exp\left(-\frac{t}{\tau_r}\right)\right] \tag{21.64}$$

The McGraw·Hill Companies

where P is the maximum polarisation attained by a dielectric due to the prolong application of the electric field, and τ_r the relaxation time for the particular polarisation process, is the average time between molecular collisions (in case of liquid), during the application of the electric field. The relaxation time is a measure of the polarisation process and is the time taken for a polarisation process to reach 67% of the maximum value.

From Fig. 21.18, it can be inferred that the space charge polarisation occurs at very low frequencies (less than 10^2 Hz). The space charge polarisation refers to the diffusion of metal ions over several interatomic spacing. It also occurs at machine frequencies in the order of 50–60 Hz. The space charge polarisation is one of the slowest process compared to the other types of polarisation.

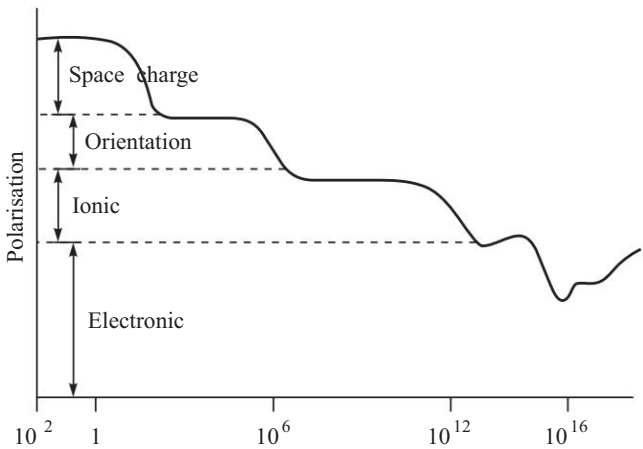


Fig. 21.18 *Frequency dependence of the polarisation processes*

The orientation polarisation mechanism arises in the polar molecules when an ac field is applied. This polarisation mechanism is faster than space charge polarisation but is slower than the ionic and electronic polarisations. In many liquid dielectrics, the orientation polarisation occurs between the radio wave and microwave frequency range at room temperature. In some polymeric liquids, it involves a limited rotation of molecules and occurs even at low frequencies.

The ionic polarisation mechanism occurs in the ionic molecules, whenever the frequency of the applied field is less than 10^{13} Hz. Ionic polarisation does not occur at optical frequencies ($\approx 10^{15}$ Hz), since the ionic molecules require time nearly 100 times greater than the frequency of the applied field. Ionic polarisation is faster than the orientation and space charge polarisations, but it is slower than the electronic polarisation.

Electronic polarisation is the fastest polarisation and it occurs immediately after the supply is switched on. This polarisation occurs even at optical frequencies.

Dielectric Constant The frequency dependence of the real and imaginary part of dielectric constant is shown in Fig. 21.19. The curve is similar to that of the frequency dependence of polarisation as shown in Fig. 21.18. Figure 21.19 shows distinct

peaks in ε_r and some distinct features in ε_r', but in reality these peaks and features are broader. The variations in ε_r' and ε_r'' are produced by the presence of different polarisation mechanisms.

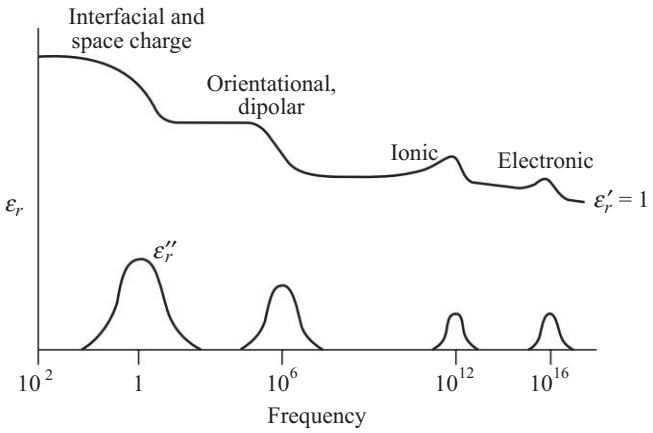


Fig. 21.19 *Frequency dependence of dielectric constant*

21.14.2 Temperature Dependence of Dielectric Properties

Orientation polarisation is produced due to the alignment of dipoles in polar molecules in the presence of electric field. The increase in temperature disrupt the alignment of dipole due to the electric field. This effect decreases the value of dielectric constant with increase in temperature. For example, the dielectric constant of solid HCl decreases from 19 to 14, when the temperature is increased from 100 to 160 K. At the melting point of 160 K, it drops to 12. This shows that the number of molecules per unit volume decreases due to the expansion on melting.

Solid nitrobenzene is a polar molecule, but it does not exhibit orientation polarisation in its solid state because the rotation of the molecules is prevented under an applied field. But in liquid state, the molecules have sufficient energy to reorient themselves in the applied field. The relative dielectric constant increases from 3 to 37 on melting. This is in contrast with the results obtained for solid HCl. The dielectric properties of four insulators are given in Table 21.2.

Table 21.2 *Dielectric properties of different insulators at two different frequencies*

Material	*f = 60 Hz*		*f = 1 MHz*	
	ε_r'	tan δ	ε_r'	tan δ
Polycarbonate	3.17	9×10^{-4}	2.96	1×10^{-2}
Silicone rubber	3.7	2.25×10^{-2}	3.4	4×10^{-3}
Epoxy with mineral filler	5.0	4.7×10^{-2}	3.4	3×10^{-2}
Alumina	8.5	1×10^{-3}	8.5	1×10^{-3}

The McGraw·Hill Companies

In case of the polymer film Polyethyleneterephthalate (PET), the dielectric constant ε_r' measured at $f = 10$ kHz is found to increase with increase in temperature as shown in Fig. 21.20.

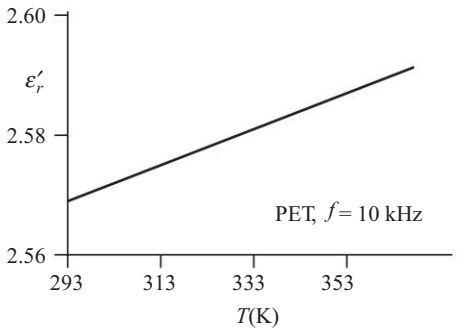


Fig. 21.20 *Variation of ε_r' with temperature in PET material*

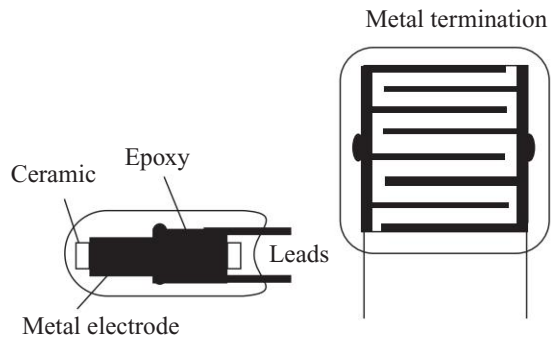
21.15 USES OF DIELECTRIC MATERIALS

Some of the most important properties of insulating materials possess excellent mechanical and dk'lectrical strength, high thermal and electrical stability, etc. In view of the above properties, dielectrical materials are used as insulating materials in applications like power distribution transformers, capacitors, cables, transmission equipment, etc.

21.15.1 Capacitors

Dielectric materials are used to manufacture capacitors of different ranges. The selection of right dielectric materials for capacitor applications is based on various parameters like value of capacitance, frequency, tolerance lens, size and operating voltage. Dielectric materials, paper and plastic film, mica film, single layer ceramic, multilayer ceramic, solid electrolytic (Al, Ta) and electrolytic (AI, Ta) are used in capacitors to obtain a range of capacitance 1 pf to 104 μf

Single and Multilayer Dielectric Capacitors The schematic representation of a single layer capacitor is shown in Fig. 21.21. It consists of a thin ceramic disk or plate and metal electrodes. The ceramic plate is placed in-between the metal electrodes. The leads for the electrical connections are taken from the metal electrode. In order to prevent the degradation of dielectric properties of ceramic plate due to moisture the whole set-up is encapsulated in an epoxy coating. The capacitance of single layer capacitor for an area of 1 cm2 is 885 pf. One can increase the capacitance of the capacitor by connecting N number of ceramic plate in parallel by providing a sufficient space as shown in Fig. 21.21(b). The capacitor with stacking of ceramic plates in different layers is known as multilayer capacitor. The capacitance of multilayer capacitor is in the range of few hundred microfarads.



(a) Single-layer ceramic capacitor (e.g. disk capacitors)

(b) Multilayer ceramic capacitor (stacked ceramic layers)

Fig. 21.21 *Dielectric capacitors*

Polymeric Film Capacitors Polymer thin film capacitors are constructed to work as mid frequency capacitors. Two separate thin polymer sheets metal coated by having a small margin on one side employing vacuum deposition. The two metal coated films are placed in parallelby keeping the margin in the opposite side. Thus, the two margins are opposite in directions. In order to get the capacitor, the two metal sheets are rolled together like Swiss roll. The electrical connections are taken from the opposite sides of the metals by proper soldering. One can also produce multilayer polymeric thin film capacitors like multilayer ceramic capacitors.

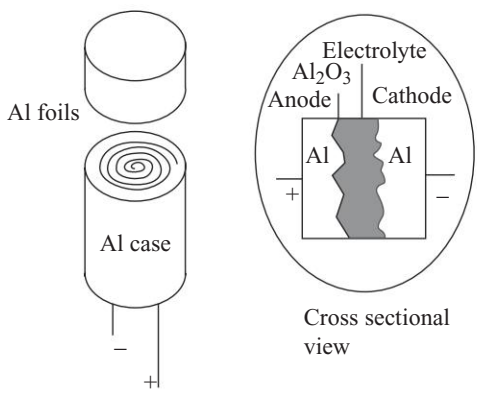
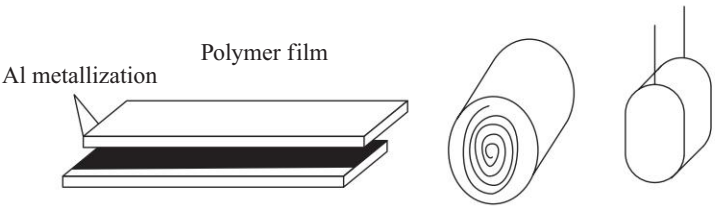


Fig. 21.22 *Aluminum electrolytic capacitors*

Electrolytic Capacitors Large values capacitors are achieved using electrolytic capacitors. Electrolytic capacitors are of different types namely, aluminum electrolytic and solid electrolyte tantalum capacitors. A cross-sectional view of aluminum electrolytic capacitors is shown in Fig. 21.22. The Al electrolytic capacitor consists of two Al foils.

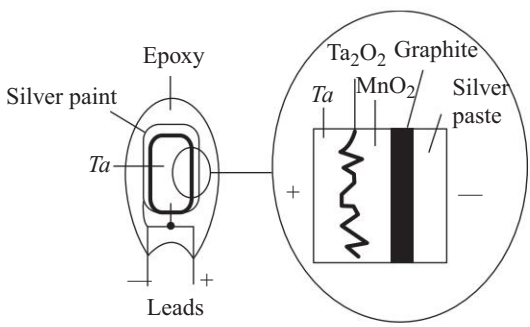
A thin layer of Al_20_3 as a dielectric medium is grown on the rough surface of the one of the foils. The two foilsarewound together in the form of disk and kept inside the cylindrical case as shown in Fig. 21.3. The AC foil coated with Al_20_3 dielectric act as an anode while /tKe~uncoatedi AlfoTacTas a cathode. The thickness of AljO, grown on the foil is 0.1 *fim* and hence it gives large capacitance value (i.e, small thickness and targe area). The AljOj is grown on the Al foil eleeterolytically and hence, this capacitor is known as electrolytic capacitor.



(a) Aluminium Coating on film *(b) Capacitor*

Fig. 21.23 *Aluminum electrolytic capacitors*

The structure of a solid electrolyte tantalum capacitor is shown in Fig. 21.4 (a). The anode is made up of tantalum (Ta) pellet. The surface of Ta pellet is anodized to get the tantalum pentoxide (Ta_2O_5). A thick solid electrolyte like MnOj is coated above Ta_2O_s The process is complete'd by coating graphite and silver paste as shown in Fig. 21.4 (b). The above system is then coated by an epoxy. Tantalum capacitors are widely used in) electronic applications.



(a) Solid electroyte tantalum capacitors *(b) Cross sectional view*

Fig. 21.24 *Aluminum electrolytic capacitors*

Power and Distribution Transformers The following are the application of insulating materials in power and distribution of transformers:

- Transformer oil is used as a liquid dielectric and coctat
- Electrical grade paper is used as layer binding insulation such as a cover to cover the cable and as a backing paper is used for applications such as top and bottom coil clamping ring, washer etc.
- The insulation tape is used for applications such as taping and banding, core insulation, banding of transformer cores, etc.
- Materials such as synthetic rubbers are used as a gasket in different places to percent oil leakages from joints

In order to meet the above requirements the insulahng/dielectric material requires the following properties:

- Excellent electrical properties to withstand high voltages.
- High dielectric strength and adequate chemical stability.
- Very high moisture resistance.

21.16 APPLICATION

Some of the applications of dielectric materials are given below.

(i) Quartz crystal is used for the preparation of ultrasonic transducers, crystal oscillators, delay lines, filters, etc.
(ii) Barium titanate is used for the preparation of accelerometers.
(iii) Lead zirconate titanate ($PbZr_xTi_{1-x}O_3$) is used for the preparation of earphones, microphones, spark generators (gas lighter, car ignition, etc.), displacement transducer, accelerometers, etc.
(iv) The insulating dielectric liquids are used in transformers, switchgears and generators.
(v) Dielectric materials are used as insulating material in power cables, signal cables, electric motors, electric iron, etc.
(vi) Dielectric materials are used in radiation detectors, thermoionic valves, strain gauges, capacitors, resistors and many other electric devices, and
(vii) The electro-optic devices are prepared using dielectric material.

Keypoints to Remember

- The force per unit charge is known as electric field strength E.
- The permittivity of free space, i.e., vacuum $\varepsilon_o = 8.854 \times 10^{-12}$ F m^{-1}.
- The electrical flux density is defined as the number of flux lines crossing a surface normal to the lines divided by the surface area.
- The dielectric constant of a material is defined as the ratio of permittivity of the medium to permittivity of free space $\varepsilon_r = \dfrac{\varepsilon}{\varepsilon_o}$.
- The product of magnitude of charge and distance of separation is known as electric dipole moment.
- Polarization is defined as induced dipole moment per unit volume.
- There are four different types of polarizations, namely, electronic, atomic, orientation and interfacial polarization.
- The electronic polarization of a solid materials which contains large number of atoms is equal to $P_e = N 4 \pi \varepsilon_o R^3 E$.
- An electronic polarization is polarization produced due to the displacement of electron, when an electric field is applied to the atom.
- An atomic or ionic polarization is polarization due to the displacement of positive ions away from the field and the negative ions towards the field due to the application of an electric field.
- The internal field or local field in a solid dielectric material is equal to the sum of the applied field plus the field produced at the location of an atom by dipoles of all other atoms.

The McGraw·Hill Companies

- Dielectric materials are classified into solid dielectric, liquid dielectric and gaseous dielectric materials.
- The Claussius–Mosotti equation for a solid material is equal to

$$\frac{N\alpha_e}{3\varepsilon_0} = \frac{\varepsilon_r - 1}{\varepsilon_r + 2}$$

- Dielectric loss is the loss of energy in the form of heat wasted due to absorption of electrical energy by a dielectric material when an ac field is applied to the material.
- Dielectric breakdown is the phenomenon in which the dielectric material fails to offer insulation resistance for a large applied voltage.
- There are five different types of dielectric breakdown, namely, intrinsic, thermal, electrochemical, defect and discharge.
- The dielectric materials which exhibit hysteresis curve similar to that of ferromagnetic materials are known as ferroelectric materials.
- Ferroelectric materials are classified into three major classes as Rochelle salt, potassium dihydrogen phosphate and barium titanate.
- Active dielectric materials are the materials which exhibit gain in current or voltage or both and have directional electronic properties.
- Passive dielectric materials do not exhibit any gain or directional properties.

Solved Problems

Example 21.1 The lattice parameter of KCl is 0.629 nm. It crystallizes like the NaCl crystal structure. The electronic polarisability of K^+ is 1.264×10^{-40} F m^2 and that of Cl^- is 3.408×10^{-40} F m^2. Calculate the relative permittivity of KCl crystal at optical frequencies.

Given Data

Lattice parameter of KCl = 0.629 nm
The electronic polarisability, α_e, for K^+ = 1.26×10^{-40} F m^2
The electronic polarisability, α_e, for Cl^- = 3.408×10^{-40} F m^2

Solution The electronic polarisability a_e for KCl = (α_e for K^+ + α_e for Cl^{-1})

$$= 1.264 \times 10^{-40} + 3.408 \times 10^{-40}$$

$$= 4.672 \times 10^{-40} \text{ F m}^2$$

The number of dipoles per m^3 (N) $= \dfrac{\text{Number of atoms}}{\text{Volume of the unit cell}}$

$$= \frac{4}{(0.629 \times 10^{-9})^3}$$

$$= 1.607 \times 10^{28} \text{ atoms m}^{-3}.$$

The dielectric constant is given by

$$\varepsilon_0 (\varepsilon_r - 1) = N\alpha_e$$

$$\varepsilon_r = \frac{N\alpha_e}{\varepsilon_0} + 1$$

$$= \frac{1.607 \times 10^{28} \times 4.672 \times 10^{-40}}{8.854 \times 10^{-12}} + 1$$

$$= 1.8479$$

The dielectric constant of KCl is 1.848.

Example 21.2 Calculate the electronic polarisability of an isolated Se atom. The atomic radius of an Se atom is 0.12 nm.

Given Data

The atomic radius of Se = 0.12 nm

Solution The electronic polarisability is given by

$$\alpha_e = 4\pi\varepsilon_0 R^3$$

$$= 4\pi \times 8.854 \times 10^{-12} (0.12 \times 10^{-9})^3$$

$$= 1.9226 \times 10^{-40} \text{ F m}^2$$

The electronic polarisability of an isolated Se is 1.9226×10^{-40} F m^{-2}.

Example 21.3 The following data refers to a dielectric material. $\varepsilon_r = 4.94$ and $n^2 = 2.69$, where n is the index of refraction. Calculate the ratio between electronic and ionic polarability for this material.

Given Data

The dielectric constant $\varepsilon_r = 4.94$

The refraction index $n^2 = 2.69$

Solution

We know that, the Clasius - Mosotti relation

$$\frac{N\alpha}{3\varepsilon_0} = \frac{\varepsilon_r - 1}{\varepsilon_r + 2}$$

(i) At optical frequencies $\varepsilon_r = n^2$, the electronic polarisation alone occurs at optical frequencies

Therefore, $$\frac{N\alpha}{3\varepsilon_0} = \frac{\varepsilon_r - 1}{\varepsilon_r + 2}$$

We know that the electrical polariability

$$\alpha = \alpha_e + \alpha_i + \alpha_o$$

Since the orientation polariability (α_0) is very less it can be neglected.

Therefore, $$\alpha = \alpha_e + \alpha_i$$

Substituting Clausius–Mosotti relation, we get,

$$\frac{N(\alpha_e+\alpha_1)}{3\varepsilon_0}=\frac{\varepsilon_r-1}{\varepsilon_r+2}$$

Simplifying the above equation, we get,

$$\frac{\alpha_e+\alpha_i}{\alpha_e}=\frac{(\varepsilon_r-1)}{(\varepsilon_r-2)}\frac{(n^2+2)}{(n^2+1)}$$

Substituting the values of ε_r and n^2, we get,

$$\therefore \quad 1+\frac{\alpha_i}{\alpha_e}=\frac{(4.94-1)}{(4.94+2)}\times\frac{(2.69+2)}{(2.69-1)}$$

(or)
$$\frac{\alpha_i}{\alpha_e}=0.5755$$

$$\frac{\alpha_e}{\alpha_i}=1.7376$$

The ratio of the electronic to ionic polariability is 1.7376

Example 21.4 The polarisability of Ne gas is 0.35×10^{-40} F m^2. If the gas contains 2.7×10^{25} atoms m^{-3} at 0 °C and 1 atmospheric pressure, calculate its relative dielectric constant.

Given Data

The electronic polarisability $\alpha_e = 0.35\times10^{-40}$ F m^2

The number of atoms $N = 2.7\times10^{25}$ atoms m^{-3}

Solution The dielectric constant is given by

$$\frac{\varepsilon_r-1}{\varepsilon_r+2}\frac{N\alpha_i}{3\varepsilon_0}$$

Substituting N, α_e and ε_0 values in the above equation, we get

$$\varepsilon_r=\frac{1+\dfrac{2N\alpha_e}{3\varepsilon_0}}{1-\dfrac{N\alpha_e}{3\varepsilon_0}}$$

Substituting the values of N, α_e and ε_0 in the above equation, we get

$$\frac{N\alpha_e}{3\varepsilon_0}=\frac{2.7\times10^{25}\times0.35\times10^{-40}}{3\times8.854\times10^{-12}}$$

$$=3.5577\times10^{-5}$$

Therefore,

the dielectric constant $\varepsilon_r = \dfrac{1 + 2(3.5577 \times 10^{-5})}{1 - 3.5577 \times 10^{-5}}$

$= 1.00010673$

The dielectric constant of Ne gas is 1.00010673.

Example 21.5 A parallel plate capacitor consists of 2 plates each of area 5×10^{-4} m^2. They are separated by a distance of 1.5×10^{-3} m and filled with a dielectric of relative permittivity 6. Calculate the charge on the capacitor if it is connected to a 100 volts d.c. supply.

Given Data

Area of the capacitor plate	$A = 5 \times 10^{-4}$ m^2
Distance between the plates	$d = 1.5 \times 10^{-3}$ m
Relative permittivity of the dielectric	$\varepsilon_r = 6$
Applied voltage	$V = 100$ V
Permittivity in free space	$\varepsilon_0 = 8.85 \times 10^{-12}$ F m^{-1}

Solution

We know that, the capacitance of the capacitor is $Q = C\,V$

$$C = \frac{\varepsilon_0\, \varepsilon_r\, A}{d}$$

$$Q = \frac{\varepsilon_0\, \varepsilon_r\, AV}{d}$$

Substituting the values, we get,

$$= \frac{8.85 \times 10^{-12} \times 6 \times 5 \times 10^{-14} \times 100}{1.5 \times 10^{-3}}$$

$$= 1.77 \times 10^{-9}$$

The charge on the capacitor is 1.77×10^{-9} C.

Example 21.6 Ar gas contains 2.7×10^{25} atoms m^{-3} at 0 °C and at 1 atmo-spheric pressure. Calculate the dielectric constant of the Ar gas at this temperature, if the diameter of the Ar atom is 0.384 nm.

Given Data

The number of Ar atoms m^{-3}	$= 2.7 \times 10^{25}$ m^{-3}
The diameter of Ar atom	$= 0.384$ nm

Solution The electronic polarisability of Ar atom

$$\alpha_e = 4\pi\varepsilon_0 R^3$$
$$= 4\pi \times 8.854 \times 10^{-12} \times (0.384 \times 10^{-9})^3$$
$$= 0.63 \times 10^{-40} \text{ F m}^{-2}$$

The dielectric constant of Ar, using Clausius–Mosotti equation is

$$\varepsilon_r = \frac{1 + \dfrac{2N\alpha_e}{3\varepsilon_0}}{1 - \dfrac{N\alpha_e}{3\varepsilon_0}}$$

Substituting the values of N, α_e and ε_0 in the above equation, we get

$$\varepsilon_r = \frac{1 + \dfrac{2 \times 2.7 \times 10^{25} \times 0.63 \times 10^{-40}}{3 \times 8.854 \times 10^{-12}}}{1 - \dfrac{2.7 \times 10^{25} \times 0.63 \times 10^{-40}}{3 \times 8.854 \times 10^{-12}}}$$

$$= \frac{1.000128078}{0.9999359611}$$

$$= 1.000192129$$

The dielectric constant of Ar is 1.000192129.

Example 21.7 A parallel plate capacitor has a capacitance of 2μF. The dielect-ric has permittivity, $\varepsilon_r = 80$. For an applied voltage of 1 kV, find the energy stored in the capacitor as well as the energy stored in polarizing the dielectric.

Given Data

The capacitance C = 2 μF = 2×10^{-6} F
The permittivity of the dielectric ε_r = 80
Applied voltage V = 1 kV = 1000 V

Solution The energy stored in the capacitor

$$E_1 = \frac{1}{2} C V^2$$

$$= \frac{1}{2} \times 2 \times 10^{-6} \times (1000)^2 = 1 \text{ J}$$

The capacitance of the capacitor when the dielectric is removed

$$C_0 = \frac{C}{\varepsilon_r}$$

$$= \frac{2 \times 10^{-6}}{80} = 2.5 \times 10^{-8} \text{ F}$$

The energy stored in the capacitor with vacuum as dielectric,

$$E_2 = \frac{1}{2} C_0 V^2$$

$$= \frac{1}{2} \times 2.5 \times 10^{-8} \times (1000)^2 = 0.0125 \text{ J}$$

Energy stored in the capacitor in polarising the dielectric,

$$C = E_1 - E_2 = \frac{1}{2}(C - C_0)V^2$$

$$E = 1 - 0.0125 = 0.9875 \text{ J}$$

The energy stored in the capacitor = 1 J
The energy stored in polarissing the capacitor = 0.9875 J

Example 21.8 A solid contains 5×10^{28} identical atoms per m^3, each with a polarizability of 2×10^{-40} F m^2. Assuming that the internal field is given by the Lorentz relation, calculate the ratio of the internal field to the applied field.

Given Data

Number of atoms present per m^3 $N = 5 \times 10^{28}$
Polarisability $\alpha_e = 2 \times 10^{-40}$ F m^2

Solution The Lorentz relation $E_i = E + \dfrac{P}{3\varepsilon_0}$

The polarisation $P = N\alpha_e E_i$

$$= 5 \times 10^{28} \times 2 \times 10^{-40} \times E_i$$

$$= 1 \times 10^{-11} \times E_i$$

The Lorentz field $E_i = E + \dfrac{1 \times 10^{-11} \times E_i}{3\varepsilon_0}$

i.e., $\left[1 - \dfrac{1 \times 10^{-11}}{3\varepsilon_0}\right] E_i = E$

or $\dfrac{E_i}{E} = \dfrac{1}{\left[1 - \dfrac{1 \times 10^{-11}}{3\varepsilon_0}\right]}$

$$\frac{E_i}{E} = \left[1 - \frac{10^{-11}}{3 \times 8.854 \times 10^{-12}}\right]^{-1} = 1.60379$$

The ratio of the internal field to the applied field = 1.6038.

Example 21.9 If a NaCl is subjected to an electrical field of 1000 V m^{-1} an dthe resulting polarization is 4.3 x 10^{-8} C/m^2, calculate the relative permittivity of NaCl.

Given Data

Applied electrical field $E = 1000$ V m^{-1}
Polarisation $P = 4.3 \times 10^{-8}$ C m^{-2}
Permittivity in free space $\varepsilon_0 = 8.85 \times 10^{-12}$ F m^{-1}

The McGraw·Hill Companies

Solution

We know that, the relative permittivity of NaCl is

$$\varepsilon_r = 1 + \frac{P}{\varepsilon_0 E}$$

Substituting the values, we get,

$$= 1 + \frac{4.3\times10^{-18}}{(8.85\times10^{-12})(1000)}$$

$$= 5.86$$

The relative permittivity of NaCl is 5.86.

Example 21.10 Calculate the electronic polariability of argon atom given $\varepsilon_r = 1.0024$ at NTP and $N = 2.7 \times 10^{25}$ atoms/m^3.

Given Data

Relative permittivity	$\varepsilon_r = 1.0024$
Number of atoms	$N = 2.7 \times 10^{25}$ atoms m^{-3}
Permittivity in free space	$\varepsilon_0 = 8.85 \times 10^{-12}$ F m^{-1}

Solution

We know that, the polarisability of argon atom is,

$$P = \varepsilon_0 (\varepsilon_r - 1)E$$

$$P = N\alpha_e E$$

$$\alpha_e = \frac{\varepsilon_0 (\varepsilon_r - 1)}{N}$$

Substituting the values, we get,

$$\alpha_e = \frac{(8.85\times10^{-12})(1.0024 - 1)}{2.7 \times 10^{25}}$$

$$\alpha_e = 7.9 \times 10^{-40} \text{ F m}^2$$

Therefore, the polarisability of argon atom is 7.9×10^{-40} F m^2

Example 21.11 The dielectric constant of He gas at NTP is 1.0000684. Calculate the electronic polarizability of He atoms if the gas contains 2.7×10^{25} atoms per m^3.

Given Data

The dielectric constant of the gas at NTP,	$\varepsilon_r = 1.0000684$
The number of He atoms,	$N = 2.7 \times 10^{25}$ m^{-3}
Permittivity in free space	$\varepsilon_0 = 8.85 \times 10^{-12}$ F m^{-1}

Solution

We know that, the electronic polarasability α_e is

$$\alpha_e = \frac{\varepsilon_0 (\varepsilon_r - 1)}{N}$$

Substituting the values, we get,

$$= \frac{8.85 \times 10^{-12} \times (1.0000684 - 1)}{2.7 \times 10^{25}}$$

$$= 2.242 \times 10^{-41} \text{ Fm}^2.$$

Therefore, the electronic polarisability of He atoms at NTP is equal to 2.242×10^{-41} Fm^2.

Example 21.12 An elemental dielectric material has a relative dielectric constant of 12. It also contains 5×10^{28} atoms m^{-3}. Calculate the electric polarisability assuming the Lorentz field.

Given Data

The relative dielectric constant of the material, $\varepsilon_r = 12$
The number of atoms in the element, $N = 5 \times 10^{28}$ m^{-3}
Permittivity in free space $\varepsilon_0 = 8.85 \times 10^{-12}$ F m^{-1}

Solution

We know that the electronic polarasability is

$$\alpha_e = \frac{\varepsilon_0 (\varepsilon_r - 1)}{N}$$

Substituting the values, we get,

$$\alpha_e = \frac{8.85 \times 10^{-12} \times (12 - 1)}{5 \times 10^{28}}$$

$$= 1.947 \times 10^{-39} \text{ F m}^2.$$

Therefore, the electronic polarisability of given element is 1.947×10^{-39} F m^2.

Example 21.13 A parallel plate condenser has a capacitance of 2 μF. The dielectric has permittivity $\varepsilon_r = 100$. For an apllied voltage of 1000 V, find the energy stored in the condenser as well as the energy stored in polarizing the dielectric.

Given Data

Capacitance of plate condenser $C = 2 \times 10^{-6}$ F
The applied voltage $V = 1000$V
The dielectric permittivity $\varepsilon_r = 100$

The McGraw·Hill Companies

Solution

Total energy stored in the capacitor $E = \frac{1}{2}CV^2$

Substituting the values of C and V in the above equations, we get,

$$E = \frac{1}{2} \times 2 \times 10^{-6} \times (10^3)^2$$

$$= 1 \text{ J}$$

We know that the energy stored in the dielectric material which is introduced in between the parallel plates of the condenser, capacitance has to be calculated removing the dielectric material.

$$C_0 = \frac{C}{\varepsilon_r}$$

Substituting the values, we get,

$$= \frac{2 \times 10^{-8}}{100} = 0.02 \ \mu F$$

Energy stored without the dielectric

$$= \frac{1}{2} C_0 V^2$$

Substituting the values, we get,

$$= \frac{1}{2} \times 0.02 \times 10^{-6} \times (10^3)^2$$

$$= 0.01 \text{ J}$$

Therefore, the energy stored in the dielectric

$$= 1 - 0.01$$

$$= 0.99 \text{ joules}$$

The energy stored in the dielectric is 0.99 J.

One Mark Question

21.1 The force acting per unit charge E is equal to

(a) $\frac{Q^2}{4\pi\varepsilon r^2}$ (b) $\frac{Q}{4\pi\varepsilon r^2}$

(c) $\frac{Q^2}{2\pi\varepsilon r^2}$ (d) $\frac{Q}{\pi r^2}$

The McGraw·Hill Companies

21.2 The value of permittivity of free space is equal to _____.

21.3 The dielectric constant of a solid material is equal to P = _____.

(a) $\varepsilon_o = \dfrac{\varepsilon}{\varepsilon r}$ (b) $\varepsilon_r = \varepsilon \varepsilon_o$

(c) $\varepsilon_o = \varepsilon \varepsilon_r$ (d) $\varepsilon_r = \dfrac{\varepsilon}{\varepsilon_o}$

21.4 The polarization of a solid material is equal to P = _____.

21.5 The unit for polarization is _____.

21.6 The polarization of a solid which contains N number of particles per unit volume is equal to

(a) $P_e = N\ \alpha\ E$ (b) $P_e = 2\ N\ \alpha\ E$

(c) $P_e = N\ \alpha^2\ E$ (d) $P_e = N\ \alpha^2\ E^2$

21.7 The electronic polarization of a solid material which contains N number of atoms are

a) $P_e = N\ 2\ \pi\ \varepsilon_o\ R^3\ E$ (b) $P_e = N\ \pi\ \varepsilon_o\ R^3\ E$

c) $P_e = N\ 4\ \pi\ \varepsilon_o\ R^3\ E$ (d) $P_e = N\ 4\ \pi\ \varepsilon_o\ R^3$

21.8 Atomic polarization is exhibited in ionic molecules like _____ and _____.

21.9 The value of ionic polarizability is equal to α_i = _____.

21.10 Dipolar or orientation polarization is produced in _____ molecules.

21.11 The value of orientation polarization is equal to _____.

(a) $P_e = \dfrac{1}{3}\left(\dfrac{\mu_F^2\ E}{RT}\right)$ (b) $P_e = \dfrac{1}{3}\left(\dfrac{\mu_F\ E}{RT}\right)$

(c) $P_e = \left(\dfrac{\mu_F^2\ E}{RT}\right)$ (d) $P_e = \dfrac{3}{4}\left(\dfrac{\mu_F^2\ E}{RT}\right)$

21.12 The value of orientation polarization is equal to _____.

21.13 The total polarization in a solid material is equal to P = _____.

21.14 The internal field ie equal to $E_i = E +$_____.

21.15 The internal or Lorentz field is equal to

(a) $E_i = E + E_c$ (b) $E_i = \dfrac{P}{3\varepsilon_0}$

(c) $E_i = E + \dfrac{P^2}{3\varepsilon_0}$ (d) $E_i = E + \dfrac{P}{3\varepsilon_0}$

21.16 Give an example for dielectric solid materials _____.

21.17 Give an example for polar solid dielectric materials _____.

21.18 The Clausius–Mosotti equation for a solid dielectric material is equal to

(a) $\dfrac{N\alpha_i}{3\varepsilon_0} = \dfrac{\varepsilon_r + 1}{\varepsilon_r - 2}$ (b) $\dfrac{N\alpha_i}{3\varepsilon_0} = \dfrac{\varepsilon_r - 1}{\varepsilon_r + 2}$

(c) $\dfrac{N\alpha_i}{\varepsilon_0} = \dfrac{\varepsilon_r - 1}{\varepsilon_r + 2}$ (d) $\dfrac{3N\alpha_i}{\varepsilon_0} = \dfrac{\varepsilon_r - 1}{\varepsilon_r + 2}$

The McGraw·Hill Companies

21.19 Dielectric materials which exhibit hysteresis curve are known as _____ material.

21.20 The Curie temperature for Rochelle salt is _____ and _____K.

21.21 The transition temperatures for barium titanate ferroelectric phases are
(a) 278 and 193 K (b) 278 and 393 K
(c) 278, 193 and 393 K (d) 278, 193 and 300 K

21.22 Examples for active dielectric materials are _____, _____ and _____.

21.23 Give two examples for passive dielectric materials _____ and _____.

21.24 The polarization is a function of time

(a) $P(t) = P\left[1-\exp\left(\frac{t}{T_r}\right)\right]$ (b) $P(t) = P\left[1+\exp\left(\frac{t}{T_r}\right)\right]$

(c) $P(t) = P\left[1-\exp\left(\frac{-t}{T_r}\right)\right]$ (d) $P(t) = P\left[1+\exp\left(\frac{-t}{T_r}\right)\right]$

KEY

21.1 (b)
21.2 8.854×10^{-12} F m^{-1}
21.3 (d)
21.4 Mass/Volume
21.5 C m^{-2}
21.6 (a)
21.7 (c)
21.8 NaCl, KBr
21.9 $\alpha_i = \frac{e^2}{\omega_0^2}\left(\frac{1}{M}+\frac{1}{m}\right)$
21.10 Polar
21.11 (a)
21.12 $\frac{\mu_F^2}{3RT}$
21.13 $P_e + P_i + P_o$
21.14 Field due to all other dipoles
21.15 d)
21.16 Diamond
21.17 $C_6H_5NO_2$
21.18 (b)
21.19 Ferroelectric
21.20 255, 296
21.21 d)
21.22 Piezoelectric, ferroelectric and quartz
21.23 Ceramic, mica
21.24 (d)

Exercises

21.1 A condenser of 1 μF contains titanium oxide as a dielectric with $\varepsilon_r = 100$. For an applied dc voltage of 1000 V, find the energy stored in polarizing the titanium oxide. Answer the same question for a 1 μF mica condenser, assuming the dielectric constant $\varepsilon_r = 5.4$ for mica.

21.2 An atom of polarisability α is placed in a homogeneous field E. Show that the energy stored in a polarised atom is equal to $1/2\alpha E^2$.

The McGraw·Hill Companies

21.3 The electronic polarisability of the Ar atom is 1.7×10^{-40} F m^2. What is the static dielectric constant of solid Ar (below 84 K) if its concentration is 2×10^{25} atoms m^{-3}?
21.4 The dielectric constant of Si is 11.9. What is the electronic polarisability due to the valence electrons per Si atom.
21.5 An elemental dielectric material has $\varepsilon_r = 12$ and it contains 5×10^{28} atoms m^{-3}. Calculate the electronic polarisability assuming Lorentz field.
21.6 The relative permittivity of Ar at 0 °C and 1 atmospheric pressure is 1.000435. Calculate the polarisability of the atom.
21.7 The dielectric constant of He at 0 °C and 1 atmospheric pressure is 1.000074. Calculate the dipole moment induced in each He atom. The electric field intensity of 3×10^4 Vm^{-1}.

Short Questions

21.1 Distinguish between a dielectric material and an insulator.
21.2 what are dielectric materials?
21.3 What is meant by polarisation in a dielectric material?
21.4 what are the different types of polarisations?
21.5 What is meant by dielectric loss?
21.6 What is meant by local field in a solid dielectric?
21.7 Explain the term total polarisation.
21.8 Explain the term dielectric breakdown.
21.9 Write the Clausius–Mosotti equation for a solid dielectric and explain the terms.
21.10 Mention the different types of dielectric material?
21.11 What are the properties of dielectric material?
21.12 What are the uses of dielectric materials?
21.13 What are ferroelectric materials? Give an example.
21.14 What is ferroelectricity?
21.15 List the different types of ferroelectric materials.
21.16 Explain the crystal structure of Barium titanate.
21.17 Explain the crystal structure of potassium dihydrogen phosphate.
21.18 Explain the crystal structure of Rochelle salt.
21.19 What are the changes observed in the crystal structure of $BaTiO_3$ with change in temperatures?
21.20 What is a domain in a ferroelectric material?
21.21 What is permanent polarisation?
21.22 What is coercive field in a ferroelectric material?
21.23 What are active and passive dielectrics?
21.24 What is electronic polarisation?
21.25 What is ionic polarisation?
21.26 What is dipolar polarisation?
21.27 What is space charge polarisation?

The McGraw·Hill Companies

Descriptive Questions

21.1 What is meant by polarisation? Write an essay about the different types of polarisation mechanisms involved in a dielectric material?

21.2 What is meant by local field in a solid dielectric? Deduce an expression for the local field in a solid dielectric and hence obtain Clausius–Mosotti relation.

21.3 Show that the imaginary part of the dielectric constant account for a condenser being loss.

21.4 Explain with a neat sketch the experemental determination of dielectric constant of the material.

21.5 What is meant by dielectric loss? Explain the phenomenon of dielectric loss in detail.

21.6 Write an essay about the different types of dielectric breakdown.

21.7 What are the different types of dielectric materials? Explain them in detail.

21.8 What is ferroelectricity? Explain the hysteresis curve exhibited by a ferroelectric material with a suitable sketch. Give examples for ferroelectric materials.

21.9 Explain with a neat sketch the frequency and temperature dependence of dielectric properties.

21.10 Explain the classification of insulating materials based on their temperature stability.

The McGraw·Hill Companies

Chapter

22

Shape Memory Alloys

OBJECTIVES

- To explore the knowledge on shape memory materials
- To discuss the shape memory effect and the special features of shape memory alloys
- To understand the different characterisation tools to explore the phase transformation temperatures
- To understand the commercial applications of shape memory alloys

22.1 INTRODUCTION

In recent decades, material science has drastically gained importance in view of the development of newer materials with exotic properties, new process techniques, smart material performance, etc. The development of *multifunctional materials* is due to the demand for materials with additional engineering functionality such as sensing, actuation, electromagnetic shielding, etc.

Some of the functional materials exhibit unique properties such as the ability to recover their shape or size when there is an increase in temperature. The material which can recover its original shape or size when it is subjected to an appropriate temperature changes, either heating or cooling, is known as a *Shape Memory Alloy* (SMA). The shape recovery of the material is possible even under high applied loads. The SMA behaves as a rubber, at lower temperatures and can be deformed even for small applied forces. However, at high temperatures, it behaves like a metal and hence, the induced strains are highly irrecoverable. The unique characteristics of SMAs lead to many industrial applications such as sensing, actuation, impact absorption and vibration damping. The application of SMAs in industrial sectors such as aerospace, automotive, biomedical and oil exploration are evidenced extensively.

Shape memory alloys are classified into two types, namely, one-way shape memory and two-way shape memory alloys. The material which exhibits the shape memory effect only upon increase in temperature is known as *one-way shape memory alloy.*

The McGraw·Hill Companies

Some of the materials can exhibit the shape-memory effect both during increase and decrease in temperatures. These materials are known as *two-way shape memory alloys.*

In order to understand the science and technology of *Shape Memory Alloys* (SMAs), detailed discussions are given in the following sections:

22.2 ORIGIN OF SHAPE MEMORY ALLOYS

In 1890, Adolf Martens discovered the existence of martensite phase in steels. The concept of thermo-elastic martensitic transformation and, thereby, the reversible transformation of martensite was explained in 1949 by Kurdjumov and Khandros. Arneolander, a Swedish physicist observed an interesting phenomenon on gold (Au) and Cadmium (Cd) alloys in the year 1932. When the Au–Cd alloy is cooled to low temperature, it exhibits a plastic feel and while heating the material, it surprisingly returned to its original dimensional configuration. The effect of thermo-elastic transformation structure during the change in temperature is known as *Shape Memory Effect (SME).* The material which exhibit this phenomena are known as *shape memory alloys (SMA).* The discovery of NiTi is a major breakthrough in SMAs for engineering applications. The NiTi material is named NiTiNOL in order to honour its discovery by Buchler and his co-workers from the Naval Ordnance Laboratory (NOL). Alloying elements such as Co or Fe are added to the existing NiTi system, resulting in an abrupt change in the SMA transformation temperatures. The NiTi material added with Co/Fe elements are extended for first commercial applications and hence, are known as Cryofit. The new materials such as TiPd, TiPt and TiAu are developed with improved transformation temperatures greater than 373 K. These materials are known as *High Temperature SMA* (HTSMA).

22.3 PRINCIPLE OF PHASE TRANSFORMATION IN SHAPE MEMORY ALLOYS

The formation and disassociation of phases of alloys at a fixed composition is mainly because of the change in temperature. The *Shape Memory Alloys* (SMA) have two phases, namely, martensite and austenite within the operating temperature region. The structure of the SMA in the respective phases is different and hence, the properties vary widely. The phase of the material is changed from austenite (A) to martensite (M) depending on the temperatures of the alloys. At high temperature, the material is in *austenite* (A) phase, while at low temperature it is transformed into *martensite* (M) phase. The austenite phase takes a cubic structure while the martensite phase takes either tetragonal or orthorhombic or monoclinic structure. The transformation of crystalline structure of the alloy occurs by shear lattice distortion, and it is known as *martensitic transformation.* The direction of orientation of each martensitic crystal is called *variant*. The assembly of martensitic variants can exist in two forms, namely, *twinned* martensite (M_t) and *detwinned* martensite. The first one is self-accommodated martensitic variants while the later is the dominant of the specific variant (M_d).

The McGraw·Hill Companies

The crystal structure change from austenite to martensite during cooling is called *forward transformation*. It is the result of the formation of several martensitic variants. When the material is heated to high temperature, the crystal structure is transformed from the martensitic phase to the austenite phase. The above transition is called *reverse transformation.*

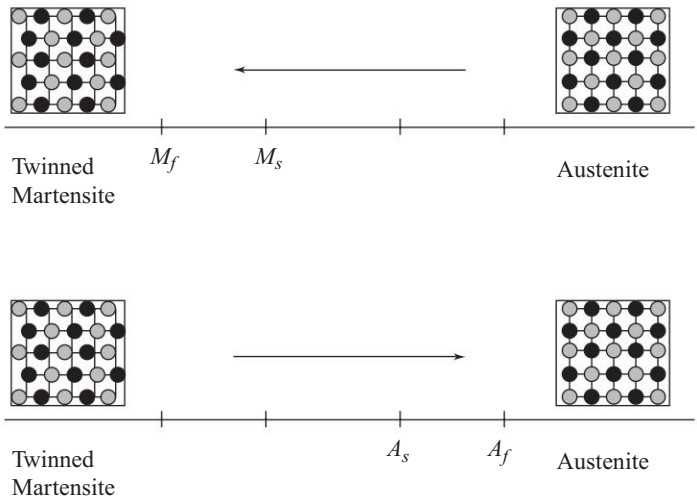


Fig. 22.1 *SMA—phase transformation*

The schematic representation of the twinned martensite and austenite crystal structures of the SMA and the corresponding phase transformations between them are shown in Fig. 22.1. The characteristic temperatures associated with the phase transformation are martensitic start (M_s), martensitic finish (M_f), austenitic start (A_s) and austenitic finish (A_f). *Martensitic start temperature* (M_s) is the temperature at which the crystal structure begins to transform to the twinned martensite state from the austenite state during the forward transformation. This transformation completes the crystal structure to martensitic phase at the temperature known as *martensitic finish temperature* (M_f). At this stage, the material is fully converted into the twinned martensitic phase. Similarly, during heating, the reverse transformation starts at the *austenitic start temperature* (A_s). The transformation is completed at the *austenitic finish temperature* (A_f) and hence, it reduces the austenitic structures.

When a mechanical load is applied to the material which is in the twinned martensitic phase at low temperature, it is possible to detwin the martensite by reorienting a certain number of variants as shown in Fig. 22.2. During this detwinning process, a macroscopic shape change occurs. The above deformed configuration is recovered to the original shape when the load is released. A reverse phase transformation takes place upon a subsequent heating of the SMA which is in the detwinned martensite phase to a temperature above the austenite finish (A_f). Thus, the material attains the complete shape recovery as shown in Fig. 22.3.

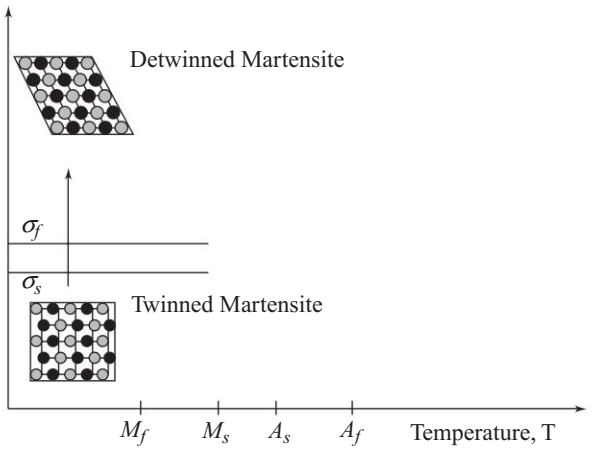


Fig. 22.2 *SMA—detwinning*

When the material is cooled again below the temperature M_f, it again leads to the formation of twinned martensite without any deformation in its shape. This process is known as the *Shape Memory Effect* (SME). During the detwinning process, the applied load must be sufficiently large so as to start the process. The detwinning initiation is termed *detwinning start stress* (σ_s), which is the minimum stress required to start detwinning. Complete detwinning of martensite occurs at sufficiently high applied load and is called *detwinning finish stress* (σ_f).

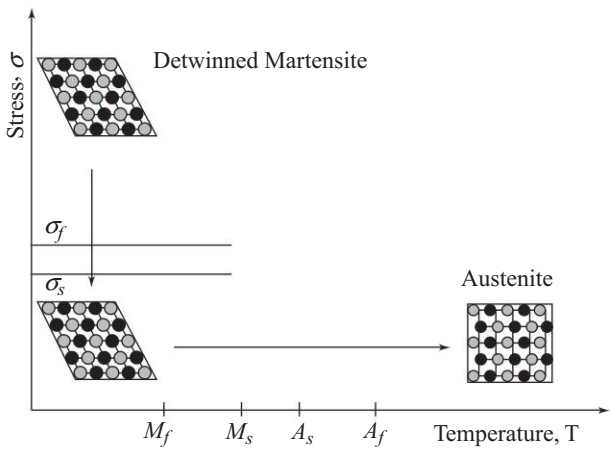


Fig. 22.3 *SMA—austenite without any applied load*

Direct formation of detwinned martensite occurs during the cooling of material which is in the austenitic phase with a mechanical load greater than σ_s. Thus, the deformation in shape takes place. The deformed shape of the material is totally recovered at the austenitic phase during the reheating process even though the applied stress σ_s are continued.

The McGraw·Hill Companies

In this case, the magnitude of the applied load is the key factor in determining the transformation temperatures. The transformation temperature increases with increase in the value of applied load as shown in Fig. 22.4.

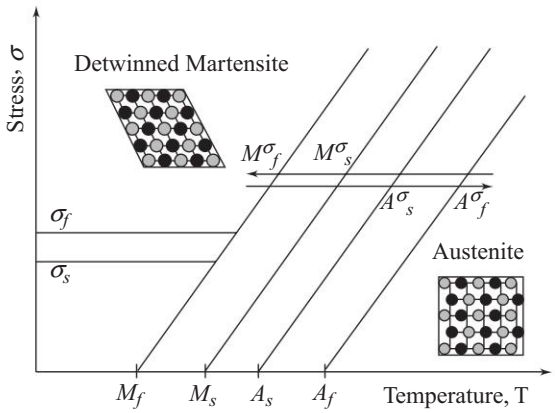


Fig. 22.4 *SMA—phase transformation during the load*

22.4 SHAPE MEMORY ALLOYS—PROPERTIES

22.4.1 Hysteresis

The types of alloys and their compositions, and the applied thermo-mechanical treatments are the key factors which induce the phase transformations in SMAs. The temperature required for the transformation of the phase changes from martensite to austenite and hence, from austenite to martensite is different. During the increase in temperature, somewhat higher temperature is required to induce the transformation from martensite to austenitic than that for the reverse transformations, i.e., from austenitic to martenstic during cooling. The difference in the temperatures during the transformation upon heating and cooling exhibits a hysteresis curve as shown in Fig. 22.5.

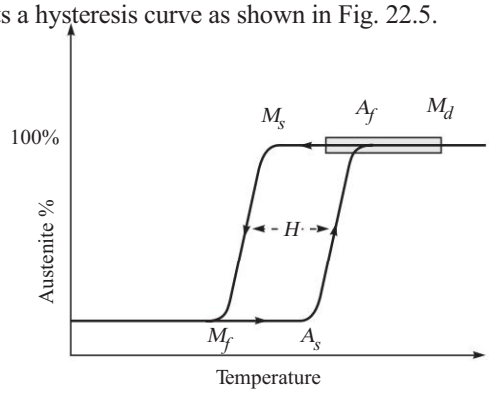


Fig.22.5 *Hysteresis curve*

The martensitic phase start and finish are represented respectively by M_s and M_f. Similarly, the austenitic phase start and finish temperature are represented respectively by A_s and A_f. The SMA exhibits a hysteresis curve, since there is no overlay of transformation temperature during heating and cooling. The hysteresis curve depends on the alloy systems. Most of the phase transformation occurs in a narrow temperature range. However, the starting and ending transformation of phases during heating and cooling occurs relatively at higher temperature range. The transformation temperature of few SMAs is shown in Table.22.1.

The difference between the temperature at which 50% of the material is transformed into austenitic during heating and 50% transformation of the material to martensite during cooling is known as the transformation temperature *T.* In most of the SMAs, the value of *T* lies between 293 and 303 K.

Table 22.1 *Transformation temperature range of few SMAs*

Alloys	*Composition*	*Transformation temperature range (K)*	*Transformation hysteresis T (K)*
Ag – Cd	44/49 wt % cd	83 – 223	≅ 15
Ni – Al	36/38 wt % Al	93 – 173	≅10
Ni – Ti	49/51 wt % Ni	223 – 163	≅ 30
Mu – Cu	5/35 et % Cu	23 – 93	≅ 25
Cu – Zn	38.5/41.5 wt % Zn	93 – 263	≅ 10

22.4.2 Two-Way Shape Memory Alloy

A repeated shape change in the SMA takes place when it is subjected to cyclic thermal conditions during the absence of any applied load. The ability of the SMA is to recover its original shape to austenite whenever the temperature is increased, and also to gain the deformed or changed shape to martensite while cooling is known as *two-way shape memory effect* (TWSME). A change in the microstructure leading to macroscopic change in the material behaviour is expected due to the repeated cycles of loading and unloading.

22.4.3 Pseudo Elastic Effect

The phase transformation of an SMA highly depends on the thermal conditions and also to the applications of high mechanical load to the material in the austenitic phase. The result of the applied load is a fully detwinned martensite phase created from austenite. When the temperature of the material is above A_f, the stress-introduced transformation leads to strain generation during loading and stress recovery during unloading. Therefore, the martensite phase totally changed upon removal of the stress and the austenite phase recovers its original shape. This is based on the stress induced martensite formation. The effect of return to its original shape upon loading after substantial deformation is called pseudoelastic effect or superconductive

The McGraw·Hill Companies

effect. The pseudoelastic effect is shown schematically in Fig. 22.6. The σM_s and σM_f are the stress levels respectively at which the martensite transformation starts and completes.

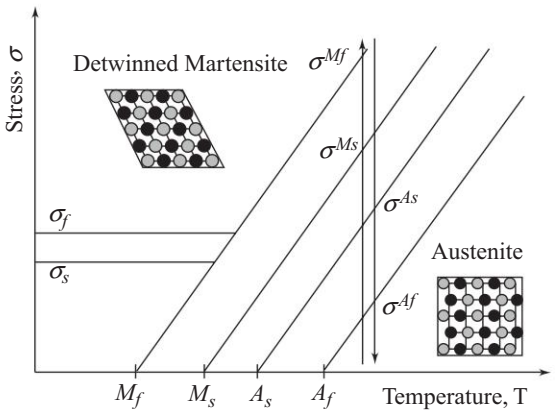


Fig. 22.6 *Pseudoelastic effect*

Similarly, a reverse transformation to austenite from detwinned materials takes place when the load is reversed from SMA. The corresponding stress levels at which the material initiates and completes are represented respectively by σA_s and σA_f.

22.4.4 Thermoelastic Properties

The elastic constants such as Young's modulus, Poisson's ratio, etc., are used to obtain the thermo-elastic properties of SMA, both at austenite and martensite conditions. These parameters are used to explore the plastic yield and failure behavior of the materials for design applications. The elastic stiffness of the material at austenite condition is taken as Young's modulus of austenite (E^A). Similarly, the elastic constant of the material after completing the transformation from austenite to martensite is taken as elastic modulus of martensite (*EM*).

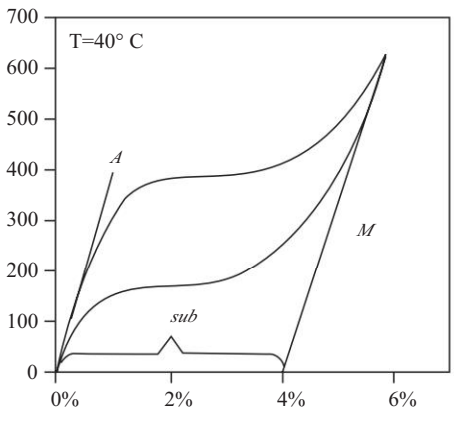


Fig. 22.7 *Dependence of strain against the stress*

The McGraw·Hill Companies

Similarly, the Poisson's ratio for each phase is measured concurrently with the stiffness. The thermally induced deformation response of austenite and martensite materials is predicted by observing the coefficient of thermal expansion.

22.5 PROCESSING TECHNIQUES

Materials such as titanium, used for the preparation of SMA, are highly reactive to any foreign materials present in the region. Therefore, instead of the normal material-processing techniques, a high vacuum or an insert atmosphere is required for the processing of SMAs. In addition, the processing temperature also needs to be very high. The commercial SMA is prepared employing techniques like plasma arc melting, electron beam melting, and vacuum induction.

22.6 CHARACTERISATION TECHNIQUES

Even though a number of materials are available for material characterisation, the following four different methods are widely used to explore phase transformations in the SMAs:

(i) Differential scanning calorimeter
(ii) Resistivity measurements
(iii) Transformation of SMA with temperature
(iv) Tensile test

22.6.1 Differential Scanning Calorimeter

The release and absorption of latent heat of the material is responsible for the observed thermally induced phase transformation from austenite to martensite and vice versa. The heat of transformation and the related transformation temperatures are obtained using the technique known as *Differential Scanning Calorimeter* (DSC). The information such as phase transformation temperature, latent heat due to transformation, and the specific heat capacity of different phases in a material are explored from these measures. In this technique, the sample material is encapsulated in an inert atmosphere to prevent oxidation.

A tangent is drawn at the start and end regions of the transformation peaks and the baseline of the heating and cooling curve. A typical DSC curve obtained for an SMA is shown in Fig.22.8. The corresponding values of the tangent give the transformation temperatures of the SMA.

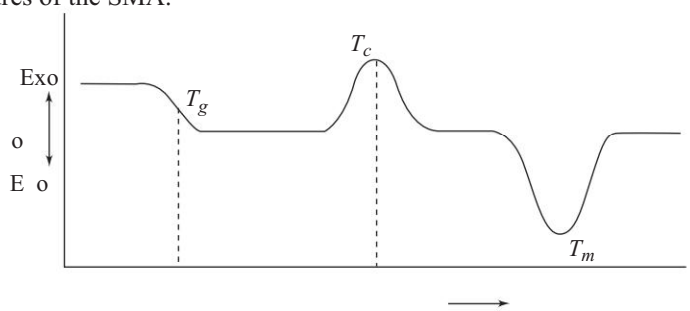


Fig.22.8 *DSC curve*

The McGraw·Hill Companies

The specific heat capacity of the material is computed by normalizing the power employed during the heating rate and the weight of the specimen. The associated latent heat for the phase transformation is obtained by integrating the specific heat over the range of the transformation temperatures. The stored mechanical energy may affect the transformation temperatures significantly and hence, the initial state of the material is to be noted accurately. Thus, the stored mechanical energy in turns can cause either a shift or widening in the transformation temperatures. From the observed endotherm and exotherm peaks of the DSC curve, one can easily measure the energy absorbed or given off at the beginning and end of the phase transformation.

22.6.2 Resistivity Measurements

It is the second method used to measure the phase transformation in an SMA by measuring the change in resistivity as a function of temperature. In order to measure the transformation temperature both ways, i.e., heating and cooling, one has to measure the resistivity of an SMA as a function of temperature over a wide range. The measured resistivity shows peaks at phase transformation temperatures. The results obtained by this method often disagree with the measured phase changes.

22.6.3 Transformation of SMA with Temperature

The most direct method employed to characterise the SMA is the stress–strain measurements. By maintaining a constant stress, the strain on the sample is measured during the transformation of both cooling and heating. The observed transformation hysteresis is shown in Fig. 22.9. The obtained hysteresis transformation is the direct measurement from the stress–strain measurements. The transformation hysteresis temperature T, martensite start M_s, austenite start A_s, martensite finish M_f, and austenite finish A_f temperatures are obtained from the hysteresis results. The obtained transformation temperatures for the M_s and A_f from the above method are slightly higher than those obtained from the DSC method. The observed higher transformation temperature is due to the stress applied on the SMA during the measurements,

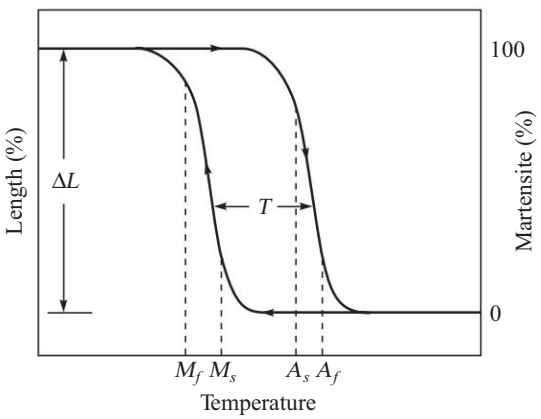


Fig. 22.9 *Transformations versus temperature*

while in the DSC measurement, the transformation is obtained without any stress on SMA. The results obtained by this method are from direct measurements. The main disadvantage of this method is the sample preparation and the method of conducting the test on the SMA.

22.6.4 Tensile Test

In this method, the SMA is subjected to the standard tensile test within the range of transformation temperature. The stress–strain properties of the SMA are obtained during both heating and cooling. From the measured properties, the approximate transformations temperature for the SMA is obtained. Even though this method is used to obtain the transformation temperatures, it is not effectively used.

22.7 COMMERCIAL SHAPE MEMORY ALLOYS

A wide variety of SMAs were investigated over the last few decades, starting from the initial investigation of AuCd and AgCd alloys. The Ni–Ti and copper-based alloys are the first developed commercial SMAs. The development and fabrication of new compositions of SMAs are being made by adding different alloy elements to the existing alloys. This is mainly used to produce different SMAs with required properties. Shape memory alloys are classified based on a variety of categories such as primary alloy elements, mode of actuation of both magnetic and thermal, and operating temperature.

22.7.1 NICKEL-TITANIUM BASED ALLOYS

(i) ***Ni–Ti alloys*** In the early 1960s, Buehler and his co-workers discovered the shape memory effect in a nickel–titanium alloy. The composition of Ni–Ti alloy is at equiatomic state, i.e., 50 % of Ni and Ti having the maximum A_f temperature of 393 K for all its Ni–Ti compositions. The properties of Ni–Ti alloys depend on the composition of Ni added to the alloys. As the composition of Ni decreases in Ni–Ti alloys, there is no change in the value of transformation temperatures. On the other hand, an increase in the percentage of Ni above 50% leads to a decrease in the transformation temperature. This results at ambient room temperature of 296 K, the existence of characteristic properties from SME to pseudo elasticity. The Ni–Ti alloy shows good result on shape memory effect in both ways. The pseudo-elastic behaviour of this alloy is studied extensively and found suitable for a variety of commercial applications. Ni–Ti alloys exhibit good biocompatibility and better resistance to corrosion and hence, prove their suitability for many biomedical applications. The thermal, mechanical and shape memory properties of the Ni–Ti alloy is shown in Table 22.2.

Knowledge of crystallography and thermo-mechanical response of Ni–Ti is equally important. The formation of intermediate phase and, thereby, a change in the crystalline structures of the Ni rich Ni–Ti alloy are obtained through aging at certain temperatures. Ni–Ti alloys exhibit fully recoverable transformation strains up to 8%. Ni–Ti alloys are available commercially in various forms such as wires, strips, rods, tubes and plates.

The McGraw·Hill Companies

Table 22.2 *Thermal, mechanical and shape memory properties of Ni–Ti alloy*

Properties	*Value*
Thermal properties	
Melting temperature (K)	1573
Density ($x10^{-3}$ kg m^{-3})	6.50
Resistivity ($x10^{-8}$ / μΩ)	70 – 100
Thermal conductivity (Wm^{-1})	8.5 – 18
Material properties	
Young's modulus (GPa)	
Austenite	83
Martensite	28 – 41
Ultimate tensile strength (MPa)	895
Shape memory properties	
Transformation temperatures (K)	73 – 383
Maximum shape memory strain (%)	8.5

When the Ni constant on Ni–Ti is increased to 55%, a change in the transformation temperature is observed in the range of 263 K to 333 K. The low transformation strains are obtained due to the multiphase structure of the alloy. This composition has proven to the superior corrosion resistanete than stainless steels, even in a harsh environment such as a salt water bath or salt fog. In addition to the above, this alloy shows an excellent thermo-mechanical stability and easier control of transformation temperatures through heat treatment.

(ii) ***Nic–Tic–Cu Alloys*** The addition of Cu to Ni–Ti replaces the Ni and forms the Ni–Ti–Cu alloys. The composition of Cu to be added in a binary system is ≤ 10%. The addition of Cu greater than 10% results in embitterment of the material. Thus, it helps to reduce the hysteresis of SMA response, transformation strain and the pseudo-elastic hysteresis. The small hysteresis makes the Ni–Ti–Cu alloy an ideal choice for actuators. The width of the pseudoelastic hysteresis reduces width addition of Cu to the Ni–Ti alloys. A reduction in pseudoelastic hysteresis from 200 Mpa to 100 Mpa while adding 10% to Ni–Ti alloys is observed. The addition of Cu greatly reduces the sensitivity of the martensitic start temperature and hence, there is a corresponding change in the phase transformation.

(iii) ***Ni–Ti-Nb Alloy*** The addition of niobium (Nb) element to Ni–Ti alloy widens the hysteresis curve. In SMAs, the existence of wide hysteresis has important practical

applications. Thus, it helps to show a minimum response to the change in the hystersis temperature.

(iv) ***Ni–Ti–X Alloys*** The demand for SMAs with high transformation temperatures results in the development of a new class of SMAs known as *High Temperature Shape Memory Alloys (HTSMAs)*. The HTSMAs are a unique class of SMAs that have the transformation temperatures greater than 373 K and are capable of actuating under high temperature conditions. Applications such as the core region of an aircraft engine or downhole applications in the oil industry require high transformation temperatures along with stable material properties. In the past few decades, SMAs were developed for commercial applications with an operating temperature less than 373 K.

One can produce a variety of novel shape memory alloys by adding the ternary elements such as palladium, platinum, hafnium, gold, and zirconium to Ni–Ti, wherein the transformation temperatures are shifted anywhere in the range between 373 and 1073 K.

22.7.2 COPPER–BASED ALLOYS

The Ni–Ti SMAs offer excellent properties like pseudoelasticity and SME but the main disadvantage of the system is the cost when compared to Cu-based SMAs. The properties such as electrical and thermal conductivity of the Cu-based SMAs show that it is the best alternative for Ni–Ti alloy. The hysteresis of copper-based alloys exhibit less value than Ni–Ti and its transformation temperatures are highly dependent on the composition. The Cu–Zn and Cu–Al alloys are the main Cu-based alloys.

(i) Cu–Zn–Al The ductility and resistance to intergranular fracture of Cu–Zn binary alloys are comparatively higher than any other Cu-based alloys. One can increase the transformation temperatures by adding aluminium. Further, one can also change the M_s temperature over a wide range from 93 K to 373 K by changing the composition of aluminum between 5 wt.% and 10 wt.%. of Cu–Zn–Al alloys. These alloys are very sensitive to heat treatments. A phase dissociation or a change in transformation temperatures is achieved by means of quenching the alloys.

(ii) Cu–Al–Ni Similar to CuZnAl, the transformation temperatures of Cu–Al–Ni alloy depends on the change in composition and the content of aluminum or nickel to be added. One can change the M_s temperature from 133 K to 373 K by changing the composition of aluminium between 14% and 14.5%. The change in the hysteresis behaviour and transformation temperature fairly depends on the change in composition. This alloy is harder to produce and hence, manganese is added to improve its ductility. Similarly, titanium is added to refine its grains.

The thermal, mechanical and shape memory properties of the Cu–Zn–Al and Cu–Al–Ni alloys are shown in Table 22.3. For commercial applications many novel SMAs with wide and varied properties are under development. Recently the materials like, CuAlBe, CuAlMn and CuAlNb have been developed. The CuAlMn alloy has very good ductility, while the CuAlNb alloy is more suitable for high temperature applications.

The McGraw·Hill Companies

Table 22.3 *Thermal, mechanical and shape memory properties of Cu–Zn–Al and Cu–Al–Ni alloys*

Properties	*Cu–Zn–Al*	*Cu–Al–Ni*
Thermal properties		
Melting temperature (K)	1223–1293	1273–1323
Density (x10^{-3} kg m^{-3})	7.64	7.12
Resistivity (x10^{s} / $\mu\Omega$)	8.5–9.7	11–13
Thermal conductivity (W m^{-1})	120	30–43
Mechanical properties		
Young's modulus (GPa)		
Austenite	70	85
Martensite	70	80
Ultimate tensile strength (MPa)	600	500-800
Shape memory properties		
Transformation temperature (K)	< 393	< 473
Maximum shape memory strain	4.0	4.0

22.7.3 Iron-Based Alloys

In recent years, iron-based ferrous alloys such as FeNiCoTi and FeMnSi have been developed. The alloy $FeNi_{31}Co_{10}Ti_3$ exhibits good shape memory effect after specific thermomechanical treatments. Similarly, the alloy FeMnSi has good commercial prospects. This is mainly due to the addition of Si content to its composition to improve the shape memory effect.

22.7.4 Additional SMAs

(i) The CoNiAl alloy is prepared with the addition of Co to Ni–Al or Ni to the CoAl binary alloy systems. The CoNiAl alloy has a very good corrosion and oxidation resistance at high temperatures. This alloy with a composition of $CoNi_{33}Al_{29}$ shows wide transformation temperatures and stability in the pseudoelastic behaviour over a wide range of temperatures. The magnetic properties of CoNiAl alloys reveal that martensite reorients when it is subjected to an external magnetic field. As a result, the material generates a magnetic-field-induced reorientation strain.

(ii) NiMnGa is the most widely investigated *Magnetic Shape Memory Alloys* (MSMAs). The NiMnGa alloy demonstates the magnetic field-controlled shape memory effect with larger field-induced strains. The other important property of this alloy is the relatively low blocking stress, i.e., the stress at which the magnetic reorientation strain is completely suppressed. This is the main limitation of Magnetic Shape Memory Alloys (MSMAs). The studies on other magnetic shape memory alloys such as Fe–Pd, Fe–Ni–Co–Ti, Fe–Pt, Co–Ni–Ga, Ni–Mn–Al and Co–Ni–Al are revealed with existence of lower field-induced strains.

22.8 SHAPE MEMORY ALLOYS-APPLICATIONS

Nowadays, newer materials with novel properties are required to meet the existing application requirements. The best active materials such as SMAs are quickly gaining momentum mainly due to their new kind of behaviours and multi-functionality. The conversion of thermal energy into mechanical work using SMAs for real has been aligned practical applications. The applications of the SMA in the present scenario are diversified and ranges from everyday consumer products to biomedical implants applications. The following are some of the applications of the safe memory and super-elastic alloys:

(i) ***Aerospace and Military*** The SMA has wide applications in aerospace industry such as in the areas of fixed-wing aircraft, rotorcraft and spacecraft. The application of SMAs in the fixed wing aircraft is to optimise the performance of lifting bodies.

Reduction in Aircraft Engine Noise

The SMA is used in the aircraft engine to reduce the noise. This is aligned by installing chevrons into the engine and to mix them with flow of exhaust gases where it reduces the engine noise. The SMA components are embedded inside the chevrons. During low-altitude flight or low-speed flight, the chevrons containing the SMA beams bend and, thereby, increase the mixing and hence, reduce the noise. Similarly, during high-altitude or high-speed flight, the SMA beam components cool into martensite state and thereby straightening of the chevrons results in an increase of engine performance. Thus, the aircraft noise levels during takeoff and landing are controlled effectively.

(ii) Transportation SMAs are extensively used is automobiles for applications ranging from impact absorption to sensing and actuation. The key properties of SMAs like pseudoelastic behaviour and hysteresis provide an effective system to dissipate vibrations and impact.

The SMA is used as an actuator in actuating the blinds which cover the fog lamp to prevent it from external damage, Whenever the fog lamp is turned on, the actuation of the SMA louvers through the electrical circuits and prevents damage to the lamps. It is also used for sensing and actuating purposes simultaneously.

(iii) Safety Applications The SME of the alloys is used for many applications. The SMA actuator is used to shut off the toxic or flammable gas flow when there is a fire. A similar actuation system is incorporated in the Shinkansen bullet train gearbox wherein the temperature in the gear box is monitored. The SMA spring actuates a valve to adjust the oil level in the gearbox. Other applications developed for trains include the thermally actuated switch for the radiator fan in diesel engines and steam traps for the steam heating system in passenger trains. In both of these applications, the shape memory effect is used. Ni–Ti alloys are used as actuators and fasteners in automobile applications. The fan clutch is used to regulate the temperature of the engine. Similarly, the SMA is used as a thermostat to open and close the valve to control the flow of water in the cooling system.

(iv) Consumer Products Most of the consumer product applications are based on the pseudoelastic properties of SMAs. Some of the applications are eyeglass frames, golf equipment, wires in retractable antennas, gained wire for steering catheters into

The McGraw·Hill Companies

vessels, working toys, etc. The SMA is used as a thermostat in coffeepots to open and close the valve to control the release of hot water to brew a perfect pot of coffee.

(v) Medical The pseudoelastic characteristics and the biocompatibility properties of the SMAs like NiTi, makes the SMA an attractive element for use in medical applications. The SMA is used for applications such as stents, filters, and orthodontic wires as well as devices for Minimally Invasive Surgery (MIS).

Blood–clot Filter

A suitable shape of Ni–Ti is used to filter blood clots when it is inserted into the vein. The SME is used to achieve this filtering. The undercooled SMA is inserted into the vein in its deformed shape. The existence of body heat helps the SMA recover its original shape due to SME and filters the trapped blood clots passing through the vein. The trapped clots get dissolved eventually.

Orthodontic

SMAs are successfully implemented in a variety of dental applications. For example, nitinol orthodontic arch wires are used as an effective material than steel wires. When a linear wire made of stainless steel is used in orthodontic archeries, it exacts a large amount of force on the tooth for a small amount of corrective motion, even for a small strain. The advantage of pseudoelastic arch wires is their ability of correction motion. When the SMA is used as an arch wire, it operates in the pseudoelastic plateau region and hence, it has a near-zero stress due to the large strain increments. Therefore, the replacement of steel wires by SMA arch wires results in a moderate force on the tooth for a long time.

In addition, SMAs find wide applications in product development such as microsurgical instruments and flow control devices, particularly for biomedical applications.

22.9 SHAPE MEMORY ALLOYS-DISADVANTAGES

The following are some of the disadvantages of SMAs:

- The manufacturing cost of SMAs is more than any other materials.
- The fatigue properties of the SMA are very poor. As a result, a load with a standing capacity of an SMA is 100 times less than that of steel components.

- Shape Memory Alloys (SMAs) are materials which can recover their original shape or size when subjected to an appropriate temperature change, either heating or cooling.
- The important two phases of the SMA are martensite and austenite.

- Martensitic start (M_s), martensitic finish (M_f), austenitic start (A_s) and austenitic finish (A_f), are the important transition temperatures.
- High Temperature Shape Memory Alloy (HTSMA) are materials with transition temperatures higher than 373 K.
- Shape Memory Effect (SME) is the process of returning to the original dimensional configuration during heating.
- The materials which exhibit SME both during heating and cooling are known as two-way SMAs.
- The material which exhibit SME upon increase in temperature are known as one way shape memory alloys.
- The difference in the transformation temperature during heating and cooling exhibits a smooth curve known as hysteresis.
- Superelasticity or pseudoelasticity is the ability of the SMA to return to its original shape upon loading after substantial deformation.
- Categories of SMAs are nickel ntitanium, copper-and iron-based alloys.

Objectives Questions

22.1 _____ and _____ alloy exhibits the first shape memory effect.
22.2 The first commercial application of SMA is achieved by adding Co/Fe elements on _____
22.3 _____ is the first commercial SMA.
22.4 The two phases of the SMA are _____ and _____.
22.5 The characteristic temperature associated with the phase transformation are
a) M_s, M_f, A_s, M_d b) M_s, M_f, A_s, Af
c) M_s, M_d, A_s, A_f d) M_s, M_d, A_s, M_f
22.6 The reverse transformation starts at _____ temperature.
22.7 The austenite structure is reduced at the _____ finish temperature.
22.8 _____ SMA is used for biomedical applications.
22.9 In SMA, the ductility is improved by adding_____.
22.10 _____ is used to refine the grains is SMA
22.11 The most widely used magnetic shape memory alloy is _____.
22.12 SMA is used in aircraft to _____ the engine noise.

KEY

22.1 A_u and C_d 22.2 NiTi
22.3 Cryofit 22.4 Martensite and Austenite
22.5 b) 22.6 Austenitic start
22.7 Austenitic 22.8 NiTi
22.9 Manganese 22.10 Titanium
22.11 NiMnGa 22.12 Reduce

The McGraw·Hill Companies

Short Questions

22.1 What is meant by multifunctional materials?
22.2 What are shape memory alloys?
22.3 What is meant by one-way shape memory alloy?
22.4 Explain the two-way shape memory alloy.
22.5 Differentiate between one-way and two-way shape memory alloys.
22.6 Define shape memory effect.
22.7 Explain high-temperature shape memory alloy.
22.8 What is meant by phase transformation?
22.9 Explain the martensite and austenite phase.
22.10 Define varriant.
22.11 Explain the twinned and detwinned martensite.
22.12 What is meant by hysteresis?
22.13 Explain two way shape memory effect.
22.14 What is pseudoelasticity?
22.15 Mention few commercial shape memory alloys.
22.16 What are the properties of SMA
22.17 What is meant by actuators?
22.18 Mention few applications of SMAs in aerospace and military.
22.19 How SMA is used in safety applications.
22.20 Mention few applications of SMAs in consumer products.
22.21 Explain how SMAs are used to filter the blood clots.
22.22 Explain the disadvantages of SMAs.

Descriptive Questions

22.1 Explain with a neat sketch the origin, principle and phase transformation of SMAs.
22.2 Describe the different properties of SMAs with suitable illustrations.
22.3 Describe with a neat sketch the different characterisation techniques used for SMAs.
22.4 Write an essay on the application of different commercial SMAs.
22.5 Explain with suitable examples the applications of SMAs in different fields.
22.6 Write notes on the following:
a) Shape memory effect
b) Hysteresis
c) Pseudoelastic effect
d) Ni–Ti alloys
22.7 Explain the shape memory effect and pseudoelasticity of shape memory alloys with necessary diagrams.

Chapter

23

Nonlinear Materials

OBJECTIVES

- To introduce the concept of nonlinear optics
- To describe the nonlinear behaviour and the properties of materials
- To understand the passive and active nonlinear materials and their properties
- To study the different types of nonlinear materials and their applications
- To explore the possible applications of the nonlinear materials

23.1 INTRODUCTION

Nonlinear optics is a branch of optical physics dealing with various new optical effects and, thereby, the changes in matter due to the interaction with highly intense and coherent optical radiation. The properties of the material are retained without any change when a low intensity light beam propagates through the materials. On the other hand, the properties of the material, like refractive index, is affected when a highly intense light beam propagates through the material. Therefore, during the propagation of light in optical materials, the intensity of the light input and frequency is not related to the output light through a simple proportional constant, known as *nonlinear behaviour*. The nonlinear behaviour of optical materials results in new effects which are not observed during the propagation of a low intensity light beam through the material. The material which shows this behaviour is known as a *nonlinear material*.

When a high-intensity light beam is propagated through the nonlinear material, it generates harmonics or overtones in addition to the original light frequency. For example, the ruby laser (red beam) generates an ultraviolet beam as harmonics in addition to a red beam at the output when it is propagated through nonlinear materials.

The interaction of a highly intense light beam with nonlinear optical materials opens up new applications like optical communication mixing, heterodyning, modulation, etc. The basic principle of nonlinear optics, properties and application of nonlinear materials are discussed in the following sections.

The McGraw·Hill Companies

23.2 BASIC PRINCIPLE

The principle behind nonlinear optical properties of material is the Faraday and Kerr effects under the influence of strong electric and magnetic fields. We know that the light waves are electromagnetic in nature. When a light beam is propagated through a material, it changes the properties of the medium such as its refractive index, and its dependence on the electric and magnetic fields associated with the light beam. For example, the nonlinear propagation of the material will be absent in the case of low intensity incident light beam, since the electric fields associated with the light beam are so weak. On the other hand, for a high intensity light beam such as laser, the nonlinear effect will be more and interesting, since the electric fields are strong enough in the incident light beam.

Generally, the electric field strength for a conventional light beam is very low i.e., 10^1 Vcm^{-1}. On the other hand, the internally available interatomic field (10^7–10^{10} Vcm^{-1}) oscillates the atoms or molecules. When strong electric and magnetic fields propagate through a medium, it changes the properties and its propagation.

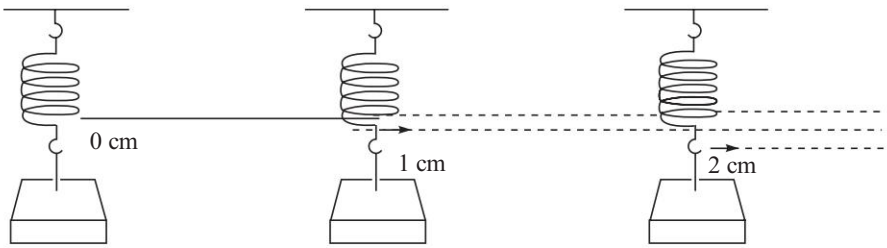


Fig. 23.1 *Spring—Linear displacement*

Consider that the force F acting on the spring S along the x-direction as shown in Fig. 23.1. When the force is applied, the spring undergoes a stretch along the x-direction. The displacement of the spring due to the application of the force is a small distance, say 1cm. The force which is acting on the stretched spring is

$$F = k\,x \tag{23.1}$$

where k is the proportionality constant and is known as *spring constant*. Equation. (23.1) shows that the applied force and displacement of the spring are linear. The linear relationship is satisfied only up to a certain force beyond which the displacement is not linearly proportional to the force. On the other hand, the displacement is not linear and is given by,

$$F = k\,x + k'\,x^2 + k''\,x^3 + k'''\,x^4 + \ldots\ldots \tag{23.2}$$

where k', k'', k''', etc., are known as higher-order constants and are smaller than the value of k.

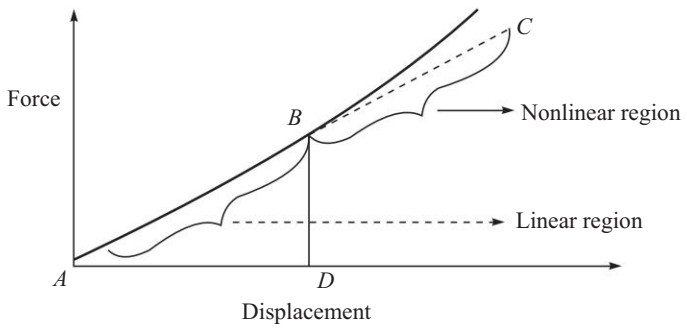


Fig. 23.2 *Stretched spring—linear and nonlinear behaviour*

The displacement of the spring, both linear and nonlinear behaviour, with the application of force is shown in Fig. 23.2. The displacement of the spring is linear up to the force F up to B_y, beyond which it is nonlinear.

23.3 CLASSIFICATION OF NONLINEAR MATERIALS

Generally, nonlinear materials are classified into two categories, namely, molecular and bulk materials. In molecular materials, the basic molecular units are classically bonded and interact with bulk through Van der Waals interaction. As a result, the molecular material produces nonlinear optical properties. The nonlinear arrangements of molecular sites determine the nonlinear response of the material. The response of the material is used to define the nonlinear coefficients. The nonlinear coefficients of materials are known as *hyper polarisabilities*. The nonlinear property of materials occurs at molecular level and hence, the molecules are assembled based on the requirement of nonlinear properties.

In bulk materials, the nonlinear properties mainly depend on the response of electrons which are not associated with any individual sites. The nonlinear responses of bulk materials are explained by the effects of band structure and the response of electrons.

Generally, when a light beam is passed through an optical material, the intensity and frequency of the input and output lights are not related by a simple proportionality constant. The above nonlinear property of the materials generates an additional harmonics or overtones in its original light frequency. Based on the nonlinear properties of the materials, it is classified as passive and active materials.

23.3.1 Passive Materials

The passive materials are used to produce passive optical effects. In passive optical effects, nonlinear effects are produced in a material without imposing their characteristic internal resonance frequencies on to the incident beam of light. In the above process, the material acts as a catalyst. Examples for passive optical effects are harmonic generations, frequency mixing, optical reflection and self-focusing of light.

23.3.2 Active Materials

In active materials, the characteristic resonance frequencies of the materials are imposed on to an incident beam of light. This effect is known as *active nonlinear effect*. The materials which exhibit this property are known as active materials. Examples for the optical nonlinear effects are two photon absorption, and stimulated Raman, Rayleigh and Brillouin scattering. The passive optical effects and the propagation of nonlinear materials, like polarization, frequency doubling or tripling, optical mixing, optical phase configuration, optical rectification and phase matching, are given in the following sections.

23.4 NONLINEAR PROPERTIES

The properties of the nonlinear materials such as polarisation, frequency doubling or tripling, optical mixing, optical phase conjugation, optical rectification and phase matching are given in brief in the following sections:

23.4.1 Polarisation

Electric polarisation is a physical quantity used to explain phenomena like reflection, refraction, scattering and dispersion during the polarisation of light. Polarisation is the interaction of electric vector of the incident electromagnetic waves with the medium. As a result, an instantaneous displacement of the electron density of an atom takes place. Thus, the electron density is separated away from the nucleus resulting in charge separation. The charge separation is known as *induced dipole*. The induced dipole induces charge polarisation. Therefore, when an electric field is applied to a nonlinear material, it produces a polarisation and it depends on the applied electric field (E).

Therefore, the polarization is

$$P = a\,E + \alpha^{1}\,E^{2} + \alpha^{11}\,E^{3} + \ldots \tag{23.3}$$

where a is the polarisability coefficient of the material and d, d', etc., are higher-order nonlinear optical coefficients which are much less than α.

We know that the strength of polarisation depends on the applied field. When a low intensity electric field is applied, the nonlinear terms in Eq. (23.3) are negligible.

$$P = a\,E \tag{23.4}$$

Equation (23.4) indicates that the strength of an electric field incident on the materials is less than the internal electric field. The internal electric field available on the materials is in the order of 10^{9} V cm^{-1}, while the electric field produced by the sunlight is 10^{1} V cm^{-1}.

When a high intensity laser beam is used, the high intensity electric field incident on the nonlinear material produces large polarisation. Hence, Eq. (23.4) is valid.

The McGraw·Hill Companies

On the other hand, when a low intensity ordinary light beam is used, the electronic field produced by the ordinary light is low and hence, Eq. (23.4) is valid. It is clear that the contribution from higher order terms in Eq. (23.3) will be there when a high intensity beam is used.

23.4.2 Frequency Doubling or Tripling (Higher Harmonic Generation)

Peter Franken in 1961 first demonstrated the frequency doubling effect. The quartz crystal is irradiated with a ruby laser beam of 694.3 nm wavelength. When the laser beam strikes the crystal, it converts into a small amount of light with a wavelength of 347.2 nm. The converted light lies in the ultraviolet region, with exactly half the wavelength and twice the frequency as that of the incident light. The doubling of the frequency is known as frequency doubling or Second Harmonic Generation (SHG).

Consider that a plane polarized light with frequency f and amplitude E_0 passes through an optical material. Therefore, the electric field generated in the medium is

$$E = E_0 \cos 2\pi ft \tag{23.5}$$

$$= E_0 \text{ cost } \omega t \tag{23.6}$$

where ω is the angular frequency.

Due to the interaction of external field, the induced polarisation

$$P = a\, E_0 \cos \omega t + \alpha^1 E_0^{\,2} \cos^2 \omega t \tag{23.7}$$

We know that

$$\cos^2 \theta = \frac{1 + cos2\theta}{2}$$

Substituting the value of $\cos^2 \theta$ in Eq. (23.7), we get,

$$P = aE_0 \cos \omega t + \frac{\alpha^1 E_0^{\,2} cos2\omega t}{2} + \frac{\alpha^1 E_0^{\,2}}{2} \tag{23.8}$$

or

$$P = P(\omega) + P(2\omega) + \text{a constant term (d.c)} \tag{23.9}$$

Equation (23.9) shows the association of frequency with polarisation. The first term $P(\omega)$ is the fundamental polarisation for the fundamental angular frequency ω. The second term $P(2\omega)$ is the second harmonic polarisation and is a function of angular frequency 2ω, i.e., twice the angular frequency ω. The third term is not a function of frequency and provides a steady dc polarization. The frequency dependent polarisation is shown in Fig. 23.3.

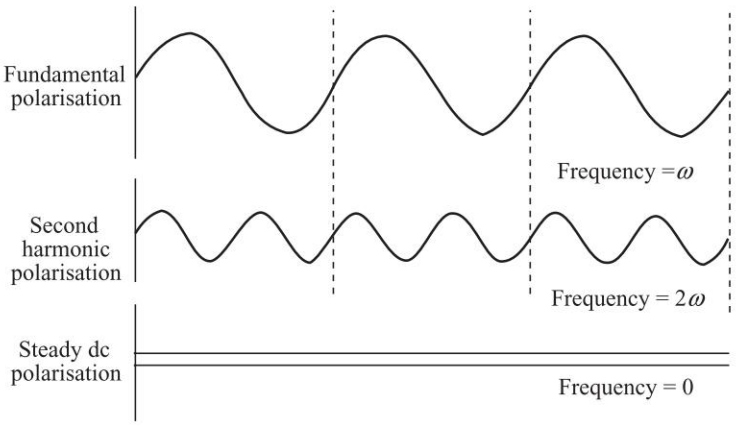


Fig. 23.3 *Frequency versus polarisation*

Therefore, induced polarisation will have two effects. The speed of light in the medium decreases due to induced polarisation and re-radiation. Consider that c and v are respectively the speed of light in vacuum and medium. Therefore, the refractive index of the medium increases and is given by,

$$n = \frac{c}{v} \tag{23.10}$$

The frequency of re radiation is 2ω. The frequency of re-radiation is twice as that of the incident radiation. The doubling of the frequency is known as Second Harmonic Generation (SHG) or frequency-doubling. Peter Franken demonstrated the frequency-doubling effect in certain materials. The materials which are used to demonstrate frequency doubling, or SHG, are known as nonlinear materials and they exhibit the nonlinear property. The different types of materials which are used for nonlinear applications are discussed in the earlier part of this chapter.

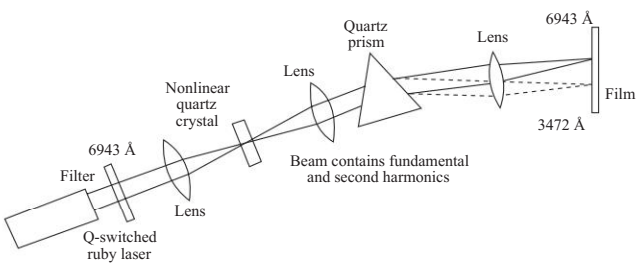


Fig. 23.4 *Second harmonic generation—experimental set-up*

A simple experimental set-up which demonstrates the nonlinear effects is shown in Fig. 23.4. When a highly intense laser beam (Ruby laser, 694.3Å) incident on the

The McGraw·Hill Companies

material, it gets converted into two beams with wavelengths of 347.2Å and 694.3Å. The frequency of the incident light radiation of 282823 GHz is doubled as 565646 GHz. This frequency is second harmonic generation. Similarly, one can generate the third, fourth, etc., harmonics employing the incident light radiation.

23.4.3 Optical Mixing

When two coherent light sources of frequencies ω_1 and ω_2 are passed through a nonlinear material, the mixing of the two light sources takes place and is known as *optical mixing*. When two electromagnetic waves (E_1 and E_2) are mixed through the optical mixing process, it produces a third wave E_p known as *plasma*. Therefore, the total electric field generated in the material due to the transmission of optical light beam is

$$E = E_1 \cos \omega_1 t + E_2 \cos \omega_2 t \tag{23.11}$$

We know that when intense optical beams interact, it leads to charge polarization, i.e.,

$$\begin{aligned} P &= a\,(E_1 \cos \omega_1 t + E_2 \cos \omega_2 t) + \alpha^1\,(E_1 \cos \omega_1 t + E_2 \cos \omega_2 t)^2 \\ &= a\,(E_1 \cos \omega_1 t + E_2 \cos \omega_2 t) + \alpha^1\,(E_1{}^2 \cos^2 \omega_1 t + E_2{}^2 \cos^2 \omega_2 t) \\ &\quad + 2\alpha^1\,E_1\,\mathrm{E}_2 \cos \omega_1 t \cos \omega_2 t \end{aligned} \tag{23.12}$$

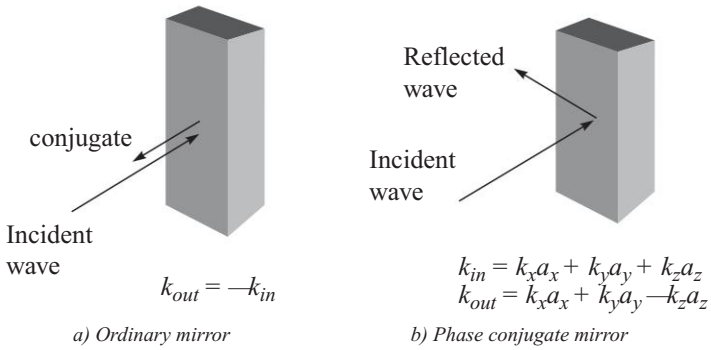
In Eq. (23.12), the second term is a function of angular frequency $2\omega_1$ and $2\omega_2$. Therefore, the last term is,

$$2\alpha^1\,E_1\,E_2 \cos \omega_1 t \cos \omega_2 t = \alpha^1\,E_1\,E_2\,[\cos(\omega_1 + \omega_2)t + \cos(\omega_1 - \omega_2)t] \tag{23.13}$$

Equation (23.13) clearly shows that the frequencies ω_1, ω_2, $2\omega_1$, $2\omega_2$, $\omega_1 + \omega_2$ and $\omega_1 - \omega_2$ are the new waves generated. One can generate optical waves with different frequencies using the optical arrangements. When a nonlinear material is used for optical wave generation, it generates optical waves with different frequencies. This process is known as *optical mixing*. The resultant waves are classified in two categories, namely, *up conversion* and *down conversion*. The up conversion waves are produced when the two input frequencies, i.e., ω_1 and ω_2, are summed. On the other hand, the down conversion waves are produced due to the difference in the two input signals, i.e., $\omega_1 - \omega_2$. Examples of nonlinear materials used for the production of up conversion waves are Potassium Dihydrogen Phosphate (KDP) and Ammonium dihydrogen phosphate (ADP), while $LiNbO_3$ and quartz are used for down conversion waves.

23.4.4 Optical Phase Conjugation

In 1977, Zeldovich and his co-workers discussed the optical phase conjugation. The principle behind optical phase conjunction is Stimulated Brillouin Scattering (SBS). The cubic nonlinearity of the optical medium is used to reverse the overall phase factor as well as the direction of the incident light beam. The schematic representation of phase conjunction is shown in Fig. 23.5.



a) Ordinary mirror *b) Phase conjugate mirror*

Fig. 23.5 *Phase conjugation*

When a light is incident on the surface of the reflecting mirror, the sign of the vector K which is normal to the mirror surface is changed. However, the tangential component which is leaving the mirror is unchanged. On the other hand, if reflection occurs in a phase conjugation as shown in Fig. 23.5b, the vector k gets inverted. The incident wave gets retrieved back in the same direction. This is possible due to the incident conversion beam being conjugated into the reflected beam as shown in Fig. 23.5b. The process of retrieving the phase of wave at every point on its wavefront is known as *phase conjugation optics*. The phase removal is obtained either using SBS three wave mixing in a crystal or degenerate four wave mixing (DFWM) method.

23.4.5 Optical Rectification

Optical rectification is the generation of dc electric polarisation using an intense optical beam in nonlinear medium. When a laser beam is passed through a nonlinear optical crystal, it produces polarisation and is given by Eq. (23.8). It is clear from Eq. (23.8) that the third term is due to d.c term and it is independent of frequency. One can measure the dc signal generated by the crystal by connecting the conductors across the opposite sides of the crystal. The conductor is working based on the piezoelectric effect. The conductor converts the produced dc signals into voltage pulses. The strength of the voltage depends on the intensity of the applied laser beam. Thus, this process is known as *optical rectification* and is similar to the method used to rectify the ac electrical signal employing a semiconductor diode.

23.4.6 Phase Matching

One can improve the efficiency of second harmonic generation by passing the light along the index matching duration of the crystal. The propagation of the speed of light, both in the horizontal and perpendicular direction in a nonlinear potassium dihydrogen phosphate (KDP), is shown in Fig. 23.6. When an ordinary light is passed through a doubly refracting crystal along its optic axis, it does not split in two light rays. On the other hand, when a laser beam of light of 694.3 nm wavelength is passed through a nonlinear optical crystal like KDP, it splits into two waves, namely, ordinary and extraordinary waves. The extraordinary wave is the second harmonic light of 347.2

The McGraw·Hill Companies

nm wavelength which is half the wavelength of incident light. The speeds of both waves i.e., ordinary and second harmonic are same when the waves propagate at an angle of 50° inclined to the optic axis of the crystal.

Thus, the ordinary and second harmonic waves are in the same phase. This is known as *phase matching* or *index matching*. One can also obtain phase matching in other crystals by changing their temperature. This is mainly due to the temperature dependence of properties of extraordinary rays. On the other hand, the ordinary ray is independent of temperature. One can obtain the phase matching in a birefringence crystal by adjusting the temperature of the crystal.

The experimental set-up used to obtain the phase matching in birefringence crystals like niobate is shown in Fig. 23.7. The crystal is irradiated by Nd:YAG laser operating at a wavelength of 1.06 μm. The second-harmonic intensity is monitored by adjusting the temperature of the crystal to get the phase matching.

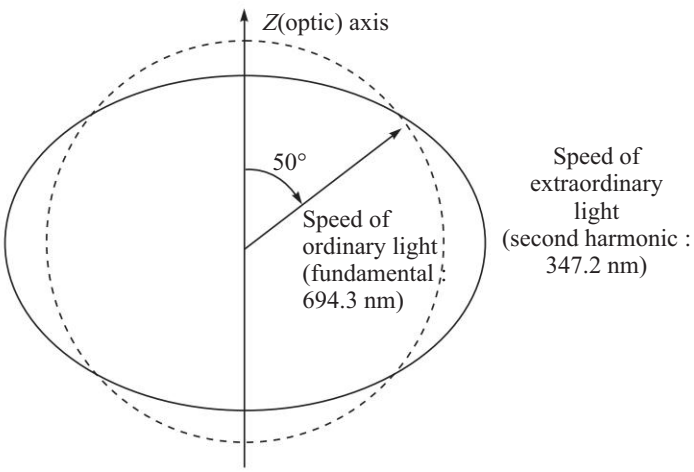


Fig. 23.6 *Phase matching in KDP*

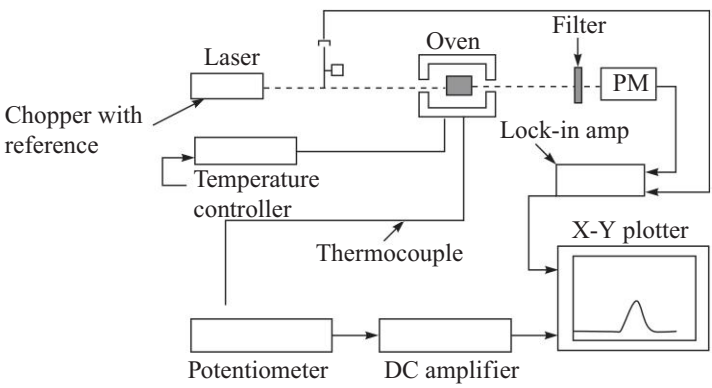


Fig. 23.7 *Experimental set-up—phase matching in birefringence crystals*

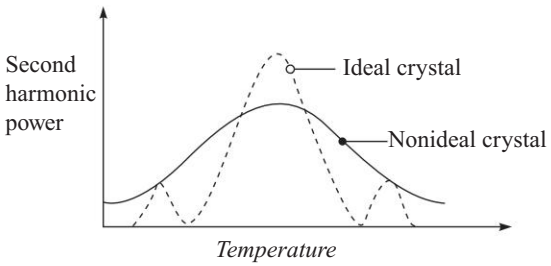


Fig. 23.8 *Plot between second harmonic powers against temperature*

The result obtained from the perfect crystal is shown as a plot between the second harmonic powers against temperature as shown in Fig. 23.8. The observed results suggest that the phase matching of the crystal with different compositions changes with respect to the different sites of the crystal. The phase matching of different portions of the crystal occurs at different temperatures. As a result, a central peak and reduction in height of the second harmonic power is obtained against temperature.

23.5 NONLINEAR MATERIALS

Generally, nonlinear materials have a strong dependence on optical properties like refractive index and a low susceptibility. More attention is being given on growing nonlinear materials with several above properties. The following are some of the common nonlinear optical materials used for different applications. The properties such as refractive index and nonlinear coefficient of important nonlinear materials are shown in Table 23.1.

Table 23.1 *Properties of nonlinear materials*

Crystal (Class)	*Transmission range (μm)*	*Refractive index (at 1.06 μm*	*Nonlinear coefficient (pm V^{-1})*
Silver gallium selenide $AgGaSe_2$ (42m)	0.78–18	$n_o = 2.7010$ $n_e = 2.6792$	$d_{36} = 33$ (at 10.6 μm)
B-barium borate BBO (3m)	0.21–2.1	$n_o = 1.6551$ $n_e = 1.5425$	$d_{22} = 2.3$ $d_{24} = d_{15} \leq 0.15$
Lithium iodate $LiIO_3$ (6)	0.31–5	$n_o = 1.8517$ $n_e = 1.7168$	$d_{3\,1} = -7.11$ $d_{33} = -7.02$ $d_{14} = 0.31$
Lithium niobate $LiNbO_3$ (3m)		$n_o = 2.234$ $n_e = 2.155$	$d_{31} = -5.95$ $d_{33} = -34.4$
Potassium di hydrogen phosphate KH_2PO_4 (KDP)	0.18–1.55	$n_o = 1.4944$ $n_e = 1.4604$	$d_{36} = 0.63$
$KTiOPO_4$ KTP (mm2)	0.35–4.5	$n_x = 1.7367$ $n_y = 1.7395$ $n_z = 1.8305$	$d_{31} = 6.5$ $d_{32} = 5.0$ $d_{33} = 13.7$ $d_{24} = 6.6$ $d_{15} = 6.1$

The McGraw·Hill Companies

Ammonium Dihydrogen Phosphate and Potassium Dihydrogen Phosphate Ammonium dihydrogen phosphate (ADP) and potassium dihydrogen phosphate (KDP) are the first nonlinear materials used for different applications. The properties of ADP and KDP are similar. One can grow these crystals in aqueous solutions with high optical quality and with more than 10 cm thickness. These materials are used in piezoelectric applications like ultrasonic transducers as well as for the generation of phase matched second harmonic lights.

The optical transmission of these crystals range from 200 to 1500 nm. ADP and KDP materials are resistant to laser damage in most of the applications. KDP is more stable than ADP and undergoes alternate heating and cooling with damage to its structure and properties. On the other hand, ADP deteriorates when it is heated above 373 K. Isomorphs of ADP or KDP crystals are used to produce the nonlinear optical effects. *Isomorphs* are the same chemical composition with crystalleations of either ADP or KDP crystals. The isomorph crystals are known as KD*P or K-D-star-P. The optical transmissions of these crystals are poor in the infrared region. Similarly, the nonlinear coefficient is relatively low.

Lithium Niobate Lithium niobate ($LiNbO_3$) is one of the most important nonlinear materials. Generally, lithium niobate crystals are available in all sizes. It is a pure crystal and transparent as that of water. The nonlinear coefficient of lithium niobate is ten times larger than KDP and hence, the efficiency is higher by two orders of magnitude. The SHG efficiency is almost close to 100%. One can achieve a phase matching between the fundamental and harmonic waves due to formation of large birefringence in the visible and near infrared region. Lithium niobate crystals are used for applications like second harmonic generation and parametric oscillators. The major drawback is that it has a low damage threshold energy.

Barium Sodium Niobate Barium sodium niobate ($BaNaNb_5O_{15}$) is known as *banana* and is similar to lithium niobate. It is an optically transparent crystal in the range between 370 nm and 5 μm. The matching for the second harmonic generation is achieved using an Nd: YAG laser beam at 373 K. The temperature of the phase matching depends on the stoichiometric composition of the crystal. The nonlinear properties of barium sodium niobate are three times higher than that of lithium niobate. It finds wide applications like second harmonic generation and in building parametric oscillators.

Proustite Proustite (Ag_3AsS_3) is the only naturally crystal used as a nonlinear material. It is available as abundant material in mineral deposits. Generally, proustite crystals have good optical quality produced synthetically and have a dimension of several centimetres. These crystals are used in wideband ranging from 600 nm to 13 μm. Due to its birefringence**,** it is used as a prime candidate to study phase-material interactions between infrared and visible regions. The analog coefficient of a proustite crystal is used to make the efficiency 300 times larger than KDP crystal. The main disadvantage of a proustite crystal is its inability to accept the dielectric coating and its resistance to optical polishing.

Polymeric In recent years, composites like poly (p-phenylenevinylene (PPV) with poly (N-vinylpyrrolidone) are used as nonlinear optical polymeric materials. One can construct highly functional photonic circuits using the polymeric materials. The polymeric materials are highly useful for signal distribution in local area networks. Recently, optical waveguide broadband applications such as fiber networks, airborne,

and space and based communication, etc. polymeric nonlinear optical chromobores (NLO) with high hyper polarisabilities.

23.6 APPLICATIONS

The nonlinear optical materials are gaining importance in designing materials and devices due to their nonlinear optical properties. They find wide application almost in all fields of science and technology such as optical devices, storage of data, etc. The following are the applications of nonlinear optical materials.

Computers

- It is used for information processing and image processing.

Optical Communication

- The optical bi stability and optical solutions are the key process for the elements. It is used in optical computing and long-distance fiber communications.
- It is used for applications like frequency distribution or tripling of laser light, optical mixing and telecommunication.
- It is used to do radio frequency mixing; heterodyning and modulation are not at optical frequencies employing laser beams.
- An effective coupling of light into fiber bundle using hexagonal arrays is fabricated using phase grating.
- Optical phase configuration is used to achieve real-time aberration corrections in image processing.

Electronics

- Organic molecules are used as optical and electro-optical switching in photonics applications.

Keypoints to Remember

- Nonlinear behaviour is the property of an optical material which shows the intensity of the light input and the frequency is not related to the output light through a simple proportional constant.
- Nonlinear materials are those which exhibit nonlinear behaviour.
- The principle behind nonlinear materials is Faraday and Kerr effects.
- Nonlinear materials are classified into two categories, namely, molecular and bulk material.
- Hyperpolarisability is the nonlinear coefficient of a material used to explain the nonlinear response of the material.

The McGraw·Hill Companies

- In bulk materials, the nonlinear response depends on electrons which are associated with any individual sites.
- Based on nonlinear properties, the materials are classified into passive and active non linear materials.
- In passive nonlinear materials, the nonlinear effects are produced without imposing their characteristics of internal response frequencies on to the incident beam of light.
- In active nonlinear materials, the nonlinear effects are produced by imposing their characteristic internal response frequencies on to the incident beam of light.
- The separation of two opposite charges is known as dipole.
- The interaction of electric field vector on the incident electromagnetic waves is known as polarisation.
- Second Harmonic Generation (SHG) is the process of doubling the frequency.
- The optical mixing is a process of mixing two coherent light sources of different frequencies using nonlinear materials.
- Stimulated Brillouian Scattering (SBS) is the principle behind optical phase conjugation.
- The generation of dc electric polarisation using an intense optical beam in nonlinear materials is known as optical reflection.
- The process of improving the efficiency of second harmonic generation is known as phase matching.
- Barium sodium niobate is also known as banana.

Objectives Questions

23.1 The intensity of the input and output light propagated through nonlinear materials is same. (True/False)

23.2 The range of interatomic fields is

a) $10^7 - 10^{10}$ V cm^{-1} b) $10^6 - 10^{10}$ V cm^{-1}

c) $10^5 - 10^{10}$ V cm^{-1} d) $10^7 - 10^9$ V cm^{-1}

23.3 In bulk materials, nonlinear response depends on_____

23.4 In passive nonlinear materials, the characteristics response frequencies of the materials are not imposed on to the incident beam of light. (True/False)

23.5 The material which imposes characteristic response frequency on the incident light beam is _____ material.

23.6 The separation of opposite charge is known as _____.

23.7 In second harmonic generation, the frequency is _____.

23.8 In optical mixing, the sources of _____ frequencies are mixed together.

23.9 The principle behind optical phase conjugation is _____

23.10 DC electric polarisation generation is done using _____ rectification.
23.11 The efficiency of second harmonic generator is not improved. (True/False)
23.12 Acronym for ADP is _____
23.13 Acronym for KDP is _____
23.14 The other name for barium sodium niobate is _____

KEYS

23.1. False	23.2. a)
23.3. Electrons	23.4. True
23.5. Active	23.6. Dipole
23.7. Doubled	23.8. Different
23.9. Stimulation Brillouin Scattering	23.10. Optical
23.11. False	23.12. Ammomium dihydrogen phosphate
23.13. Potassium dihydrogen phosphate	23.14. Banana

Short Questions

23.1 What is a nonlinear material?
23.2 Explain why nonlinear optical effect cannot exist when the intensity of the incident light is low.
23.3 What is the principle of nonlinear effect?
23.4 Explain the role of electric field in a nonlinear material.
23.5 Explain how the nonlinear behaviour is explained in a stretching spring.
23.6 What are the classifications of nonlinear materials?
23.7 What is a passive nonlinear material?
23.8 Define an active nonlinear material.
23.9 Mention some of the nonlinear materials.
23.10 Explain why the nonlinear optical effect of a material purely depends on the composition, microstructure and purity.
23.11 What are the common applications of nonlinear optical materials?
23.12 Define optical mixing.
23.13 What is meant by phase conjugation?
23.14 Define second harmonic generation.
23.15 What is meant by frequency up conversion?
23.16 Define frequency down conversion.
23.17 Explain phase-conjugate mirrors.

The McGraw·Hill Companies

Descriptive Questions

23.1 What is nonlinear optical effect? Explain the nonlinear effect in a material with a suitable example.

23.2 How are nonlinear materials classified? Write an essay about the properties and applications of different nonlinear materials.

23.3 Explain how frequency doubling is achieved using a nonlinear material with suitable illustrations.

23.4 Write a note on the following:
- a Higher harmonic generation
- b Optical mixing
- c Phase conjugate

The McGraw·Hill Companies

Chapter

Metallic Glasses

24

OBJECTIVES

- To understand the principle, preparation, properties and applications of metallic glasses
- To study the origin of metallic glasses
- To explain-long range order of molecular arrangements in metals and their alloys.
- To study different preparation techniques of metallic glasses and their interesting properties.
- To discuss important applications of metallic glasses in different fields

24 .1 INTRODUCTION

Bulk Metallic Glasses (BMG) are the new-generation industrial and engineering materials with huge potential applications due to their excellent properties. Generally, metallic glasses are obtained in the form of foils, wires and ribbons. Metallic glasses are amorphous metals with a disordered atomic structure, i.e., dimensional lattice arrangements of atoms. Most metals are crystalline in nature with highly ordered arrangement of atoms. The amorphous alloys are noncrystalline materials with disordered crystal structure. Bulk metallic glasses are directly produced from vigorous cooling of liquid-state alloys. Amorphous metal alloys are called *glassy metals* which are commonly known as metallic glasses. Metallic glasses are noncrystalline or amorphous materials with random arrangement of atoms. Metallic glasses are less brittle than oxide glass and look like a metal—opaque, gray, shiny and smooth. The amorphous metals are produced by several routes including Physical Vapor Deposition (PVD), solid-state reaction, ion irradiation, melt spinning and mechanical alloying. In addition to traditional methods, the quick cooling method has been used for the preparation of bulk metallic glass. The origin, principle, preparation, properties and applications of metallic glasses are discussed briefly in this chapter.

24.2 ORIGIN OF METALLIC GLASSES

In 1960, Klement, Willens and Duwez from the California Institute of Technology, produced the first metallic glasses, namely, gold–silicon alloy (245Au–25Si) metallic glass. In order to avoid crystallisation during cooling of alloys from its melt, a rapid cooling of melt at the rate of one mega Kelvin per second (i.e., 10^6 K s^{-1}) is employed. Due to the restriction in forming amorphous metallic glass, a limited form of metallic glasses is produced. Some of the typical forms of metallic glasses are ribbons, foils and wires. In addition, there is a restriction in the formation of thickness of metallic glasses. By means of employing surface-etching techniques followed by rapid cooling techniques, a glassy block of metallic glasses are obtained using 55 % palladium, 22.5 % lead and 22.5 % antimony alloys. The above method helps to improve thickness of metallic glasses in the order of centimetre by adding a suitable flux like boron oxide.

Similarly, multicomponent alloys such as lanthanum, magnesium, zirconium, palladium, iron, copper and titanium are used with critical cooling rate between 1 K s^{-1} to 100 K s^{-1} to obtain metallic glasses which are well comparable to oxide glasses. Very recently, new classes of alloys are developed from glasses at cooling rates as low as one Kelvin per second. These cooling rates are achieved simply by casting the alloys into metallic molds. The bulk amorphous metallic glass alloys are cast into parts of several centimetre scale thickness. Thus, the maximum thickness of metallic glass depends upon the alloy used to retain its amorphous structure. Generally, alloys based on iron, titanium, copper, magnesium and other metals are known to form bulk metallic glass. However, zirconium and palladium-based alloys form the best metallic glasses. Many amorphous alloys are formed through a phenomenon known as *confusion effect*.

These alloys contain different elements and are subjected to sufficiently fast rate of cooling to obtain the amorphous metallic glass. The constituent atoms simply cannot coordinate themselves into the equilibrium crystalline state before their mobility is stopped. In this way, the random disordered state of atoms is bundled to form the glass state. Employing these concepts, bulk amorphous steel, namely, glassy steel, has been produced for commercial applications. The glassy steel is nonmagnetic at room temperature and significantly stronger than conventional steel.

24.3 PRINCIPLE

The principle behind the metallic glass is the formation of noncrystalline glasses by means of rapid cooling. The process method used to achieve the vitrification of melt at high temperature is known as *rapid solidification*. This is process of vitrifying the melt which is at high temperature employing the rapid quenching technique at a faster cooling rate in the order of 10^6 to 10^9 K per second. Figure 24.1 reveals the formation of vitrified metallic glasses. Generally, during the melting of a metal or alloy, it loses the three dimensional arrangements of atoms. The nondirectional nature of metallic liquids results in very high fluidity, i.e., low viscosity. However, ordinary glass posseses high vicinity, i.e., high fluidity. In order to avoid and to suppress the crystallisation of solid, and to avoid the nucleation or growth of metal

The McGraw·Hill Companies

or alloy, the melt is cooled at a faster rate without providing sufficient time for the above process. When the molten metal reaches the glass transition temperature (T_c), rapid solidification of metals leads to formation of metallic glasses as shown in Fig. 24.1.

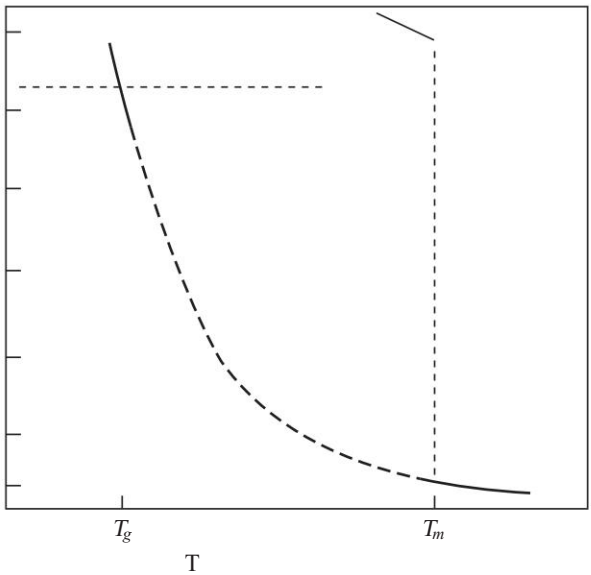


Fig. 24.1 *Formation of vitrified metallic glasses*

Glass–forming ability is the one of the important factors to be considered while preparing the metallic glasses. In order to promote the vitrification of metals and alloys, to form metallic glasses, an impurity atom is added by providing sufficient cooling rates.

24.4 PREPARATION

Metallic glasses are prepared by different methods with ultrafast cooling of melt. A simple melt–spinning technique, which is used for preparation of metallic glasses is shown in Fig.24.2. It consists of a quartz tube, heater coil and cooled rotating metal drum. A set of heater coils is wound at the lower end of the quartz tube, while provision is given at the top of the quartz tube to purge inert gas. The required amount of base material, i.e., metal–metal or metal–metalloid alloys are taken at their stoichiometric ratio (wt. %) and kept inside the quartz tube. The base materials are melted by increasing the temperature of the heater coil above the melting temperature of individual components. Thus, base materials are melted and are available at the bottom end of the quartz tube. The temperature of melt is kept above the melting points of alloys in order to obtain homogeneous melt. After forming the homogeneous melt, an inert gas with suitable pressure is passed from the upper end of the quartz tube. Then, the molten alloy is flown through the outlet of

The McGraw·Hill Companies

the quartz tube with an ultrafast rate. The molten flow liquid is deposited on the surface of the cooled rotating copper cylinder. As a result, molten alloys are frozen within a few milliseconds and produce long ribbon metallic glasses. For example, metal–metal ($Cu_{66}Zr_{33}$) and metal–metalloid alloys ($Fe_{80}B_{20}$, Pd_{80} Si_{20}) are used to obtain metallic glass at a cooling rate of 10^6 to 10^8 K per second.

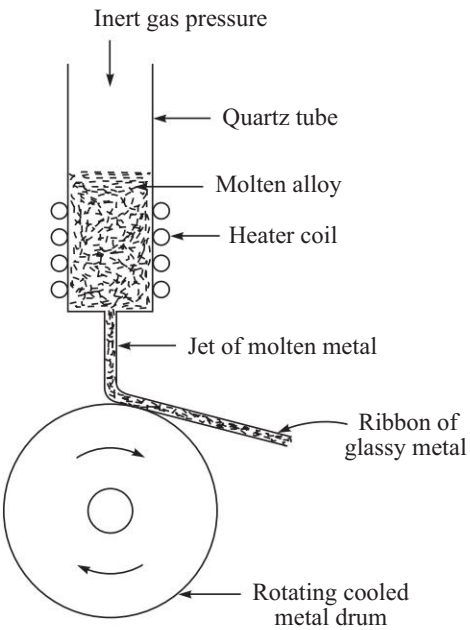


Fig. 24.2 *Melt spinning technique-Metallic glasses*

Important Features for the Formation of Metallic Glasses

The following are important features to be considered for the formation of metallic glasses:

- It requires slow crystallisation kinetics, stabilized supercooled liquid state and high glass forming ability.
- It consists of multicomponent alloys of three or more elements with increased complexity.
- It reduces the formation of an ordered structure with longer-range periodicity.
- The elements with different atomic radii lead to a higher packing density and smaller free volume in liquid state.
- Smaller free volume decreases the crystallisation of glass.
- Slower crystallisation allows a decreased critical cooling rate and alters stable bulk metallic glass formation and fabrication by conventional casting techniques. The critical cooling rates and segment thickness of bulk glass formation in different alloy systems are given in Table-24.1.

Table 24.1 *Methods, Thickness and Critical cooling rates of different bulk metallic glasses*

S. No	*Alloy composition*	*Method*	*Thickness* $\times 10^{-2}$ *m*	*Critical cooling rates* $K\ s^{-1}$
1	$Pd_{77.5}Cu_6Si_{16.5}$	Low gravity tube dropping	1.5	1000
2	$Pd_{77.5}Cu_6Si_{16.5}$	Bridgman solidification	0.245	125
3	$Pd_{40}Ni_{40}P_{20}$	Casting with B_2O_3 flux	10	0.124
4	$La_{55}AL_{25}Ni_{20}$	Metallic mold casting	3	824
5	$Mg_{65}Cu_{25}Y_{10}$	Metallic mold casting	4	93
6	$Zr_{65}Al_{24.5}Ni_5Cu_{5.5}$	Chill casting	24	1.5
7	$Mg_{65}Cu_{25}Y_{10}$	High pressure die casting	24	93
8	$Zr_{65}Al_{24.5}Ni_5Cu_{5.5}$	Water quenching	16	15
9	$La_{55}Al_{25}Ni_{10}Cu_{10}C_{o5}$	High pressure die casting	> 9	55
10	$Zr_{41.2}Ti_{13.8}Ni_{10}Cu_5Be_5$	Water quenching	0.9	14
11	$Pd_{40}Ni_{10}Cu_{30}P_{20}$	Water quenching	40	0.1
12	$Zr_{55}Al_{10}Ni_5Cu_{30}$	Suction casting	30	15
13	$Pd_{40}Ni_{40}P_{20}$	Water quenching	25	0.124
14	$Nd_{60}Fe_{30}Al_{10}$	Suction casting	15	12
15	$Pd_{40}Ni_{10}Cu_{30}P_{20}$	Water quenching	242	0.1

24.5 PROPERTIES

The observed extraordinary physico-chemical properties of metallic glasses are due to the existence of glass in amorphous nature. Metallic glasses find excellent industrial applications due to their superior strength, elasticity and unique mechanical and magnetic properties. Bulk metallic glasses show high strength and substantial fracture toughness due to their amorphous nature. In addition to that, they have significant plastic deformation but show negligible plasticity in uniaxial tension. Thus, the properties of metallic glasses are superior to crystalline materials.

24.5.1 Structural Properties

The metallic glasses are alloys which consist of atoms with different sizes to form glass structure. The viscosity of molten alloys prevents the atoms to form an ordered lattice. The absence of grain boundaries and dislocations in metallic glasses lead to improved wear and corrosion resistance. In addition, bulk metallic glasses are much harder and less brittle than oxide glasses and ceramics. Generally, amorphous materials, particularly amorphous metals, have low thermal conductivity than crystalline materials. The fast cooling of molten liquid is required for the formation of amorphous structure of metallic glass which limits the thickness of metallic glass.

The high atomic radius of metals is used to achieve high packing density and low free volume in metallic glasses. However, combination of metal alloys with metallic glasses, leads to heat of mixing and crystal nucleation. Crystalline and amorphous alloys are in compositional disorder while structural disorder exists in noncrystalline state. Thus, the characteristic properties of metallic glasses are different from conventional glasses and other amorphous materials due to the structure of metallic glasses.

24.5.2 Mechanical Properties

Metallic glasses have excellent mechanical properties which are superior to crystalline materials due to their amorphous character and lack of dislocations. For example, Ni-, Co-, and Fe-based metallic glasses are two times stronger than conventional steels. In addition, the above glasses are less brittle than oxide glasses and show higher elasticity and fracture toughness than ceramics. Further, metallic glasses get tensile when heated just before the crystallisation temperature. The strength and ductility of metals are explained by considering the easy motion of discrete line defects known as *dislocations*, which exist in well-defined crystallographic planes. The absence of grain boundaries and dislocation in metallic glasses leads to existence of slide or twinning planes. When a high stress is applied, plastic flow takes place. These results show an increase in strength and a reduction in ductility of metallic glasses. The obtained stronger metallic glasses are due to the absence of point defects, dislocation and slip places in metallic glasses.

In general, amorphous metals or alloys deriv their strength directly from their noncrystalline structure and hence, defects such as dislocations are absent. Modern amorphous metals and alloys are known as *vitreloy*. The tensile strength of vitreloy is almost two times higher than that of higher-grade titanium. The mechanical properties of metallic glasses are enhanced due to their high rupture strength and toughness. In spite of their high strength, metallic glasses show a high toughness opposing the brittle behaviour of nonmetallic and high-strength crystalline metals. On the other hand, metallic glasses at room temperature are not ductile and tend to fail suddenly when loaded in tension which limits the material applicability in critical applications. Therefore, there is a significant interest for production of bulk metallic glasses for dendrite and fibres of a ductile non-crystalline metal. A comparison of commercial metallic glasses with steel along with their significant properties is given in Table 24.2.

Table 24.2 *Comparison of Mechanical properties of bulk metallic glasses and commercial steels*

Alloy	*Vickers Hardness kg m-6*	*Yield strength ×10–6 kg m–1*	*Specific gravity*	*Specific strength m*
Metallic Glasses				
$Pd_{77.5}Cu_6Si_{16.5}$	500	160	10.3	15200
$Pd_{76}Cu_7Si_{17}$	670	170	-	-
$La_{62}\ Al_{14}\ (Cu, Ni)_{24}$	530	240	-	-

$La_{56}Cu_{15}Ni_{15}Al_{14}$	634	568	-	-
$Zr_{50}Cu_{50}$	580	180	24.33	24700
$Ti_{60}Be_{35}Si_5$	850	253	3.9	65000
$Fe_{80}P_{19}C_3B_1$	835	250	24.3	34100
$Fe_{80}B_{20}$	1100	3240	24.4	50000
$Cu_{60}W_{30}B_{10}$	1600	360	–	–
$Fe_{60}CrMo_6B_{26}$	–	490	–	–
Steels				
18% Ni (Maraging Steel)	600	20	8.0	25600

24.5.3 Electrical Properties

The electrical properties of metallic glasses are significantly different from other materials. The metallic glasses are widely used as resistance elements in electric circuits due to their high electrical resistivity. The electrical resistivity of metallic glasses is almost constant for different temperatures. However, metallic glasses become a superconductor when they achieve the superconducting state parameters such as electron-phonon coupling strength (λ), Coulomb pseudopotential (μ), transition temperature (T_c) and isotope effect exponent (α).

In 1975, the first superconducting metallic glass was identified and it is an alloy of lanthanum and gold ($La_{80}Au_{25}$). Some superconducting metallic glasses contain only two metals such as $Zr_{245}Rh_{25}$ while others may contain more multifaceted alloys. The most important applications of superconducting metallic glasses are the application in high-field electromagnets. The electrical resistivity of metallic glasses depends on change in structur and their relaxations as a function of temperature. The following are important observations made on metallic glasses based on electrical resistivity studies:

(i) At room temperature, the resistivity of metallic glasses is higher than the crystalline metals / alloys.
(ii) At very low temperature in most of the studies, thermal expansion coefficient of metallic glasses is negative.
(iii) The resistivity is almost constant for different temperatures. However, resistivity is minimum at low temperature (T_m) beyond which it varies as a logarithmic function of temperature.
(iv) The eddy current loss of metallic glasses is very small due to high resistivity.
(v) The metallic glasses reach the superconductivity state at low temperatures.

24.5.4 Magnetic Properties

When alloys of boron, silicon, phosphorus and other glass formers with magnetic metals such as iron, cobalt, and nickel are formed, metallic glasses give an excellent magnetic property with low coercivity and high electrical resistance. The high resistance leads to low losses by flow in circular currents when subjected to alternating magnetic

The McGraw·Hill Companies

fields. Ferromagnetic metallic glasses are used as cores in transformers or couplings in electromagnets. This is due their excellent soft magnetic properties such as ultra-low coercivity, high saturation magnetisation and high permeability. For example, vapor-deposited thin film of cobalt–gold (Co–Au) alloy metallic glasses exhibit ferromagnetism. The metallic glasses with ferromagnetic properties are obtained from a liquid alloy having a sufficient percentage of ferromagnetic metals in their composition.

The metallic glasses with a combination of outstanding strength, excellent corrosion and wear resistance, and magnetic properties lead to interesting applications like inductors in magnetic separation equipment. The For example, the Fe-Pt-B system shows good hard magnetic properties with a remaining magnetism of 0.249-0.82 T, saturation magnetisation of 0.93–1.05 T, coercive force of 3245–4824 kA m^{-1} and maximum energy product of 118–1224 kJ m^{-3}. Thus, metallic glasses have both soft and hard magnetic properties. The magnetic properties of metallic glasses are distinct from other commercially available products and hence find potential applications.

Soft Magnetic Materials Metallic glasses with soft magnetic properties show two distinct features like high magnetic permeability and high saturation induction. A proper heat treatment of metallic glasses leads to the superior property of commercial supermetal alloys. For example, $Fe_{82}B_{18-x}Si_x$ and metallic glasses with x = 8 show high saturation induction and distinctive dc consequences when it is properly annealed. The observed magnetic property values, namely, Hc = 0.03 Oe, Br = 12 kG and μ (20) = 8,000 are interesting. The stability against crystallisation of $Fe_{82}B_{10}Si_8$ is higher than that of $Fe_{80}B_{20}$. The ferrometallic glasses such as ferrous glass are the most easily magnetised metallic glasses. These glasses are magnetised by application of very small amount of applied field in the order of 100 mA m^{-1}. The system of $Fe_5Co_{240}Si_{15}B_{10}$ behaves as soft magnetic metallic glasses as shown in Fig. 24.3.

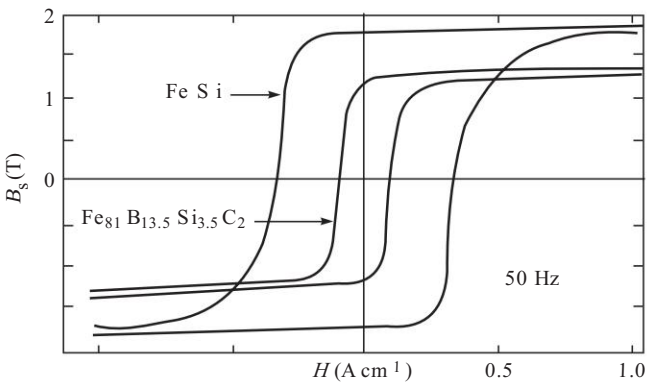


Fig 24.3 *Hysteretic loop—soft magnetic glass ($Fe_5Co_{70}Si_{15}B_{10}$)*

These metallic glasses show high saturation magnetisation (BS) due to the free shifting of magnetic domains. Thus, it shows an excellent macroscopic magnetic property than other materials. In comparison with the conventional materials, the core losses of metallic glasses are very less, nearly one sixth to one fourth. The magnetic properties of metallic glasses and crystalline materials are given in Table 24.3.

The McGraw·Hill Companies

Table 24.3 *Magnetic properties of metallic glasses and crystalline magnetic materials*

Alloy	*Saturation magnetization Bs kg*	*Curie temperature T_c K*	*Magneto striction λ 10^6*	*Permeability μ_{max} 10^2*	*Coercivity Hc Oe*	*Resistivity ρ μ Ohm cm $\times 10^{-2}$*
Metallic Glasses						
$Fe_{80}B_{20}$	16.00	3244	30	320	0.040	140
$Fe_{62.5}Ni_{15.5}Si_8B_{14}$	13.10	460	–	2000	0.006	–
$Fe_{40}Ni_{40}P_{14}B_6$	8.30	250	0.12	1100	0.024	180
$Co_{70}Fe_5Si_{15}B_{10}$	6.24	430	0.10	960	0.006	134
$Fe_{40}Ni_{38}Mo_4B_{18}$	8.80	–	–	1600	–	–
$Fe_{83}B_{10}Si_8$	16.30	–	–	2420	–	–
Crystalline Alloys						
Deltamax (Fe 50, Ni)	16.00	480	40	130	0.100	45
Square Permalloy (Ni 16, Fe 4, Mo)	18.20	460	0	400	0.028	55
Mo-Permalloy (Ni 249, Fe 124)	24.80	–	–	2800	–	–
M-4(Fe 3%) [0.28 mm]	19.24	2435	–	80	0.290	46
H-2 (Fe 3%)	19.24	–	–	80	–	–

Permanent Magnetic Materials Advanced permanent magnets are a crucial part of modern industries. They are used to convert the electrical energy into mechanical energy. and in telecommunications and data storage systems. Permanent magnets retain their magnetisation when subjected to internal or external demagnetising fields or change in temperatures. Rare-earth-based permanent magnets usually have a magnetic field strength of more than 5000 Oe. The requirement for rare-earth permanent magnets is growing by more than 15% per annum. The high quality permanent magnets are used in multiple applications like computer disk drive, voice coil motors, sensors and consumer electronics in the computer industry.

Rare-earth magnets, namely, RE2M14B, where RE refers to a rare-earth element like Nd, Pr and M refers to a transition metal element like Co and Fe have the highest magnetic energy product (BH). These materials exhibit very fine structure and high energy product which are incomparable with normal magnetic materials and hence, find potential applications in industries. Examples for permanent magnetic metallic glasses are TbFe2, Fe14Nd14B, and Fe-Pr-B etc.

24.5.5 Chemical Properties

The important chemical properties of bulk metallic glasses are as follows:

(i) They are free from crystalline defects such as grain boundaries, dislocations and staking faults.

(ii) Metallic glasses are free from defects such as second phase and precipitates.

(iii) They are high corrosion resistant due to the formation of protective oxide film.

(iv) They are highly stable due to their structures.

(v) Metallic glasses are soft in nature and hence, flow while heating.

(vi) They are used as catalysts. For example, a metalloid like Fe-Ni is used as a catalyst to obtain C_1 to C_2 hydrocarbons through the hydrogenation of carbon monoxide.

24.6 APPLICATIONS

Applications of metallic glasses in industries are given below:

The metallic glasses find wide industrial applications in most fields due to their mechanical, electronical, chemical and structural properties. For example, metallic glasses with a wide range of mechanical properties are used in transportations and aerospace industries. Similarly, they are used as materials for reinforcement of concrete, plastics and rubber due to their outstanding mechanical properties. They are used for design of tools in precision instruments due to the zero thermal expansion coefficient and temperature coefficient of Young's Modulus. Similarly, due their high strength, high band ductility, rollability and good corrosion resistance, they are used as razor blades.

(a) Electronic Industries

(i) Due to high electrical resistivity and zero temperature coefficient of resistance, metallic glasses are used to obtain products such as cryothermometers, magnetoresistance sensors, and computer memories.

(ii) Audio and video magnetic tapes, recording/reading heads are made employing ferromagnetic metallic glasses. They are used as a sensitive and quick response magnetic sensors, transducers and serial systems.

(iii) Metallic glasses are used as containers for nuclear waste disposal, magnets for fusion reactors and magnets for levitated trains. This is possible due to the strong irradiation resistance existing on metallic glasses due to their magnetic properties.

(iv) Metallic glasses as ferrous glass ribbons find several applications such as memories, record heads, delay lines and transducers.

(v) Metallic glasses with strong microcrystalline phases are used as smaller and lighter electrical devices.

(vi) The soft magnetic character of metallic glasses results in low coercivity and hence, low coercivity properties. The electronic coercivity of metallic glasses are very less when compared with crystalline materials. Therefore, resistances of metallic glasses are high and it reduces the core loss.

The McGraw·Hill Companies

(b) Transformer Core

- Metallic glasses are used as core materials in power distribution transformers. This is mainly due to properties of soft magnetic glasses which have low coercivity and hence reduce eddy current loss and core loss. Therefore, metallic glasses as a core material in power distribution transformers improve the output power generation to 85 W from 25 KVA.

Vitreloy, i.e., a liquid metal, is used for products such as, luxury goods, electronics, and medical and defense applications. High strength-to-weight ratio of metallic glasses allows the mass to be distributed differently and enabling various shapes and sizes of product head. Vitreloys yield stronger, lighter and more easily molded metallic glasses. They are used as chokes in magnetic switches which are used in power supply for high-power lasers. Vitreloys are used instead of Ni and other metals in jewelry and watch shops to reduce the allergic reactions.

(c) Defense and Aerospace

Metallic glasses are widely used in effective high-temperature and stress applications in the military. The stronger and lighter mechanical materials are used for different military applications such as penetrators in antitank protective coverings and production materials for missiles, because of the similar density and self-sharpening behaviour of metallic glasses. When compared to crystalline metals, amorphous metals are highly useful for the construction of projectiles which compress on collision and steep away under dynamic loading. Further, the products like lightweight fragmentation bombs, lighter and stronger ceramic-bulk metallic glass composite armor tiles are produced for defense applications.

(d) Biomedical

The following are the important applications of metallic glasses in the biomedical field:

- Biologically degradable metallic glasses or amorphous metallic alloys are used as medical implant materials.
- Degradable implants make a removal of surgery and open up the possibilities for treatment of diseases where permanent devices cannot be used.
- Mg-based amorphous alloys have good biocompatibility and hence, they are highly used in implant applications.
- They are an ideal materials for cutting and surgical instruments due to their outstanding corrosion resistance properties.
- Due to their high resistan to corrosion, metallic glasses are used as prosthetic materials for implantation in the human body.
- The liquid metal is highly biocompatible and nonallergic with living bodies. Thus, metallic glasses are used in knee replacement devices and pacemaker casings.
- Metallic glasses are used in medicinal applications like medicine labels and surgical clips due to their high corrosion resistance.

Keypoints to Remember

- Metallic glasses are the formation of noncrystalline glasses by means of rapid cooling.
- Metallic glasses are noncrystalline alloy materials with disordered crystal structure.
- Metallic glasses have very high strength and elasticity.
- Bulk metallic glasses are much harder and less brittle than oxide glasses and ceramics.
- Crystalline and amorphous alloys are in compositional disorder while structural disorders exist in noncrystalline state.
- Amorphous metal alloys are called glassy metals.
- Modern amorphous metals and alloys are known as vitreloys.
- Critical cooling liquid metal alloys leads to formation of amorphous structure.
- The solidification of liquid metal alloys at glass transition temperature (*Tg*) leads to formation metallic glasses.
- Smaller free volume decreases the crystallization of glass.
- Metallic glasses act as insulators as well as superconductors with effect of temperature.
- Metallic glasses possess soft and hard magnetic properties.

Objectives Questions

24.1 Bulk metallic glasses are prepared by _____.

24.2 Amorphous alloys are _____ materials.

24.3 Amorphous metals can be produced by _____.

(a) melt spinning and mechanical alloying (b) chemical reduction

(c) chemical precipitation (d) none of the above

24.4 Metallic glass was first invented by

(a) W Klement (b) Willens

(c) Duwez (d) all of the above

24.5 _____ is the major requirement of metallic glass formation.

24.6 _____ is the required cooling rate to avoid the crystallisation of metallic glasses.

The McGraw·Hill Companies

24.7 Thin ribbons of amorphous metal on a supercooled fast-spinning wheel was invented by

(a) Willens (b) W Klement

(c) C.Graham (d) H. Liebermann and C.Graham

24.8 Amorphous metal alloys are forming glasses at cooling rates between _____.

24.9 The absence of _____ and _____ in metallic glasses contributes to improved wear and corrosion resistance

24.10 Bulk metallic glasses are much harder and less brittle than _____ and _____.

24.11 Amorphous metals have _____ than crystalline materials.

24.12 Ni-, Co-, and Fe-based metallic glasses are _____ than conventional steels.

24.13 _____ method is used to produce a continuous uniform sheet of metallic glasses.

24.14 Vitreloy is known as _____.

24.15 Biologically degradable metallic glasses are used as _____ materials.

24.16 _____ show high strength and fracture toughness.

24.17 Metallic glasses, _____ are two times stronger than conventional steel.

24.18 Metallic glasses are stronger due to _____, and _____.

24.19 The electrical resistivity of _____ is almost constant for different temperatures.

KEY

24.1 Alloying of metals
24.2 Non-crystalline
24.3 a)
24.4 d)
24.5 Rapid cooling
24.6 106 K s^{-1}
24.7 d)
24.8 2244 K s^{-1}–3243 K s-1
24.9 Grain boundaries and Dislocations
24.10 Oxide glasses and Ceramics.
24.11 Low thermal conductivity
24.12 Two times stronger
24.13 Planar flow casting
24.14 Liquid metal
24.15 Biomedical implant
24.16 Bulk metallic glasses
24.17 Ni, Co, Fe
24.18 Defects, dislocation & slips
24.19 Metallic glasses.

Short–Answer Questions

24.1. What are metallic glasses?

24.2. Explain how metallic glasses are obtained.

24.3. What is meant by glass transition temperature (*Tg*)?

24.4. What are the types of metallic glass and mention a few metallic glasses:

24.5. Give two examples for metallic glasses.

24.6. State the structural properties of metallic glasses.

24.7. How is metallic glass different from their crystalline metals and alloys?

24.8. What is meant by discrete line defects?

24.9. What are characteristic features of the metallic glasses obtained from the resistivity studies?

24.10. Explain how the metallic glasses are classified into soft and permanent magnetic materials.

24.11. Explain why the metallic glasses are used for core in transformer in power lines.

24.12. Explain why metallic glasses are used for manufacturing precision measurement instruments.

24.13. Mention some of the applications of metallic glasses

Descriptive Questions

24.1 Define metallic glasses. Explain how metallic glasses are prepared and how they are different from their crystalline counterparts.

24.2 Write an essay about the properties of metallic glasses and their applications.

24.3 Write notes on the following:

a) Amorphous metallic glasses
b) Properties of metallic glasses
c) Applications of metallic glasses

Chapter

Biomaterials

25

OBJECTIVES

- To understand about biomaterials
- To study the development of various stages of biomaterials
- To classify biomaterials based on their properties
- To understand the properties of biomaterials and their applications

25.1 INTRODUCTION

Biomaterials are special materials which provide an intimate contact with living tissues when they are implanted into the body tissues or parts. Biomaterials are used for different biomedical applications in the field of medicine. They are used to repair or replace damaged or diseased body parts in the human/animal body. The different biomedical applications are surgical sutures and needles, catheters, orthopaedic hip replacements, vascular grafts, implantable pumps, cardiac pacemakers, etc. Even though newer materials like nuclear, magnetic, shape memory alloys, composites, etc., are being developed for advanced applications, biomaterials require important physico-chemical properties to be used for implant applications.

In recent years, some of the newer biomaterials have come up due to requirements in the medical field for different applications. In this chapter, the classification of biomaterials, their properties and applications are given in detail.

25. 2 BIOMECHANISM

The implants which are used for improvement of the quality of life produce a constant challenge to develop new biomaterials. The implant materials (biomaterials) react with tissues in different ways depending on the material type. Thus, the tissue attachment mechanism depends on the response of the tissue to the implant surface. In general, one can classify the mechanism as inert, bioresorbable and bioactive. The inert materials such as titanium, alumina, etc.,

are nearly chemically inert in the body and show minimum chemical interactions with the adjacent tissues. The bioresorbable materials such as tricalcium phosphate, polylacetic polyglycolic acid, polymers, etc., are designed to be replaced slowly by the tissues (bone). The third class of materials such as glass, ceramics and glass-ceramics which contain oxides of silicon, sodium, calcium and phosphorous form a chemical bond with living bone leading to the formation of a strong mechanical implant / bone bond. These materials are known as bioactive materials, since they bond to living bone.

Biomaterials are classified into three groups, namely, bioinert, bioactive and biodegradable materials based on their interactions. The different categories of biomaterials based on their interactions are shown in Fig.25.1.

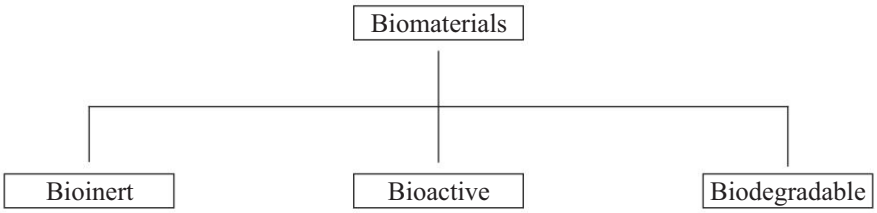


Fig. 25.1 *Biomaterials—classification*

25. 2. 1 Bioinert Materials

When a synthetic or natural biomaterial is placed in a human body as an implant, the interactions between the implant and the host tissue will be minimum, i.e., the interaction of implant with the host tissue is chemically nil. Therefore, the implant material behaves as an inert materials. On the other hand, when a foreign substance comes in contact with the organism, the system initiates its host defense mechanism involving an inflammatory response. Examples of bioinert materials are stainless steel, titanium, alumina and polythene. Bioinert materials are generally used for orthopaedic applications like fraction, fixation, wires and fixing screws.

25.2.2 Bioactive Materials

Whenever there is an interaction between the implant and the host tissue, the implant material is known as a bioactive material. A bioactive material is also defined as a material that elicits a specific biological response at the interface of the material, which results in the formation of a bond between the tissue and the material. When a biomaterial is implanted in to the body, the interaction is initiated by a time-dependent kinetic modification of the surface. Thus, it leads to ion-exchange reaction between the implant and body fluids and hence, a bioactive layer, i.e., an *apatite layer* is formed on the surface of the implant. The formation of an apatite layer enhances the tissue interactions. Examples of bioactive materials are hydroxyapatite (HAp), bioceramics, bioactive glass and glass-ceramics.

25.2.3 Bioresorbable or Biodegradable Materials

When a bioresorbable material is placed in a human body, it starts to dissolve slowly and hence, replaces the tissues. Thus, bioresorbable materials degrade or dissolve on implantation in the body. It is essential that the degradation rate of a bioresorbable material is equal to the regeneration rate of the host tissue. Bioresorbable materials are used to provide significant advances in drug delivery systems and medical implants. Examples of bioresorbable materials are tricalcium carbonate, gypsum, calcium oxide and polyacetic-polyglycolic acid polymers. The different categories of biomaterials and their interactions with host tissues along with applications are given in Table 25.1.

Table 25.1 *Biomaterials properties and their applications*

Biomaterials	*Examples*	*Definition*	*Applications*
Bio-inert	Stainless steel, titanium, alumina	Minimal interaction with surrounding tissues	Implant applications
Bioactive	Hydroxyapatite, Bioactive glass, Glass ceramics	Interact with surrounding hard(bone) and soft tissues	Bone filling, Dental filling applications
Bioresorbable	Polylactic–polyglycolic acid copolymers	Material slowly dissolved into body and replace the damaged tissues/ organs	Scaffold application drug delivery

25. 3 DEVELOPMENT OF BIOMATERIALS

Based on the development of biomaterials for different biomedical applications since the 1950s, one can categorise the biomaterials as first-, second- and third-generation biomaterials as shown in Fig. 25. 2. Biomaterials and their general classifications, namely, first, second and third generations are given in Table 25.2 along with the important properties of biomaterials and their applications.

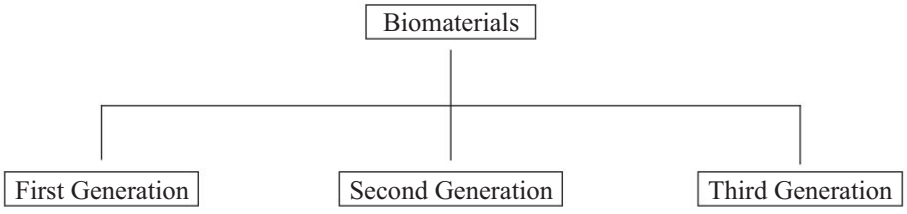


Fig. 25. 2 *Development of biomaterials*

Table 25.2 *Development of biomaterials and their applications*

S. No.	*Period*	*Category of biomaterials*	*Applications*
1	1950–1969	First Generation	• Fracture fixation • Bone and joint replacement • Dental implants • Surgical instruments
2	1970–2000	Second Generation	• Coatings for tissue ingrowth • Dental implants • Temporary bone space fillers • Percutaneous access devices • Artificial tendon and ligament • Biodegradable polymers and hydro gel
3	2000 to till date	Third Generation	**MEMS and NEMS devices** • Blood pressure sensors • DNA array systems • **Tissue engineering** • Tissue repair/regeneration • Cell encapsulation • Bio artificial skin and liver construction **Nanostructured biological systems** • Nanoparticles-based scaffolds Encapsulation on pharmaceuticals • Biological structure imitations at atomic and molecular scales.

25.3.1 First–Generation Biomaterials

Materials which were developed during the period of 1950s to 1960s mainly as implants to the human body for multiple disciplinary applications like orthopaedics, cardiovascular surgery, ophthalmology, and wound healing are known as first-generation biomaterials. Generally, materials like cellulose, acetates and polymers which are available in nature were used in the first-generation biomaterials. The important properties of the first–generation biomaterials to be considered are biocompatibility, bio-functionality and practicability. The foremost and the desired characteristic of a biomaterial is its biocompatibility. *Biocompatibility* is the ability of the surrounding tissues and the body as a whole to accept an artificial material when it is implanted in the body.

Similarly, *biofunctionality* is the second important property of biomaterials. It is the ability of the biomaterial to exhibit adequate physical and mechanical properties to augment or replace the body tissues. The practicability is the amenability of the biomaterial to machine, or to form, different shapes for ready usage at low cost. The first-generation biomaterials and their applications are given in Table 25.3.

Table 25.3 *First-generation biomaterials and their applications*

S. No.	*First-generation biomaterials*	*Applications*
1.	**Metals and Alloys**	
	• 316L stainless steel • Ti alloys, Co-Cr alloys • Gold alloys • Silver products • Platinum	• Joint prostheses • Artificial joints • Fracture fixation • Surgical instruments • Bone and joint replacement • Dental implants • Bone plates • Cardiovascular surgery • Dental and maxillofacial implants
2.	**Ceramics**	
	• Alumina • Zirconia • Calcium phosphates • Porcelain • Carbons	• Joint replacement • Dental implants • Bone replacement • Dental restorations • Heart valves • Percutaneous devices
3.	**Polymers**	
	• Polyethylene • Polypropylene • Polyamides • Polytetra fluoroethylenes, • Polyesters • Polyurethanes • Polymethylmethacryble (PMMA) • Silicones • Hydrogels	• Joint replacement • Vascular prostheses • Blood-contacting devices • Dental restorations • Intraocular lenses • Ophthalmology
4.	**Composites**	
	• BIS-GMA-quartz/silica filler • PMMA- glass fillers	• Dental restorations • Dental cements

25.3.4 Second-Generation Biomaterials

Biomaterials developed in the period between 1970s and 2000s are termed second-generation biomaterials. The basic functional properties of biomaterials, namely biocompatibility and biofunctionality, were retained while developing the second-generation biomaterials. However, biomaterials with bioinert tissue responses were replaced by the newly developed bioactive, biodegradable/bioresorbable materials. The important properties of second-generation biomaterials is their ability to be bioactive and biodegradable/ bioresorbable. The bioactive or biodegradable materials are widely used in clinical applications and there is a huge demand for these materials. The bioactive materials exhibit specific and controlled interactions with the surrounding tissue leading

to a chemical bond with them. On the other hand, the interactions of bio-inert materials lead to little or no tissue response. The second-generation biomaterials are widely used in orthopaedic and dentistry applications. The examples of second-generation biomaterials and their applications are given in Table 25.4.

Table 25.4 *Second generation biomaterials and its applications*

S. No.	*Second generation biomaterials*	*Applications*
1.	**Bioactive Materials**	
	• Bioactive glasses—silica and phosphate based glasses	• Bioactive fixation • To promote the chemical bonding between the implant and bone
2.	**Bioceramics**	
	• Calcium phosphate • Aluminium calcium phosphate • Coralline • Tricalcium phosphate • Zinc calcium phosphorous oxide • Zinc sulphate calcium phosphate • Ferric calcium phosphorous oxide • Bioactive glass ceramics • Ceravital	• Bone plates and screws • Femoral heads • Middle ear ossicles • Hip replacement prosthesis • For repairing and fusion of spinal and lumbo-sacral vertebrae • Replace and correct diseased bone and its defects
3.	**Biodegradable polymers**	
	• Polyglycolic acid • Polylactides • Polycaprolactone • Poly(ortho esters) • Polyanhydrides • Polyphosphazenes, etc.	• Soft and hard tissue repair-hydrolytic and enzymatic degradation • Orthopaedic implants • Artificial vascular grafts • Heart valve
4.	**Hydrogels**	
	• Poly(vinyl alcohol) • Polyacrylamide • Poly(N-vinyl pyrrolidone) • Poly(hydroxyethyl methacrylate) • Poly(ethylene oxide) • Poly(ethylene glycol) • Poly(ethylene glycol) monomethyl ether • Cellulose • Poly(hydroxyethyl methacrylate), etc.	• Soft contact lens material • Orthopaedic implants • Ophthalmic implants • Controlled drug systems • Biosensors

25.3.3 Third-Generation Biomaterials

The third-generation biomaterials open up the possibilities to cure complex problems and hence, improve the quality of human life through their clinical applications. The new generation biomaterials developed beyond the year 2000 to the present day are known as third-generation biomaterials. The third-generation biomaterials have been

developed with converged properties. In third-generation biomaterials, the important characteristics of second–generation biomaterials, namely, bioactivity or biodegradable properties are converged and retained. The advances in the field of tissue engineering, microfabrication and nanofabrication are taken into consideration to develop third-generation biomaterials. The main objective of the development of third-generation biomaterials is to regenerate the host tissue rather than to replace the damaged or diseased tissues. In third-generation biomaterials, materials like smart biomaterials and implants are developed using micro and nano fabrication techniques. The examples of third-generation biomaterials and their applications are given Table 25.5.

Table 25.5 *Third-generation biomaterials and their applications*

S. No.	*Third generation biomaterials*	*Applications*
1.	**Tissue engineering**	• Tissue repair/regeneration • Cell encapsulation • Tissue scaffold for cell growth and organisation • Bone morphogenic protein infusion into orthopaedic implants • Artificial skin construction • Bioartificial liver construction
2.	**Micro electro mechanical systems (MEMS)**	• MEMS accelerometer • Blood pressure sensors • Blood chemistry analysis systems • DNA array systems • Telemetry monitoring systems • Implantable drug delivery
3.	**Nanoelectroemechanical systems (NEMS)**	• Nanoparticle based scaffolds • Nanostructure encapsulation on pharmaceuticals • Biological structure imitations • Self-assembled structures at atomic and molecular scales
4.	**Nanobioactive glass** • Nanometric BG 45S5 • BG nanopowders – Coating	• Better antimicrobial effect • Improved corrosion resistance and bioactivity

Microtechnology Microfabrication is one of the new techniques used for the creation of new third generation biomaterials. The biomaterials generated from this technique are smaller in size with more active and dynamic implants and enhanced biocompatibility for a range of biomedical applications. The dimension of the devices created from this method is in the range of micrometres to millimetres. The method is based on the top-down approach. The small structures or devices are created on the surface of the bulk material substrate using different techniques for micromachining such as photolithography, etching, etc. Then, the micro electromechanical systems

(MEMS) are used to fabricate and to integrate the electrical components and mechanical components on a small device.

Nanotechnology Nanotechnology is an enabling technology which deals with atomic and molecular levels in the length scale of 1–100 nm range. It is a useful technology to create structures, devices and systems with novel properties and functions employing atoms/molecules at the nanoscale length. The novel properties of the materials at their nanoscale length are totally different from that of the same in its bulk form. The comparison of the structures of the nanostructured materials with that of biological structures elucidates that most of the biological structures are at the nanoscale level. A comparison of the size of nanostructured materials and the biological tissues are given in Table 25.6.

Table 25.6 *Comparison of the size of nanostructured materials and biological systems*

Nano structured Materials		*Biological structures*	
Materials	Size nm	Element	Size nm
Nanoparticles	1–100	Atom	0.1
Fullerene (C60)	1	DNA-	
		• Width	2
		• Thickness	0.34
		• Pitch length	3.4
Quantum dot (CdSe)	8		
		Protein	2.2–150
Dendrimer	10	Virus	60–140
CNT	2-10	Bacteria	1000–10,000
		White blood cell	10,000
		Nucleotides	0.81 – 0.95
		Amino acids	0.42 – 0.67

25.4 CLASSIFICATION OF BIOMATERIALS

Biomaterials are classified based on the developments of three different generations, origin (i.e., natural or synthetic), compositions (i.e., polymers, metals and alloys and ceramics) and the interactions and tissue response (bio-inert, bioactive, bioresorbable). The significant development in biomaterials includes the recent booming technology, namely, nanotechnology. The various clauses of biomaterials, their properties and applications are highlighted in the following sections.

25.4.1 Metals and Alloys

From ancient times, metals have been used for orthopedic applications. Pure metals such as gold, silver and copper were used for different medical applications. Later, the usage of the above metals was restricted due to the poor surgical conditions.

The McGraw·Hill Companies

In view of the requirements of suitable materials for medical applications, alloys such as 316L stainless steel, Co-Cr alloys and Ti-6Al-4V alloys have been developed for orthopedic applications. These materials fix to a bone by mechanical interlocking and they do not form any chemical bonds with the living bone. At a later stage, the above materials are coated with bioactive glasses and glass-ceramics to make them bioactive with the natural bone, since the implants (interlocking) will loosen after a long period of time. The degradation of the coating layer may take place due to the peel off from substrates, low strength, etc., resulting in the requirements of high-mechanical-strength materials that can bond directly to the natural bone.

The different types of metals and alloys used as biomaterials for medical applications are given in Table 25.3.

25.4.2 Polymers

The common polymers that are used in biomedical applications are linear chains of repeating subunits. The property of this polymer depends on the chemical structure and macromolecular organisations of the polymer chains. The change in the properties of the polymers plays a vital role in different medical applications. A class of polymers are known as *biodegradable* or *resorbable*. When it is implanted in a human body, the polymer is gradually absorbed by the human body without permanently retaining any trace of residue in the implantation site and hence, they regenerate tissues through interaction of their degradation products. The following are some of the polymers which are well studied for biocompatibility, namely, polylactide acid, polyglycolic acid and their copolymers.

The common applications of these polymers are implantable devices, catheters and tubing, vascular grafts, injectable drug delivery and imaging systems. The different types of polymers and applications are given in Table 25.3.

25.4.3 Hydrogels

Hydrogels are basically cross-linked polymers with hydrophilic property. They can either absorb water or any other biological fluids due to their hydrophilic properties. The important characteristic properties like swelling, diffusive and surface properties of hydrogel polymers make them different from other biopolymers. The hydrogels are used in a variety of clinical applications since the 1950s. The notable biomedical applications of hydrogels are implants, drug delivery systems and biosensors. In view of the excellent properties of hydrogels, they are categorised as second-generation biomaterials.

25.4.4 Composites

The composite biomaterials are hybrid products of two or more distinct constituent materials possessing the desired physical and mechanical properties to meet specific clinical requirements. The composite biomaterials are significantly different from homogeneous biomaterials in terms of their physical and mechanical properties. The biocomposite materials are new biomaterials which offer a great promise to improve the quality of human life. Biocomposite materials offer many advantages in comparison to homogeneous materials.

One can achieve the required strength by reinforcement of bioactive glasses into the polymer matrix. The resultant materials are known as *biocomposite materials*. The biocomposite materials with improved strength find wide applications in industries due to the close agreement of the material stiffness with that of bones. The essential requirement of biocomposites is that each constituent material should be biocompatible, and the interface between constituents should not be degraded in the event of biological environments. Examples of applications of biocomposites are dental filling, bone cement and orthopaedic implants, which are given in Table 25.3.

25.4.5 Ceramics

Ceramics are polycrystalline refractory materials including organic metal oxides, silicates, carbides, hydrides, sulphides, selenides and nonmetallic elements. Glasses are often considered to be a subset of ceramics. Glasses are amorphous and have only a short-range crystalline order, which makes them brittle. In recent years, high-technology ceramic materials, namely, advanced ceramics with improved compressive strength and bio-inertness find potential applications as implant biomaterials. Bioceramics are classified into three major categories based on biological environments, namely, nonabsorbable or bioinert, bioactive or surface reactive, and biodegradable or resorbable bioceramics. These bioceramics are developed in the first, second and third generation of biomaterials. The classification of bioceramics and its applications are given in Table 25.5.

(a) Bioinert or Nonabsorbable Ceramics Examples of non-absorbable or bio-inert bioceramics are alumina, zirconia and silicon nitrides which belong to the first generation of bioceramics. These materials are able to maintain their mechanical and physical properties without any alteration, even in biological environments. These bioinert materials are used for typical structural support implants such as bone plates and bone screws. Biologically inert materials can survive for a long period in a highly corrosive environment that the human body provides. Examples of bioinert ceramics and their applications are given in Table 25.5.

(b) Bioactive Glasses and Glass Ceramics Bioactive glasses and glass ceramics play an important role in biomedical applications as bone-repairing materials. They include a broad range of inorganic/nonmetallic compositions. These materials are used for their ability to provoke responses on surrounding bones and tissues after implantation. These materials are known as *surface reactive ceramics*, since they form a strong chemical bond with adjacent living tissues after implantation. All the bioactive glasses and glass ceramics are bonded to living bone through the formation of an apatite layer which is formed on the surface of the bioactive glasses. These bioactive glasses and glass ceramics are used both for orthopaedic and dental applications and are given in Table 25.5.

(i) Nanobioactive Glass Among biomaterials, Nanobioactive Glasses (NBG) have become essential biomaterials in view of their bone-bonding ability and load-bearing applications. The transition from bulk bioactive glass to nanobioactive glass is used to increase the rate of formation of hydroxyapatite layer and its thickness. This is possibly due to the higher surface-to-volume ratio of NBG when compared to bulk glass. The increased higher surface area results in higher rate of the interactions between the nanoparticles and simulated body fluid.

The McGraw·Hill Companies

(ii) Hydroxyapatite The bone serves as a reservoir for minerals, particularly calcium and phosphate. Hydroxyapatite (HAp) is a class of calcium-phosphate-based bioceramics. HAp plays an excellent role in biomedical applications due to its close chemical and physical resemblance to mineral components of bone tissue, enamel and dentin. According to calcium phosphate ratio, it is classified into different phases, namely, tricalcium phosphates (TCP), tetracalcium phosphates (TTCP), amorphous calcium phosphate (ACP), brushite, monetite, hydroxyapatite (HAp) and octa-calcium phosphate (OCP). The chemical component of HAp is $Ca_{10}(PO_4)_6(OH)_2$ with a stoichiometric Ca/P ratio of 1.67. HAp is derived either from natural or from synthetic resources and is one of the most important bioactive materials, since it forms a strong chemical bond with the host bone tissue.

(iii) Ceravital Ceravital glass is one of the popular glass ceramics used as a biomaterial. The chemical composition of ceravital glass is the same as that of bioglass and contains equal amount of SiO_2. However, elements such as Al_2O_3, TiO_2 and Ta_2O_5 are added in order to control the dissolution rate and to improve its bioactivity. Some of the important properties of ceravital glasses are very low thermal expansion coefficient, high tensile strength, and high scratching and abrasive resistances. These properties make ceravital glasses superior to glass and ceramics.

(c) Bioresorbable or Biodegradable Ceramics Resorbable ceramics are the ceramics which degrade upon implantation in the host. Some of the examples of bioceramics are aluminum calcium phosphate, hydroxyapatite and tricalcium phosphate which from third-generation biomaterials. The biodegradable bioceramics are different from bioactive bioceramics due to their high porous structure high solubility. They are highly degradable by the surrounding tissues due their high solubility. The resorbable material is replaced by endogenous tissues. The rate of degradation varies from material to material. The resorbability of bioceramics depend on the content of calcium phosphate present in then.

25.5 PROCESSING AND PROPERTIES

Biomaterials such as metals and alloys, glasses and glass ceramics, polymers and composites are prepared using the conventional techniques. In order to use biomaterials for different biomedical applications, the study of biocompatibility and mechanical properties and their interaction with the natural tissues are more important. The mechanical properties of the biomaterials such as fracture strength, compressive strength, fracture toughness, elastic moduli, etc., are essentially required. The above mechanical properties of the biomaterials can be measured using the conventional testing and measuring methods. The evaluation of the mechanical properties and their relations with their natural tissues will help to study the biocompatibility of the biomaterials for proper orthopaedic and dental-restoration applications.

25.6 APPLICATIONS

The following are some of the applications of biomaterials in the field of medicine:

Metals and Alloys 316L stainless steel, Co-Cr and Ti-6Al-4V alloys are used for orthopaedic applications. Co-Cr-Mo alloys have been used as stem and heads of

implanted hip endoprosthesis, and Co-Ni-Cr-Mo alloy has been used for hip joint. Ni-Ti shape memory alloy has been used in dental arch wires, microsurgical instruments, blood clot filter, guide wires, etc.

Bioactive Glasses and Glass Ceramics In view of the high mechanical strength and biocompatibility of the bioactive glasses and glass ceramics with living tissues, they find wide applications in orthopaedic and dental restoration applications. Bioactive glasses made up of main components of oxides of silicon, calcium, sodium, phosphorous and calcium fluoride have been used for implant applications, orthopaedic devices such as knee and hip joint replacements, spiral implants, and bone fixtures.

Dental implants Dental restorations	Replacement of teeth / root system Crown, craps, fillings
Heat treated SiO_2-Al_2O_3-Na_2O-K_2O-CaO-P_2O_5 with B_2O_3, Li_2O, CeO_2, TiO_2, ZrO_2 and F^-	Restorative dental applications
MgO-CaO-SiO_2-P_2O_5 glass ceramics	Load bearing prostheses, tight chemical bonds with living bone
Calcium phosphate glasses with high CaO/P_2O_5 ratio modified by additives such as Na_2O, MgO, TiO_2, etc.	Implants for bone replacement / regeneration and drug delivery carriers

Polymers

Polymethylmethacrylate (PMMA)	As bone cement
Polytetrafluroethylene (PTFE)	Artificial organs responsible for blood circulation
Polyurethene	Blood / device interfaces
Polyesters	Vessels prostheses such as blood vessels
Nylon	Blood vessels, joints
Polyvinylchloride (PVC)	Blood vessels, heart components

Biopolymers are used for a variety of surgical applications such as blood vessel prostheses, tissue adhesives, heart valves, lenses and sutures.

Biocomposites The following are some of the applications of biocomposites.
Carbon fibre reinforced hip replacement

Oxirane / Polyol / dental composites—coupling of dental filler materials and matrix resin

Bioceramics The following are some of the applications of bioceramics:

Tricalcium phosphate—bone repair

Al_2O_3 – SiO_2—metal oxides – femoral head

Apatite ceramics—synthetic bone

Porous ceramics—mitral valve prostheses

Keypoints to Remember

- Biomaterials are materials which provide intimate contacts with living tissues when they are implanted into the body tissues or plants.
- Biomaterials are classified as bioinert, bioactive and biodegradable.
- When there is no interaction between the implant and host tissue, the implant material is known as a bio inert material.
- When there is an interaction between the implant and host tissue, the implant material is known as a bio active material.
- The material which start to dissolve slowly and hence replace the tissues are known as bioresorbable or biodegradable materials.
- There are three different generations of biomaterials, namely, first, second and third generations.
- Important properties of first-generation biomaterials are biocompatibility, biofunctionality and practicability.
- First-generation biomaterials are developed during the period of 1950–1969.
- The ability of the surrounding tissues and the body as a whole to accept an artificial implant material is known as biocompatabilty.
- Materials developed during the period of 1970–2000 are termed second-generation biomaterials.
- Materials developed beyond the year of 2000 are known as third-generation biomaterials.
- Biomaterials are classified based on origin, compositions and interaction, and tissue response.
- Hydrogels are cross-linked polymers with hydrophilic property.

Objectives Questions

25.1 _____ materials have minimal interactions with surrounding tissues.
25.2 _____ materials have interaction with surrounding bone and soft tissues.
25.3 Materials which slowly dissolved into body and replace the damaged tissues/organs are called _____materials.
25.4 First-generation biomaterials were developed during _____to _____.
25.5 Second-generation biomaterials were developed during _____to _____.
25.6 Third-generation biomaterials were developed beyond _____year.
25.7 Examples for biomaterials and alloys are_____, _____ and _____.

The McGraw·Hill Companies

25.8 Examples for biopolymers are _____, _____, _____ and _____.
25.9 _____, _____ and _____ are examples for hydrogels.
25.10 _____, _____ and _____ are examples for bio ceramics
25.11 _____ and _____ based glasses are examples for bioactive glass.

KEY

25.1 Bioinert
25.2 Bioactive
25.3 Bioresorbable or biodegradable
25.4 1950, 1969
25.5 1970, 2000
25.6 2000
25.7 316L SS, Co-Cr and Ti-6Al-4V
25.8 PMMA, PTFE, PVC and Nylon
25.9 Cellulose, Polyacrylamide and PVA phosphate,
25.10 Calcium Phosphate, Tricalcium and Ferric calcium
25.11 Silica and Phosphate

Short Questions

25.1 What are biomaterials?
25.2 What is meant by biomechanism?
25.3 What is meant by bioactivity?
25.4 What are the classifications of biomaterials?
25.5 Define bioinert materials.
25.6 Explain bioactive materials.
25.7 What is meant by biodegradable materials?
25.8 Differentiate bioactive and bio-degradable biomaterials.
25.9 Explain the different stages of development of biomaterials.
25.10 What is meant by bio-functionality?
25.11 What is practicability?
25.12 Explain nanobioactive glasses.
25.13 What is meant by microelectro mechanical system?
25.14 How are biomaterials classified based on the development of three different generations?

Descriptive Questions

25.1 What are biomaterials? Explain the different types of biomaterials and their applications in medical field.
25.2 Write notes on the following:
(a) Bioactive glass
(b) Bioceramics
(c) Biocomposites
(d) Biometals and alloys

The McGraw·Hill Companies

25.3 Write notes on the following:
(a) Bioinert
(b) Bioactive
(c) Bioresorbable

25.4 Write an essay about the first, second and third generations of biomaterials.

25.5 Write an essay about the classification of biomaterials and their applications.

Chapter

26

Nanomaterial Synthesis

OBJECTIVES

- To explore the knowledge on the preparation, properties and their applications of nanomaterials
- To understand different experimental methods such as physical, chemical, biological and hybrid methods used to prepare the nanomaterials and their limitations
- To discuss the size-dependent property of nanomaterials
- To study the influence of nanotechnology on environment
- To understand the applications of nanomaterials in different areas

26.1 INTRODUCTION

In recent years, nanoscience and technology has emerged as one of the most important and exciting forefront areas of interest in almost all fields of science and technology. This technology provides the path for many breakthrough changes in the near future in many areas of advanced technological applications. The prefix word *nano* means *dwarf* in Greek language. It can be used as a prefix to any unit like a second or a metre and is equal to a billionth of that unit (1 nm = 1×10^{-9} m). Nanoscience and technology is an interdisciplinary area of research and development. Nanotechnology means the construction of various structures of matter having dimensions of the order of a billionth of a metre for useful applications. The nanostructured materials are materials where the dimensions and tolerances are in the range of 0.1 to 100 nm (i.e., from the size of the atom to the wavelength of the light).

There are many definitions for the terms nanoscience and nanotechnology in literature. Prof. Richard Feynman, a physics Nobel Laureate, in his lecture in late 1959, stated that there is *plenty of room at the bottom* and made initiative in this innovative field before the existence of the word nanotechnology. Feynman got this idea from the biological systems. For example, the size of proteins is 4 to 50 nm, the diameter of DNA molecule is 2 nm. The nanostructured materials may be metals, alloys, inter metallics, ceramics or biological materials. Generally, the nanostructured materials exhibit greatly altered

The McGraw·Hill Companies

properties like physical, chemical and mechanical, when compared with their normal bulk materials having the same chemical compositions. To quote the exotic properties of the nanostructured materials, the nano gold particles exhibit 100% coefficient of absorption of light, while the bulk gold particles glittering effect is only due to their reflection properties.

In the following sections, let us discuss briefly the different synthesis methods of nanostructured materials along with their potential applications.

26.2 SYNTHESIS OF NANOSTRUCTURED MATERIALS

The nanostructured materials with ultra fine grain size are in the order of 1 to 100 nm and can be produced with different dimensionalities. The different forms of nano materials are zero, one, two and three dimensions. The above classifications are based on the crystallite grain size and shape which are given in Table 26.1.

Table 26.1 *Classification of nanomaterials*

S. No.	*Types*	*Crystallite/grain size/shape*
1	Zero dimension	Clusters or powders (Quantum dots)
2	One dimension	Multilayers
3	Two dimension	Ultrafine grained over layers or buried layers
4	Three dimension	Equiaxed nanometer sized grains

As we know, the applications of nanostructured materials are diverse and hence, one has to select an approximate method for synthesis of nanomaterials to suit their requirements. However, in nanomaterials, the particles in the ultra fine structure are in the order of nano scale lengths. Therefore, the nano particles can be synthesised in two broad ways namely, top-down and bottom-up approaches.

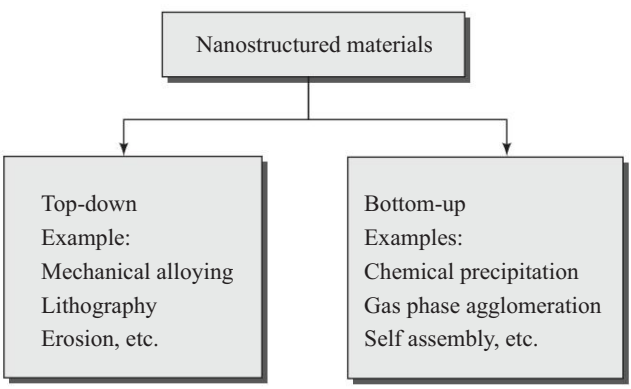


Fig. 26.1 *Synthesis of nano structural materials*

In the bottom-up approach, the nanostructured materials are synthesised by assembling the atoms or molecules together to form the nano materials. On the other

The McGraw·Hill Companies

hand, in top-down method, the bulk solids are dis-assembled (broken or dissociated) into finer pieces until the particles are in the order of nanometre. The schematic representation of the synthesis and building of nanostructured materials are shown in Figs 26.1 and 26.2 respectively.

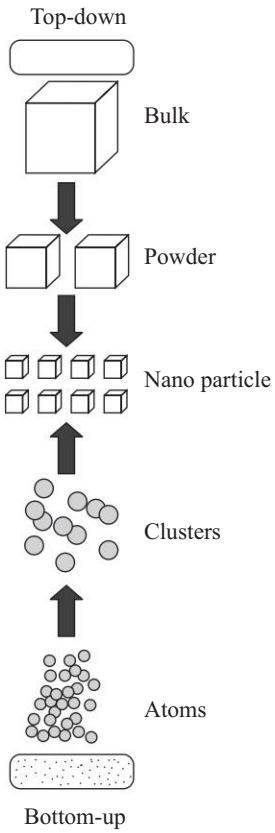


Fig. 26.2 *Schematic representation of top-down and bottom-up nanostructural materials synthesis*

The top-down and bottom-up method consists of different techniques for the synthesis of nanostructured material. For example, mechanical grinding, lithography and erosion are some of the techniques used under top-down approach for the production of nanomaterials. The techniques like chemical precipitations, gas phase agglomeration and self-assembly are used to obtain the nanomaterials under bottom-up approach. However, one has to select a right technique for the synthesis of nanomaterials for specific applications.

It is essential to produce an optimised novel nanostructured material for potential applications. In view of the requirement of optimised nanomaterials, in addition to the existing techniques like physical, chemical and biological techniques, hybrid techniques are also developed. The various techniques available to synthesis the nanomaterials are classified as shown in Table 26.2.

The McGraw·Hill Companies

Table 26.2 *Classification of synthesis of remove the nano materials*

Techniques	*Method*
Physical method	Mechanical
	High energy ball milling
	Lithography
	Machining process
	Vapour
	Physical vapour deposition
	Chemical vapour deposition
Chemical method	Colloids
	Sol gel
	Inverse micelles
Biological	Using biomembranes
	DNA
	Enzymes
Hybrid	Electrochemical
	Chemical vapour deposition
	Micro emulsion

26.3 TOP-DOWN APPROACH-NANOMATERIALS SYNTHESIS

One of the main problems for the researchers is the production of large quantity of nanomaterials for diversified applications. Some of the methods which are used in top-down method can be easily scaled up to produce large quantity of nanomaterials. In the following sections, techniques like mechanical grinding, lithography and machining are explained briefly for the production of nanomaterials.

26.3.1 Mechanical Grinding

The mechanical grinding method is a typical example of top-down method of synthesis of nanomaterials. The mechanical grinding method is very popular, simple, inexpensive and extremely scaleable method to synthesis all classes of nanomaterials. This method is used to produce nanocrystalline or amorphous materials.

The mechanical attrition mechanism is used to obtain nanocrystalline structures either from single phase powders or dissimilar powders or amorphous materials. The schematic representation of the principle of mechanical milling is planetary ball or a tumbler mill as shown in Fig. 26.3.

One can use either refractory balls or steel balls or plastic balls depending on the materials to be synthesised. When the balls rotate at a particular rpm, the necessary energy is transferred to the powder which in turn reduces the powder of coarse grain seized structure to ultra fine nano range particles. The nano particles are produced due to the shear action as shown in Fig. 26.3. Thus, it is clear that the energy transferred to the powder from balls depend on many factors such as rotational speed of the balls, size of the balls, number of the balls, milling time, the ratio of the ball to powder mass and the milling medium/atmosphere. One can also use cryogenic liquids to increase the brittleness of the powders. During the milling process, one has to take necessary steps to prevent oxidation.

The McGraw·Hill Companies

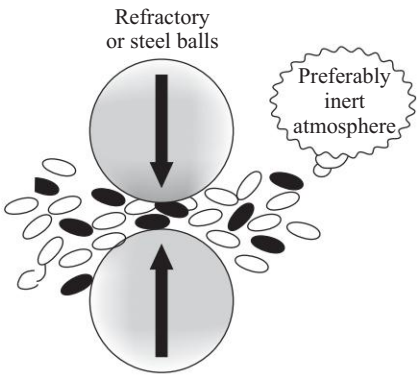


Fig. 26.3 *Schematic representation of the principle of mechanical milling*

The selection of materials of the balls plays an important role. Generally, a harder material will be selected to synthesize the softer materials. For example, α-alumina and zirconia are used widely as a ball for synthesizing the nanomaterials in view of their high grinding resistance values. The relationship between the hardness and grinding resistance is shown in Fig. 26.4.

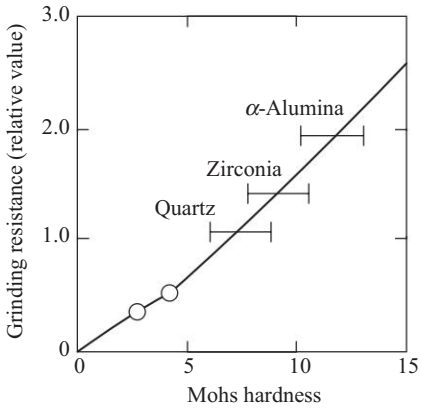


Fig.26.4 *The relationship between the micro hardness and grinding resistance*

The main advantage of this method is the scaling upto tonnage quantity of materials for wider applications. The disadvantage of this method is the contaminations of milling media and or atmosphere and is restrictive for the production of non-metal oxides. The non-metal oxides require an inert medium for milling and vacuum or glove box to handle the powder particles.

26.3.2 Nano Lithography

Lithography means carving or writing on a stone. It is one of the techniques used to pattern the flat surface by removing some part of it or to organize some material on a suitable substrate. It involves the transfer of design of patterns on a semiconductor (or) suitable substrate at suitable high speed radiation.

Lithography is one of the key technologies in the field of advanced semiconductor manufacturing. It plays an active role in the technology sectors like integrated circuit (IC) manufacturing or very large scale integration (VLSI), flat panel displays, optoelectric components and advanced electronic packaging. The progress of lithographic systems due to the advancement in technologies lead to tremendous potential in microelectronics industry. One can produce the patterns even at the scale of atomic distances due to advancement of the lithographic systems. Industries like circuit manufacturing requires uniformity, reproducibility, line width control and overlay accuracy of their products. The above requirements can be achieved employing the nanolithography techniques. Nanolithography is basically a top-down method. Thus, using the nanolithography, the atoms or molecules can be removed from the bulk materials and hence, the nanostructured patterns are obtained. The modern electronic devices like cell phones and ATM require miniaturised components like computers, resistors, etc. These components require very less power for operation, small amounts of materials and low cost. On the other hand, their performance and operations are faster.

The miniaturisation of structures such as, FET, surface gated quantum devices, quantum dots, wires, grating zone plates, and mask making can be achieved employing the nanolithography technique upto deep submicron scale with enhanced performance.

(a) Types of Lithography The important lithographic techniques which are commonly used are optical lithography, X-ray lithography, electron beam lithography, etc. The lithographic techniques are classified based on the types of radiation used for carving the material or transferring the pattern.

The electromagnetic radiation or a particle is used to expose the material at the selected regions of resist. The material which is sensitive to radiation is known as *resist*. The selection of the desired portion in the material is made by mask. The material is transparent in some region while it is opaque in other regions due to its exposure to radiation. Thus, the radiation-exposed resist becomes weaker or stronger compared to the unexposed resist. The desired pattern is obtained by resolving the exposed or unexposed material using a suitable chemical or plasma process. The above process can be repeated to obtain the desired pattern on the materials.

A schematic representation of photolithography which is used to obtain the pattern on a semiconductor surface is shown in Fig. 26.5. A thin film of a metal like chromium is coated on a glass or silicon substrate. A positive or negative photoresist material like polymer is coated above the metal surface. When the positive photoresist material is exposed to radiation, it degrades the materials or some chemical bonds are broken. On the other hand, when the negative resist material is exposed, it hardens the materials. A mask is placed over the resist coated surface. Then, the UV radiation is exposed to the mask, resulting in weaker or stronger regions on the resist, depending on the opaque and transparent regions of the mask. The image is developed using a suitable chemical (i.e., developer). The remaining unexposed material is recovered by using an appropriate chemical treatment. The etched pattern on the materials is obtained as shown in Fig. 26.5.

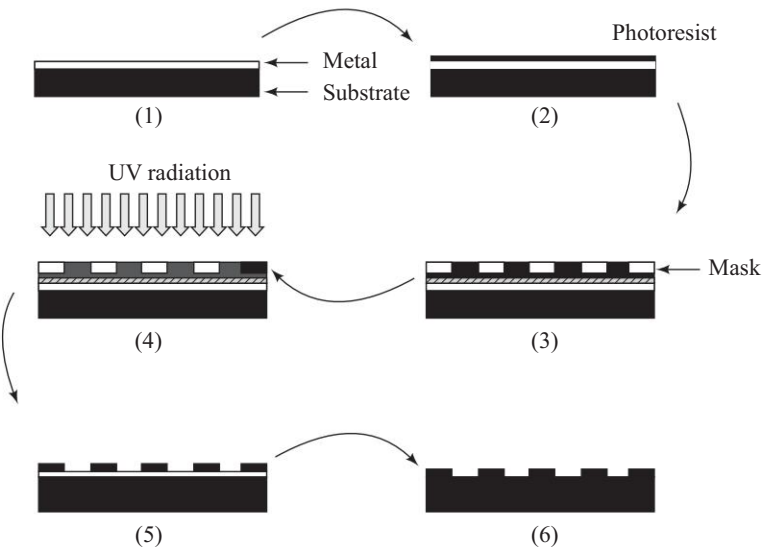


Fig. 26.5 *Photolithography process*

The Lithography is classified into the following types based on the type of radiation used:

(i) Lithography using photons like UV light, laser and X-rays
(ii) Lithography using particle beams
  a. Electron beam lithography
  b. Ion beam lithography
  c. Neutral beam lithography

In addition to the above techniques, the probe or fine tip of the scanning tunneling microscope (STM) and atomic force microscope (AFM) are used to write or scratch on the materials. These techniques are quite inexpensive and used to achieve the pattern like lab-on-chip. The techniques developed using the above concept is scanning probe lithography (SPL) and soft lithography (SL).

Let us discuss briefly the lithography based on photons, particles and probes in the following sections:

(b) Lithography Using Photons The electromagnetic radiation like visible, ultraviolet (UV-visible) or X-rays are used to obtain lithography. During the process of obtaining the pattern, the glass lenses and masks are used to obtain a pattern in the visible range of the order of 700 nm to 400 nm. The fused silica or calcium fluoride lenses are used in case of UV range. There are three methods to obtain the pattern using electromagnetic radiations, namely, proximity, constant and projection method. The schematic representation of the three methods are shown in Fig. 26.6.

In case of proximity method, the mask is held in close contact to the photoresist coated metal substrate, while the mask is in contact with the photoresist in case of contact method. A parallel beam of UV radiation falls on the mask in both the proximity and contact method. The radiation is transmitted in some regions while it is blocked in the opaque region. The obtained resolution of the pattern is more in contact

method than proximity method due to the close contact of the photo resist and mask. However, the mask gets damaged more in the contact method. In order to reduce the damage of the mask, the projection method is used. In this method, a focused beam is scanned through the mask as shown in Fig. 26.6. The above method also gives a better resolution than contact method.

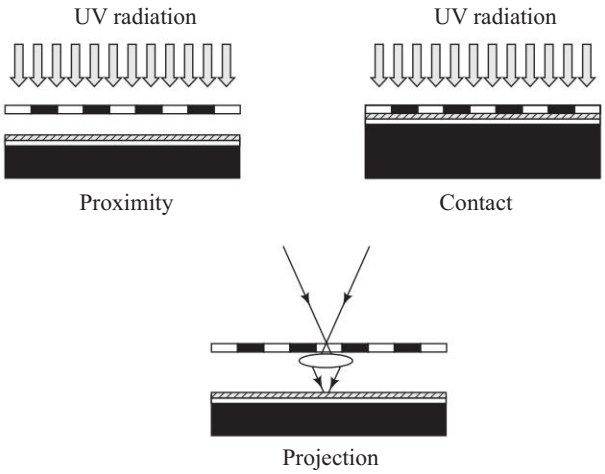


Fig 26.6 *Lithography using photons*

c) Lithography using UV Light and Laser Beam A monochromatic beam of light from mercury lamp in the visible UV range (436 nm) is used to obtain the pattern. The laser beam from KrF (248 nm) or ArF(193 nm) is used to obtain the pattern respectively with a size of 150 nm. It is clear that one cannot obtain the pattern below 100 nm using the photon as source.

(d) Lithography using X-rays X-rays are used to obtain the pattern below 100 nm but with a suitable mask for X-ray lithography. The main problem one should consider in X-ray lithography is the absorption of X-rays. Therefore, the metal mask should be fabricated with thin and thick which helps higher rate of transmission at thin and absorption at thicker regions. The gold masks are used in X-lithography.

(e) Lithography using Particle Beams We know according to de-Broglie, the wavelength (λ) of the particles depends on the mass (m) and velocity (v) as

$$\lambda = h/\, mv \tag{26.1}$$

where h is Planck's constant.

According to this principle, one can use atoms or ions or electrons to reduce the wavelength associated with the particle as small as 0.1nm. Therefore, larger particle mass and larger velocity increases the resolution. The resolution of the pattern depends on the interaction of the incident particle with resist material. In view of easy generation, acceleration and focusing, electrons are used in lithography.

(f) Electron Beam Lithography High energetic electrons are made to be incident on the positive or negative photoresist. The electron beam lithography is operated in two modes, known as *vector scan* and *raster scan*. During vector scan, the electron

The McGraw·Hill Companies

beam is made to write on a specified region and during the scanning of other region, the electron beam is put off. In raster scanning, the material is scanned continuously line by line, but the positions of the sample are moved at right angles to the beam according to the pattern to be obtained. The electron beam lithography is slower than other optical lithographic techniques.

Ion-beam lithography helps to obtain very small patterns. The resist materials are sensitive to ion beams rather than to electron beams and exhibit low scattering in the resist. Commonly used ions are He^+, Ga^+ etc., in ion beam lithography. The schematic representation of the electron beam lithography is shown in Fig. 26.7.

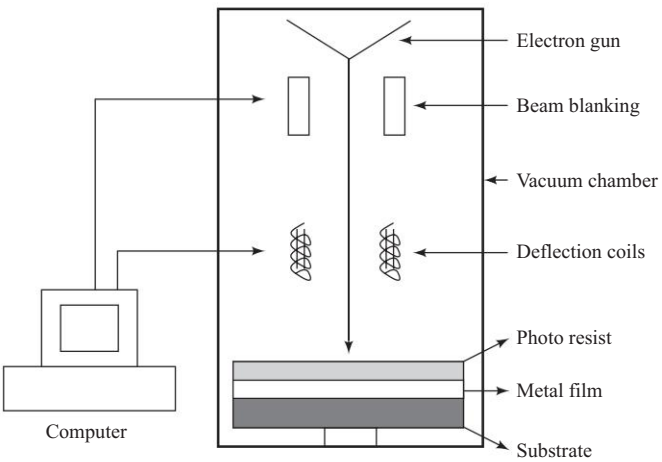


Fig. 26.7 *Electron beam lithography*

(g) Neutral Beam Lithography In particle beam lithography, neutral atoms are used to obtain patterns of the images. Here, the use of masks is essential, since the neutral atoms like argon or cesium are made incident on the material through the mask. Deposition of neutral atoms is possible over the material in this technique. The following are the applications of particle beam lithography:

(i) A Very Large Scale Integration (VLSI) of circuits on a small piece of semiconductor can be obtained using nanolithography.
(ii) Using nanolithography technology, a wide variety of electronic devices such as integrated circuits, photonic integrated circuits, etc., in smaller size and increased performance can be prepared.
(iii) The development of broadband communication mainly depends on the advanced lithography.
(iv) In nanolithography, the images can be patterned at the scale of 100 nm.

26.3.3 Scanning Probe Lithography

The development of scanning tunneling microscope (STM) and atomic force (AFM) microscope leads to a new era in the history of lithography. The new branch of lithography is known as scanning probe lithography (SPL). In SPL, the probe (or) tip is used to obtain the pattern at the nano scale level. The different methods used in SPL to obtain the pattern are, namely, mechanical scratching or movement,

optical, thermo-mechanical and electrical methods. Among the different methods, the mechanical method is discussed in detail.

The mechanical method includes different modes like scratching, pick-up and pick-down or dip-pen lithography. The schematic representation of the mechanical scratching method is shown in Fig. 26.8. The AFM tip or STM tip is used to scratch the surface of the bulk materials or thin films to obtain pits or lines. The scratched materials are piled up around the indented region. The dimension of the pits is as small as 30 nm in diameter and 10 nm in depth.

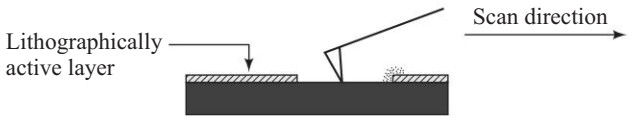


Fig. 26.8 *Mechanical scratching – Microscope tip*

In the pick-up and pick-down method, the AFM/ STM tip is moved close to the atoms which are organised on the surface of the metal substrate. The process of the method is shown schematically in Fig. 26.8. The AFM/ STM tip picks up the atoms which are having loosely absorbed surfaces. Later, the picked up atom is arranged in a desired pattern by moving the AFM/ STM tip as shown in Fig. 26.9. Using this method, attempts have been made to write any words (or) letters on the thin film or bulk materials at the nanoscale length. The range of patterns (i.e., characters or words) produced by this method ranges from 30–50 nm height and 10 to 60 nm width.

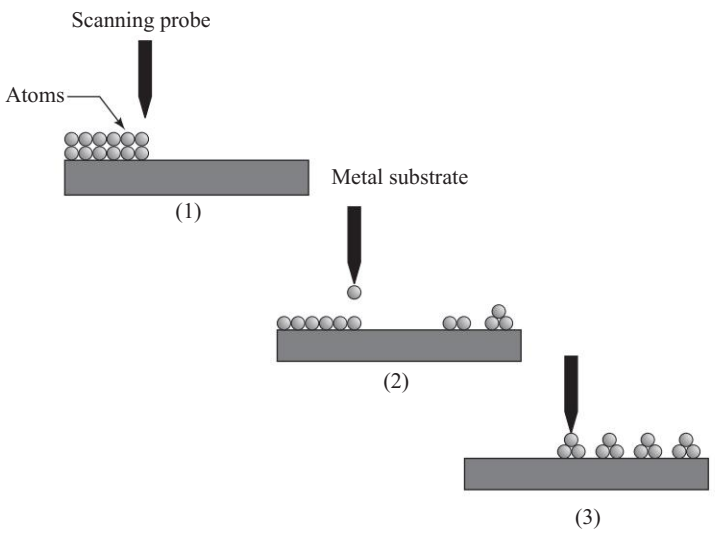


Fig. 26.9 *Pick-up and pick-down-Lithography*

Similar to pick-up and pick-down method, an another method used to obtain lithography is known as *dip-pen lithography,* as shown in Fig. 26.10. It is similar in nature as that writing to write in a paper by ink. In this method, the AFM tip is used as a pen while the molecules are used as ink. The desired molecules are picked up by the AFM tip from the source of molecules and then transported to the required place on the substrate. One can write the letters up to a line thickness of 60 nm with very

high clarity. The unique feature of this method is the overwriting and erasing, which cannot be done by any other methods.

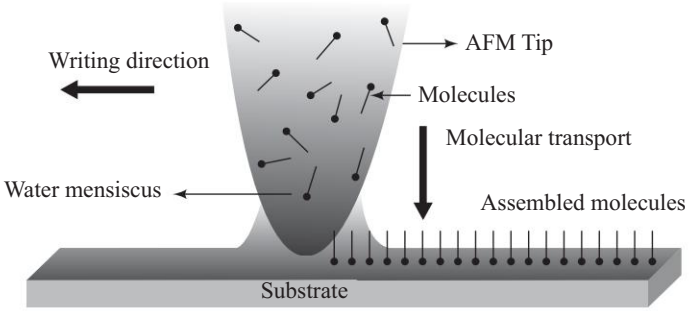


Fig. 26.10 *Dip-pin lithography*

26.4 BOTTOM UP PROCESS—SYNTHESIS OF NANOPARTICLES

In recent years, many techniques have been developed for the synthesis and formation of nanostructured materials under the bottom-up process. The following are some of the common process techniques used for the synthesis of nanostructured materials under bottom-up approach.

(i) Physical methods (Vapour phase deposition techniques and Epitaxial techniques)
(ii) Chemical methods (Colloids and sol gel)

Let us discuss the different methods in physical and chemical process methods in the following sections in details:

26.5 VAPOUR PHASE DEPOSITION

The vapour phase deposition (VPD) is the most preferred technique for the synthesis of metal, insulator and semiconductor nanostructured materials. There is an increasing interest in the vapour phase deposition technique due to elegant way to control the process parameters. This will help to produce size, shape and chemical composition controlled nanostructured materials. This technique is more suitable for the production of large scale nanomaterials. The materials of interest are evaporated and brought into the gaseous phase atoms or molecules. The gas phase atoms or molecules form clusters and thus they are deposited on a substrate. One can obtain a single layer of film or multilayer of films or powders of nanostructured materials, depending on the potential applications. It is essential to maintain an adequate pressure in the sample chamber for the synthesis of nanostructured particles. The adequate pressure helps in melting the materials to evaporate. The vapour phase technology is broadly classified into three categories, as given below:

(i) Physical Vapour Deposition (PVD)
(ii) Chemical Vapour Deposition (CVD), and
(iii) Plasma Enhanced Chemical Vapour Deposition (PECVD)

26.5.1 Physical Vapour Deposition

The materials of interest are evaporated and hence, the atoms or molecules are in gas phase. The gas phase atoms or molecules are used to obtain the nanostructured materials in any one of the methods, namely, (i) evaporation, (ii) sputtering, (iii) ion plating, and (iv) laser ablation. Let us discuss the experimental set-up used for the synthesis of nanostructured materials in the evaporation and sputtering method.

(i) Evaporation The schematic representation of the experimental set-up used for the synthesis of nano materials by evaporation is shown in Fig. 26.11. It consists of a bell jar, in which an inert gas or reactive gas is filled after vacuum. The materials to be evaporated are placed in the crucibles and are heated either by resistance or an electron gun until sufficient vapour develops. The evaporated atoms or molecules are allowed to condense on a cold finger which is cooled externally by liquid N_2. The nanoparticles on the cold finger is scraped by the scraper and then collected to the piston anvil through a funnel. The piston anvil is used to obtain the compacted nano powders. The desired purity of the nano powder is obtained since the evaporation is done at the vacuum chamber with the pressure of an inert or reactive gas. This method is more suitable for non-conductive materials or high melting materials.

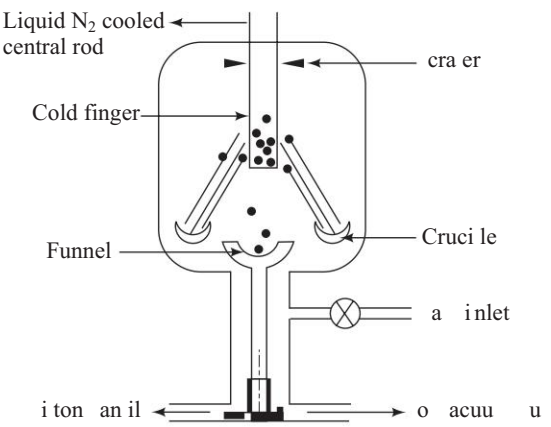


Fig. 26.11 *Physical vapour deposition – Evaporation*

(ii) Sputtering In case of compound materials, the materials will dissociate before evaporation and hence, the proportion of the constituents stoichiometric is not maintained in the final products. Therefore, it is advisable to use the sputtering technique for the deposition of stoichiometric products.

An inert gas like Ar is incident on the target material. The atoms (or) molecules in the ionized form hit the target material and knock out the surface atoms. The knocked out atoms are deposited on the second solid surface known as *substrate*. The removal of the atoms from the first solid state, i.e., target is known as *erosion*. The sputtering is achieved by two different method, namely, DC voltage and RF voltage.

Glow discharge The cross sectional view of the experimental set-up used for glow discharge, i.e., DC sputtering is shown in Fig. 26.12. The target is held at the cathode while the substrate is held at the anode or ground or floating potential. After attaining a

The McGraw·Hill Companies

suitable pressure in the chamber, the argon gas is passed into the chamber at a pressure of < 0.1 torr. When a suitable voltage is applied between anode and cathode, a small current flows over the electrodes. The observed current is due to the pressure of ions and electrons that exists in gas and the secondary electrodes which leave the target after bombardment. As voltage increases, the above contribution also increases. However, at a certain high voltage, the plasma region is obtained. The plasma region consists of mixture of particles link electrons, ions, neutrons and protons. This method fails when the target is insulating in nature, rather it requires a very high voltage.

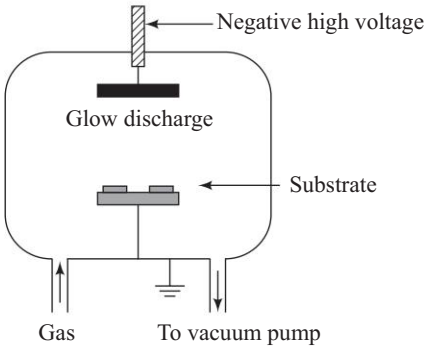


Fig. 26.12 *DC sputtering*

RF sputtering A high voltage (>10^6 V) is used to discharge the electrode when a target material is insulating in nature. When a high frequency-low voltage is applied to the system, it will result in a change in the polarity of the cathode and anode alternatively. As a result, sufficient ionization is achieved by oscillating electrons. A cross sectional view of the RF sputtering system is shown in Fig. 26.13. One can use the sputtering technique to deposit number layers on the substrate.

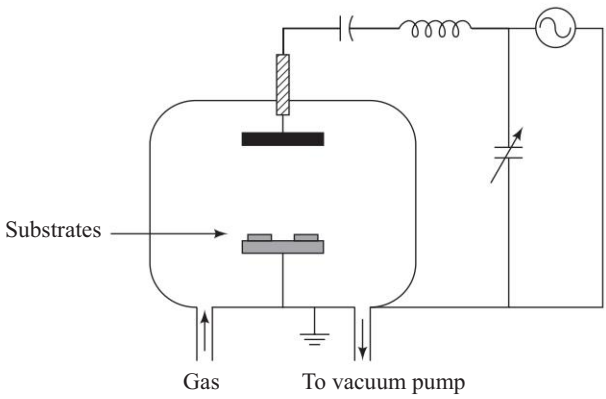


Fig. 26.13 *RF sputtering*

26.5.2 Chemical Vapour Deposition

In chemical vapour deposition (CVD), the atoms or molecules which are in the gaseous state either are allowed to react homogeneously or heterogeneously depending on

the applications. In case of homogeneous CVD, the particles or atoms / molecules in the gas phase are diffused towards the cold surface due to the thermophoric forces. The diffused particles can be scrapped from the cold surface to give nano powder (or) deposited onto a substrate to form a film known as *particulate film*. However, in case of heterogeneous CVD, a dense film of nanoparticles is obtained on the substrate surface. CVD is an excellent method which is used to control the particle size, shape, crystallinity and chemical compositions. One can use this method to obtain high purity nanomaterials and multicomponents systems as well by controlling the chemical reactions.

The schematic representation of the CVD experimental set-up used to synthesis the nanoparticles are shown in Fig. 26.14. The metal-organic precursor is introduced into the hot zone of the reactor employing flow controller. The precursor is vaporised employing the inductive or resistive heating method. An inert gas like Ar or Ne is used as carrier gas. The evaporated matter consists of hot atoms which undergoes collision with the atoms in the cold gas and hence, loses its energy. Thus, the colloidal atoms undergo condensation into small clusters through a homogeneous nucleation. The clusters continue to grow in supersaturated region. Other reactants are added to the clusters to control the chemical reactions. In order to control the size, the clusters are removed from the supersaturated environment through a carrier gas. The cluster size is controlled by controlling the following parameters, namely, rate of evaporation (energy input), rate of condensation (energy removal) and rate of gas flow (cluster removal).

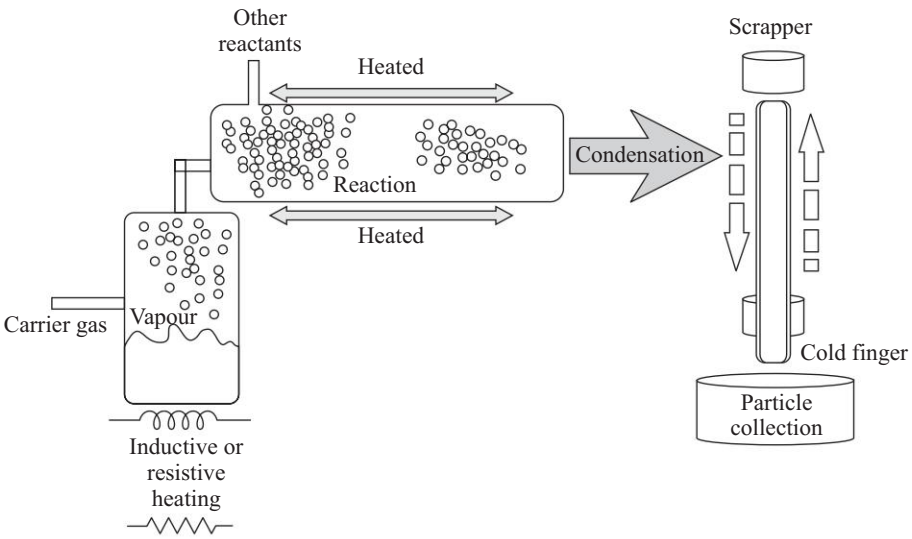


Fig. 26.14 *Chemical vapour deposition*

The condensed clusters are allowed to pass through the cold finger. The nanoparticles are collected using the scrapper as shown in Fig. 26.14. In addition to the formation of single phase nanoparticles, two phase (or) doped nanoparticles can also be synthesised using the CVD method. This can be achieved by supplying two precursors at the first end of the reactor. The CVD method is used to produce defect

The McGraw·Hill Companies

free nano particles. Due to the simplicity of the experiment, the scaling up of the unit for mass production in industry is achieved without any major difficulties.

26.5.3 Plasma Enhanced Chemical Vapour Deposition

A typical experimental set-up used in plasma enhanced chemical vapour depositors (PECVD) is shown in Fig. 26.15. The PECVD method is used to synthesize either amorphous or microcrystalline nanoparticles. The major difference between the conventional CVD and PECVD is the addition of Ar for the ignition of the plasma and of H_2. In PECVD, the source material is sputtered either by DC or RF sputtering method. The synthesis of amorphous or microcrystalline nanoparticles is determined by the degree of SiH_2 content in H_2. Thus, a high concentration of H_2 etches the amorphous phase and gives out the micro crystalline nanoproducts. One can also have a more favored etching process using a high frequency in the order of 100 MHz rather than the usual low frequency, i.e., 13.56 MHz. The PECVD method is used to produce ultra pure and non-agglomerated nanoparticles.

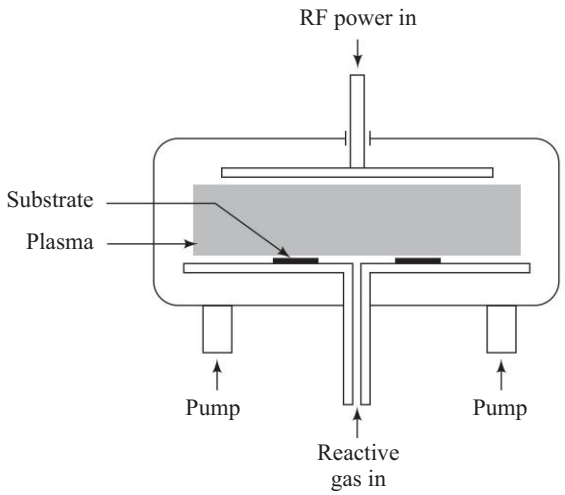


Fig. 26.15 *PECVD method*

26.6 EPITAXIAL TECHNIQUES—SYNTHESIS OF NANOMATERIALS

In the bottom-up method, techniques like molecular beam epitaxy (MBE), metal organic vapour phase epitaxy (MOVPE) and liquid phase epitaxy (LPE) methods are discussed briefly in the following sections:

26.6.1 Molecular Beam Epitaxy

Molecular beam epitaxy (MBE) is basically a sophisticated ultra high vacuum (UHV) based technique used for producing high quality epitaxial structures with mono layers

(ML) control. It is used as a tool for growing high purity semiconductor film by means of producing epitaxial layer of metals, insulators and as well as superconductors. This technique enables a sharp interface between one type of alloy and the next during the epitaxy growth.

The principle behind this technique is very simple. The solid material source is evaporated from solid ingots either by direct heating or by an electron beam. The evaporated atoms or clusters are migrated into an UHV environment and impringe on a hot surface. The atoms or clusters are diffused and eventually incorporated into growing film. The rate of growth depends on the flux of material in the molecular beams and it can be controlled by the shutters with on/off control. The principle of the epitaxial growth can be explained by considering Fig. 26.16 in a simple way. The atoms on a clean surface are free to move around until they find a correct position in the crystal lattice to bond. The atom experiences more binding forces at the step edges than on the surface as shown in Fig. 26.16. Therefore, a growth of the atom takes place at the step edge. The mobility of the atom on the surface increases with increase in temperature and hence, the efficiency depends on temperature. Typical growth rate is 1 ml per second or 1 micron per hour, which is at a pressure of 10^{-6} m bar.

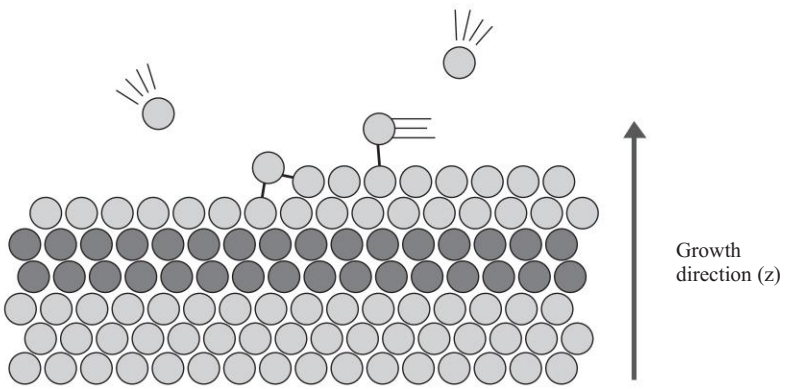


Fig. 26.16 *Molecular beam epitaxy growth*

In order to get good quality of epitaxy growth, the following extra care should be taken while doing the experiment. The substrates should be carefully prepared by ensuing negligible quantities of impurity and cleaned. The reaction chamber should be evacuated not less than 10^{-11} mbar with a proper cooling of the chamber walls. In addition, the source should be of extra pure quality.

A schematic representation of the MBE system growth chamber is shown in Fig. 26.17. The important components of the MBE system are effusion cells, substrate manipulators, pumping systems, vacuum systems, liquid nitrogen cryopanels and analysis tools:

The effusion cell provides an excellent flux stability, uniformity and material purity. The above system can withstand up to a very high temperature (1673 K) for a longer period. The cells are placed on a source flange and are focused to the substrate heater, which gives optimised flux uniformity. The homogeneity of the epitaxy growth is improved with the help of a continous azimuthal rotation (CAR)

The McGraw·Hill Companies

about its own axis. A beam flux monitor (BFM) is located opposite to the substrate holder. The molecular beam intensity is calibrated by rotating BFM through an angle of 180°C through the CAR assembly. A stainless steel chamber is used for sample preparation. The necessary UHV is provided in the chamber using a turbo molecular pump. The liquid nitrogen cryopanels are surrounded internally both at the main chamber and the source flange. The cryopanels provide the necessary thermal isolation between the different cells and for the additional pumping of the residual gas.

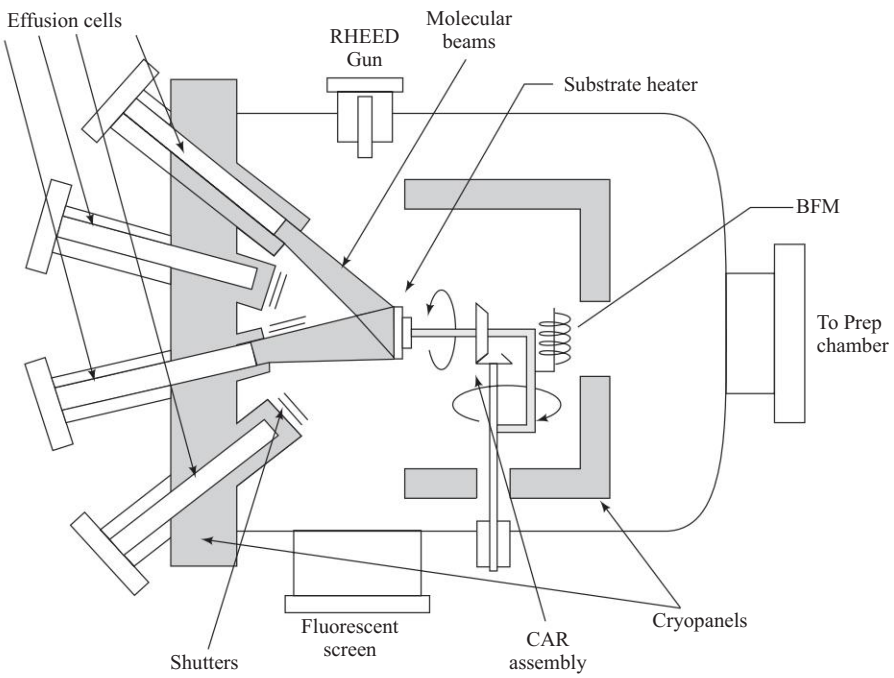


Fig. 26.17 *Typical MBE growth chamber*

The epitaxy growth can be monitored by a number of methods, namely, reflection high energy electron diffraction (RHEED), low energy electron diffraction (LEED), Auger electron spectroscopy (AES) and modulated beam mass spectroscopy (MBES). The RHEED method is used to explore the completed monolayer and partial layer, respectively, from the maximum and minimum diffracted electron beam patterns due to the forward scattering at grazing angle. The surface morphology of the MBE growth can be studied using LEED which takes place in backscattering geometry. The type of atoms present in MBE is studied by AES while MBES technique is used to explore the chemical species and the reaction kinetics. The characterisation techniques such as photoluminescence and electron microscopy are used for post growth analysis.

The following are the salient features of MBE. The epitaxy growth rates are typically in the order of a few Å per second. At any function of a second, one can shut down the molecular beam which facilitates an abrupt transition from one material to another. The main drawback of the MBE is the lower yield.

26.6.2 Metal Organic Vapour Phase Epitaxy

The metal organic vapour phase epitaxy (MOVPE) system is used to produce the complex semiconductor film systems which are used in the modern devices like lasers and transistors for mobile phones. These semiconductors consist of two or more components, e.g., gallium arsenide (GaAs), aluminium nitride (AlN) etc. The formation of the epitaxy layer at the substrate takes place due to the pyrolysis action of the constituent chemicals. Thus, in the MOVPE, the crystal growth is only by chemical reaction and not by physical depositon as in MBE. The sample chamber does not require any vacuum. However, it requires a moderate pressure of the order of 2–100 Kpa. The principle behind the MOVPE technique is shown in Fig. 26.18a.

(a) MOVPE- Principle
(b) MOVPE- Experimental set-up

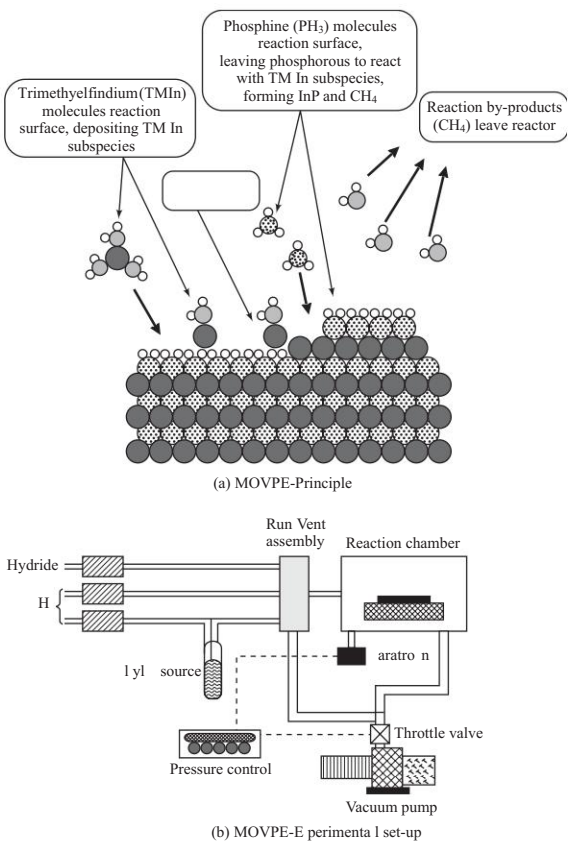


Fig. 26.18 *MOVPE*

The schematic representation of the MOVPE system is shown in Fig. 26.18. The sample chamber which is made up of stainless steel or quartz does not react with

The McGraw·Hill Companies

chemicals. The process chamber includes reactor walls, a susceptor, gas injection and temperature control unit. The necessary water circulations are provided within the reactor walls through the channels to avoid overheating. The reactor wall and the susceptor are separated by liner made up of quartz or ceramics. The materials like graphite, which are resistant to metal organic compounds, are used as a material for susceptor. The susceptor is located at controlled temperature region. The wafers which come out during the heating of the sample sit at the susceptor. The susceptor also requires special coating to prevent corrosion by ammonia for growing materials like nitrides. The gas inlet and switching system provides the required gas through a device known as *bubblers*. The bubbler is used to bubble the carrier gas through the organic metal liquid. The carrier gas picks up the metal organic vapour and then transports it to the reactor. The carrier gas flow and the bubbler temperature will help to control the rate of transport of metal organic vapour. The toxic waste products after conversion, either in liquid or solid state, are removed from the chamber through the gas exhaust and cleaning systems.

The starting material, namely, the precursors, are fed into the reactor with the aid of a carrier gas (nitrogen or hydrogen). The reactor contains the substrate which is in the form of very thin single crystal wafer. The sample chamber is heated in the temperature range of 773 to 1773 K depending on the material used. The metal organic gas is liberated from the material due to the heating and then mixed with the carrier gas. The above fragment together moves to the substrate surface and settles and migrates over the wafer, resulting in the growth of wafer monolayer by monolayer. The scrubber is used to filter, neutralize or dilute the carrier gas, like hydrogen, before pumping into the exhaust air.

The MOVPE technique can be used to grow semiconductors from a variety of organo metallic chemicals. The MOVPE is the dominant process technique to manufacture laser diodes, solar cells, LEDs, etc.

26.6.3 Liquid Phase Epitaxy

Liquid phase epitaxy (LPE) method is used for the growth of semiconductor layers from the melt on solid substrates. The LPE grown semiconductors are used for the manufacturing of optoelectronic devices like LEDs, lasers and photodiodes. The LPE technique is used to grow materials with very low unintentional impurity density, high mobility and good luminescence properties. The LPE techniques involve the near equilibrium growth of a material from a supersaturated solution which is placed in contact with polished solid substrates. The LPE is mainly used to grow very thin, uniform and high purity ternary and quaternary III- V components of GaAs substrates. Arsenide is dissolved in a Gallium solvent. The temperature is lowered from the equilibrium temperature at a controlled rate when the substrate is just in contact with solution and substrate. This results in a supersaturation in the solution and hence, the dissolved material grow epitaxially on the substrate. The equilibrium condition depends on the temperature and the concentration of the semiconductor melts.

There are different techniques for the LPE growth of the semiconducting materials, namely, tipping, dipping and sliding techniques. The different techniques are due to the development of the growth of materials from single to multi components. Let us discuss briefly the sliding boat LPE technique used for the growth of the nano particles in detail.

The sliding boat technique is an improvement of the slide technique which was mainly developed for the multi layer growth. The schematic diagram of a simple sliding boat system used for the LPE is shown in Fig. 26.19. It consists of a resistance heated type furnace. The system consists of number of boats which contain different semiconducting components. For the uniform LPE growth, a flat temperature profile is maintained at the centre of the zone. In addition, a three-zone system is also maintained to provide sufficient temperature variations. In order to improve the purity of the growth, flushing with H_2 is done. The solution and substrate are brought in contact by moving the slider in one direction. After allowing some time for the growth, the slider is moved in the same direction so that the substrate meets another solution. Hence, a new layer is grown. Once the desired structure is made, the growth can be terminated by removing the substrate. Thus, by employing this technique, one can make one or several layers of growth at the same time.

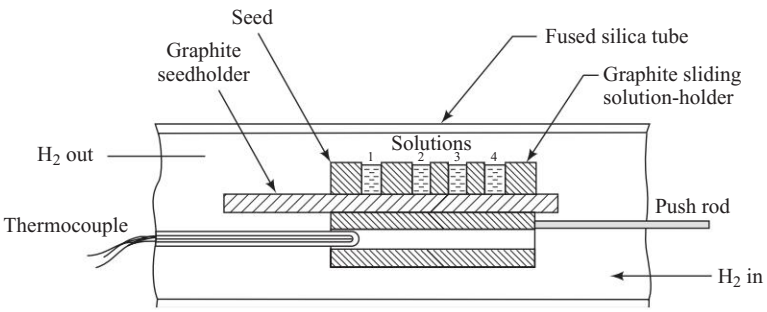


Fig. 26.19 *LPE–Slider system*

26.7 CHEMICAL METHODS—NANOMATERIAL SYNTHESIS

Compared to physical methods, the chemical methods are very simple techniques, inexpensive, and require fewer instruments to obtain the nanoparticles. In the chemical methods, the resultant materials are obtained in the form of liquid. However, one can convert into any form like dry powder or thin films or gel, depending on the applications. One can obtain colloidal particles in solutions. Later, it can be converted into a powder, after proper filtration and drying.

In chemical methods, different methods such as colloids, micro emulsion and sol gel method are used for the synthesis of nanoparticles. In the following sections, a brief description on the processes, namely, colloids and sol gel are given in brief to understand the processing of nanostructured materials.

26.7.1 Colloids

Colloids are a material which are made of two or more phases of same or different materials. Out of the three dimensions, one dimension is less than a micrometre. There are different forms of colloids, namely, particles or plates or fibers as shown in Fig. 26.20. An example for the colloid is the liquid in gas, i.e., fog.

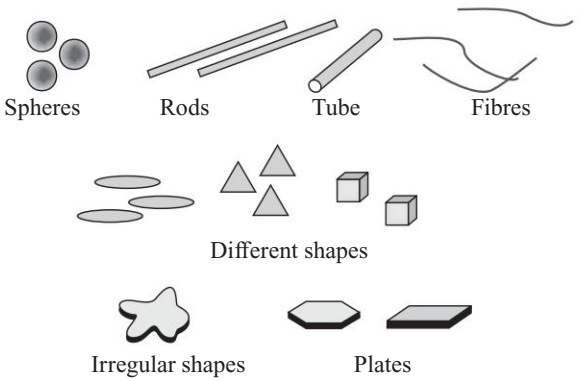


Fig. 26.20 *Types of colloids*

(i) Synthesis of Nanoparticles Generally, the colloidal particles are suspended in some host matrix. One can synthesise metal, alloy, semiconductor and insulator particles of different size and shape in either aqueous or non-aqueous media.

A typical chemical reactor chamber used to synthesise the nanoparticles is shown in Fig. 26.21. Glass reactor chamber is usually used for chemical reactions in which the colloidal particles are obtained. The reactor has a provision to introduce the precursors and gas. The provisions to measure temperature and pH during the chemical reactions are included. Inert gas, like argon or nitrogen gas is admitted, depending on the requirements through the separate control systems. The necessary stirring arrangements, i.e., teflon-coated magnetic needle has been provided. One can take the products after the reaction is over. One can use the above chemical reactor, to synthesise metal or semiconductor or alloy nanoparticles at the required size and shape. For example, gold nano particles are synthesised by the reduction of chloroauric acid ($HAuCl_4$) with tri sodium citrate ($Na_2C_6H_5O_7$).

$$HAuCl_4 + Na_3C_6H_5O_7 \rightarrow Au^+ + C_6H_5O_7 + HCl + 3NaCl \qquad (26.2)$$

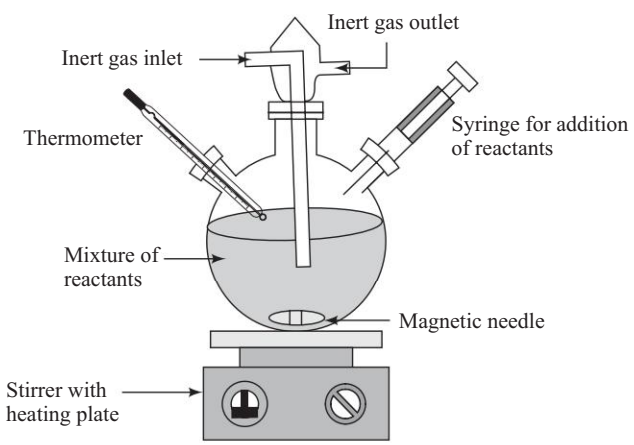


Fig. 26.21 *Typical chemical reactor chamber*

The reduction reaction takes place under four stages to yield the products. The gold atoms are formed by nucleation and condensation. The obtained gold atoms are grown in bigger size by the reduction of Au ion on the surface. These bigger atoms are stabilised by oppositely charged citrate ions. Thus, the stabilised nano gold particles are formed and are represented in Fig. 26.22.

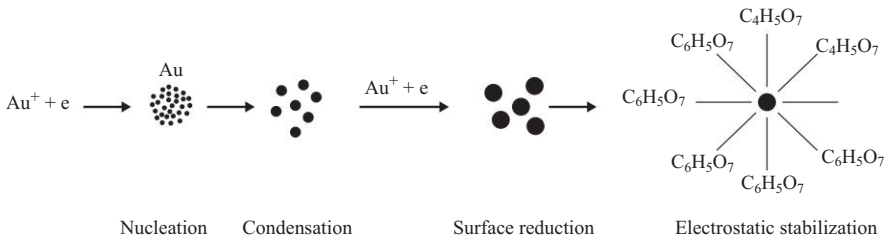


Fig. 26.22 *Gold nanoparticles*

The above reaction is carried out in water. Thus, the resultant nanoparticles are different in colour depending on their size. This reveals that optical properties of the nano particles are size dependent.

26.7.2. Sol Gel Method

The sol gel process is a wet-chemical technique, i.e., chemical solution deposition technique used for the production of high purity and homogeneous nanomaterials, particularly metal oxide nanoparticles. The starting material from a chemical solution leads to the formation of colloidal suspensions known as *sol*. Then, the sol evolves towards the formation of an inorganic network containing a liquid phase called the *gel*. The removal of the liquid phase from the sol yields the gel. The particle size and shape are controlled by the sol/gel transitions. The thermal treatment (firing/calcinations) of the gel leads to further poly condensation and enhances the mechanical properties of the products, i.e., oxide nanoparticles.

The precursors for synthesising the colloids are metal alkoxides and metal chlorides. The starting material is processed with water or dilute acid in an alkaline solvent. The material undergoes a hydrolysis and polycondensation reaction which leads to the formation of colloids. The colloid system composed of solid particles dispersed in a solvent contains particles of size from 1 nm to 1 mm. The sol is then evolved to form an inorganic network containing liquid phase (gel). The schematic representation of the synthesis of nanoparticles using the sol gel method is shown in Fig. 26.23.

The sol can be further processed to obtain the substrate in a film, either by dip coating or spin coating, or cast into a container with desired shape or powders by calcinations. The chemical reaction which takes place in the sol gel metal alkoxides $M(OR)_2$ during the hydrolysis process and condensation is given below:

$$M\text{-}O\text{-}R + H_2O \rightarrow M\text{-}OH + R\text{-}OH \text{ (Hydrolysis)} \quad (26.3)$$

$$M\text{-}O\text{-}H + R\text{-}O\text{-}M \rightarrow M\text{-}O\text{-}M + R\text{-}OH \text{ (Condensation)} \quad (26.4)$$

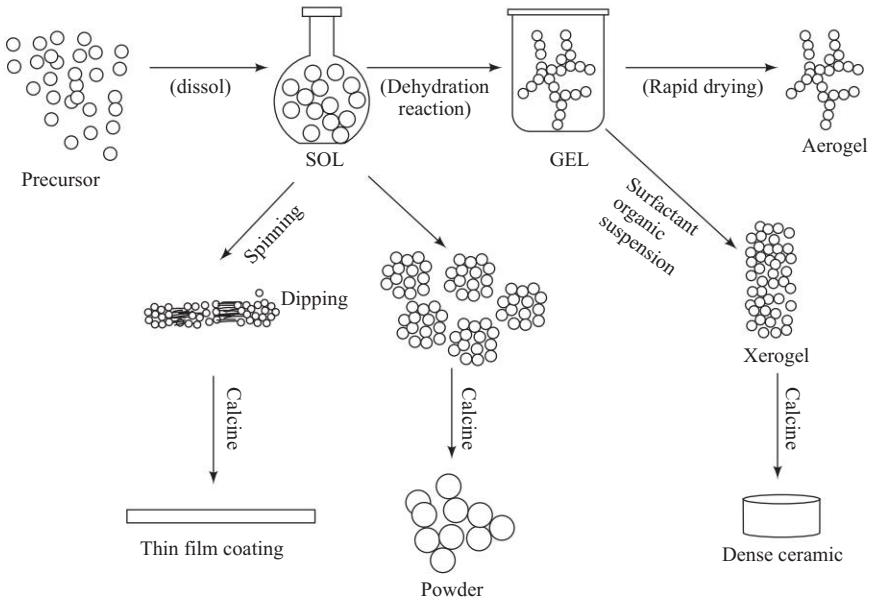


Fig. 26.23 *Schematic representations of sol gel synthesis of nano particles*

The sol gel method is an interesting, cheap and low temperature technique which is used to produce a range of nanoparticles with controlled chemical compositions. One can produce the aero gel, a highly porous material like glass and glass ceramics, at a very low temperature by controlling the process parameters. The sol gel derived nano particles finds wide spread applications in various fields like optics, electronics, energy, space, bio sensors and drug delivery.

26.8 HYBRID METHODS—SYNTHESIS OF NANOMATERIALS

In recent years, many new methods are emerging along with the existing methods for the synthesis of nanoparticles and nanostructured materials. In this series, template based nanomaterials synthesis, self-assembly and self-organisation are some of the new routes developed for the nano synthesis. In the following sections, brief descriptions of the above methods are given for a better understanding:

26.8.1 Template Based Synthesis

The synthesis of nanoparticles on templates is a very general method, but it has gained momentum in recent years. This process is used for the fabrication of nanorods, nanowires and nanotubes of polymers, metals, semiconductors and oxides. The template growth of nano rods and nano tubes has been obtained using different templates with nanosized channels. These templates will have a desired pore (or) channel size. The template also requires some additional parameters like morphology, size distributions and pore densities. Anodized alumina membrane and radiation track etched polymer membranes are the most commonly used and commercially available templates.

Three important points are to be considered while using the template based synthesis of nanoparticles. (i) The template should be chemically and thermally inert during the synthesis process. (ii) The depositing material or solutions must wet the internal pore walls. (iii) The deposition of nanoparticle should start from the bottom end of the template and hence, it has to proceed to the other end on the same side. The above procedure is adopted for the synthesis of nanowires or nanorods. On the other hand, for the growth of nanotubules, the deposition should start from the pore walls and proceed inwardly. The pore blockage is produced during the inward growth. This pore blockage is avoided in nanorods or nanowires. One can remove the nanorods or nanowires from the template after the experiments.

The nanorods or nanowires are synthesised by electrochemical deposition. This technique is also known as *electro deposition*. The following steps are involved during the electro deposition. When an external field is applied to the electron, the oriented diffusion of positively charged cations through the solution takes place. The reduction of positively charged cation at the growth or the deposition surface will act as an electrode. After this process, the electrode is completely separated from the solution due to the deposition of positively charged cations. The deposition process will continue, since the deposited cations permit the electrical current to pass through it. The process of coating the metal on the template through electrochemical deposition is known as *electroplating*. The deposition is confined inside the pores of the template membranes and hence, the nano composite are produced. The template membrane is removed from the experimental set-up and the nanoparticles are scrapped from the templates. In recent years, a hybrid of the template synthesis with other chemical process, the nanoparticles with the core-shell and onion structures are produced.

26.8.2 Self-Assembly

We knew that the cells and tissues of the human body consist of proteins. The proteins are larger molecules which are formed by the successive addition of hundreds of amino acids. When an amino acid is added to another, it makes a bond by transferring the RNA molecules. The sequence of above arrangements is known as *polypeptide chain*. An increase in the sequence will increase the length of the polypeptide chain. The increased sequence is known as *protein*. The above process is known as self *assembly* which occurs in all living tissues. Thus, the self assembly process at small length scale is important in biology, but has an analog in nanotechnology.

In nanotechnology, the molecules or atoms at their nano scale length are formed in a well defined stable and ordered molecular complex systems for different applications. A weak reversible interaction between the molecules leads to the creation of an ordered structure at equilibrium during the above process. Generally, the larger molecules, known as *macromolecules*, take number of time consuming steps to break and remake the strong covalent bonds during the organic synthesis process. The above reaction is carried out under kinetic control systems. The resultant products or a yield is very small and hence, the recognition or the removal of the error is not possible. The self assembly is the strategy used for the nano fabrication which involves the design and synthesis of the desired structures. The synthesis process involves non-covalent bonding interactions like hydrogen bonds and Van der Waals forces. The above synthesis is carried out under thermodynamic conditions. One can rectify the errors to obtain small individual molecules under thermodynamic equilibrium conditions.

The McGraw·Hill Companies

It is clear that a regular high perfect structure by way of controlling the growth materials under equilibrium condition facilitates self assembled structures. Thus, self assembly of atoms, molecules and nanoparticles leads to the desired structure without any top-down tool for making. The self-assembly has a number of advantages over the conventional methods for making nanostructured assemblies. The difficult steps in nano fabrication are the atomic level modifications of the structures employing the synthetic chemistry. The development of functional and complex structures in biology gives strong inspirations. The above method produces defect free nanostructured materials.

(a) Self-assembly in Inorganic Materials We know that individual nanoparticles have exotic physico-chemical properties which are size dependent. Thin films of the above nanoparticles are required for many practical applications. The important process towards the novel application of the nanostructured materials are the surface modifications, assembly, patterning, orientations, and alignment into a functional network. The self assembly provides an immense technique towards the products of functionalised nanostructured materials for potential application.

The self assembly of inorganic materials is achieved by the deposition of highly structured process known as *coherent process*, i.e., defect or *dislocation-free coherent process*. One can produce the quantum dots of germanium (Ge) or silicon (Si) employing the above process as shown in Fig. 26.24. The germanium and silicon have roughly 4% lattice mismatch. Employing the hetero epitaxial growth, the germanium is deposited on the single crystal silicon in a coherent manner. This results in three to four layer deposition of Ge on Si without any defects or dislocations. The further deposition of Ge on Si is not accommodated due to the induced lattice strain. The deposition results in the spontaneous formation of quantum dots or nano sized islands. The size of the quantum dots is controlled by the growth temperature, and the smoothness of the surface of the substrate.

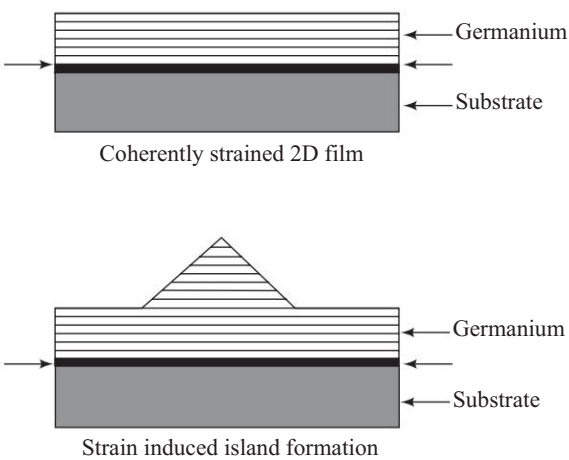


Fig. 26.24 *Island formations by the deposition of Ge on Si*

One can produce inorganic nanoparticles employing the self-assembly method. Consider that the silica particles are synthesized by employing the sol-gel method.

The silica particles which are in aqueous form are placed in a glass substrate. The aqueous medium is evaporated from the substrate. The particles self assemble after some interval of time due to the weak Van der Waals interaction among the particles. The particles are allowed to form a uniform two–dimensional structures.

(b) Self-assembly of Nanoparticles in Organic Molecules The self assemblies of the prepared inorganic nanoparticles are assembled on solid substrates employing organic molecules. The organic molecules which are used in the self assembly process help the nanoparticles to transfer to the solid substrate by binding with the solid substrate and attaching with the nanoparticles. The self-assembly of CdS nanoparticles on a solid substrate is shown in Fig. 26.25.

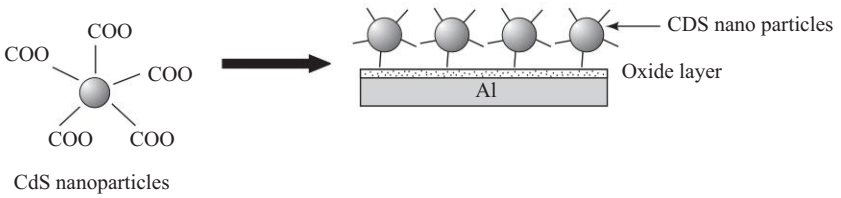


Fig. 26.25 *Self assembly of nanoparticles—organic molecules*

The CdS nanoparticles are functionalized with carboxylic group (COO) as shown in Fig. 26.25. The CdS nanoparticles are transferred into aluminum thin films employing organic molecules.

(c) Self-assembly using Biological Templates The S-layer extracted from the bacterial cells has been transferred into the metallic substrates or grids as shown in Fig. 26.26. The nanoparticle to be assembled on the biological substrate is heated and hence, the liquid or vapour phase is deposited on the biological substrate. The substrate is initially treated with cadmium salt followed by Na_2S, and results in the formation of CdS nanoparticles.

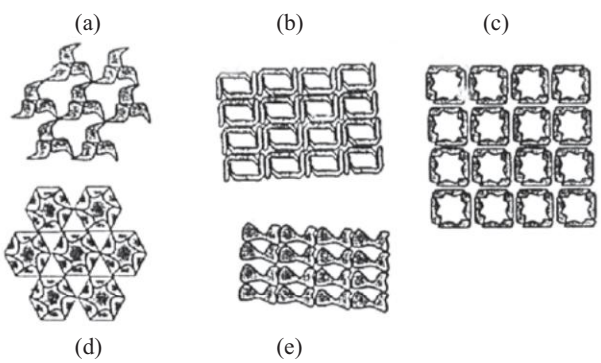


Fig. 26.26 *S-layer lattice—different structures*

Similarly, the gold nanoparticles can be embedded on the biological templates as shown in Fig. 26.27.

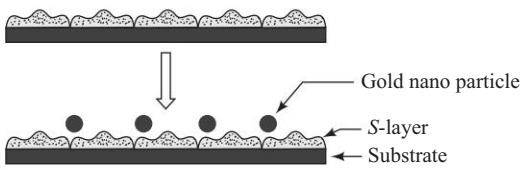


Fig. 26.27 *S-layer—Assembly of CdS or Au nanoparticles*

26.8.3 Self-Organisation

Self-organisation method is used for the fabrication of nanomaterials for a variety of applications like electronic, chemical / biological sensors and membranes. This method helps in the mass production of products like lithographical tools without using an expensive technique. One can produce highly ordered nanohole arrays on several tens of nanometers employing the self organisation method. The highly ordered nanohole arrays find potential applications such as high density storage media, high density chemical sensors, nano electronic devices and functional biochemical membranes. The mechanism for nanohole self-organisation is still at the infant stage. Even though several methods have been proposed, let us discuss the highly ordered nanohole array obtained by two-step method anodization proposed by Masuda and Fukuda.

In the two step anodization method, a high purity aluminium with 0.3 M oxalic acid solution under a constant voltage of 40 V at 273 K has been used to obtain the nanohole array. The first anodization was carried out for a period of 160 h. This helped to obtain an excellent regularity at the bottom of the nanohole. The nanohole regularity can be increased by increasing the film thickness. A P-C etch solution was proposed with a mixtures of 35 ml/litre of 85% H_2PO_4 and 20g/litre CrO_3 at 353 K as proposed by Schwartz and Platter. The above solution is used to etch the first obtained anodic alumina film. The post etched aluminum surface has a periodic surface roughness with a regular array of nanohole bottoms. Due to this initial interface, an excellent regularity is obtained after the second anodic oxidation. The SEM micrograph of the alumina nanohole array formed by two step oxidation at 40 V using 0.15 M oxalic acid is shown in Fig. 26.28.

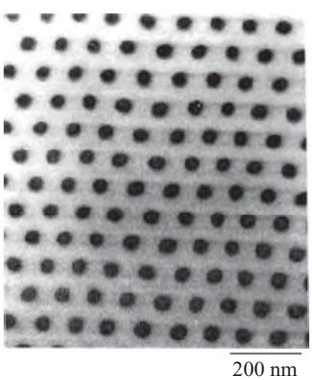


Fig. 26.28 *Alumina nanohole array – Plan view*

The plan view of the alumina nanohole array reveals a trigonal lattice with an average diameter of 36 nm. One can also enlarge the nanohole array by chemical etching with diluted phosphoric acid. Hence, the distance between the hole increases to 90 nm. The cross sectional view of nanoholes formed due to the long anodization (30 s) is shown in Fig. 26.28. The depth of the nanohole is 220 nm, while the neighbouring nanoholes are separated by 50 nm thick alumina side wall. The hole bottom is closed by barrier film of 30 nm thickness.

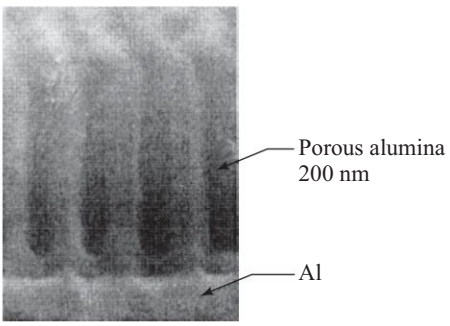


Fig. 26.29 *Alumina nanotubes array—cross-sectional view*

The self-organised porous alumina nanohole arrays have been used to fabricate a variety of nanomaterials. Some of the methods which are used to fabricate nanostructured materials are etching semiconductor substrate using a porous alumina film as a mask, pattern transfer using porous alumina as a template, etc. Self-assembly and self-organisation are important concepts, in both nanotechnology and biology.

26.9 NANOTECHNOLOGY AND ENVIRONMENT

When a new technology emerges, there is more concern about its impact on social life, health and environment. A common man would like to know the development of science and technology behind nanotechnology and the impact in social and environment life. Therefore, the environment protection of nanotechnology is more essentially required since the atoms or molecules are controlled at the nanoscale length to develop new systems or products. Thus, the environmental free nanotechnology, i.e., green nanotechnology has a pivotal role against the environmental issues.

There are few questions which arise immediately. What is the effect on nanomaterials on human bodies, animal or plants when it is inhaled or in contact? Do the nanomaterials pollute water, air and food? What is the role of nanomaterials on global climate changes?

In recent years, more work is under progress to investigate the above issues almost in all countries worldwide. In general, very small tiny particles of the order of less than 0.1 mm will affect the respirator system and other organs. There are a few findings that the nanoparticles with size around 50 nm can affect the cells. However, the detailed studies of the nanomaterials and its size-dependent effects on human bodies and animals

The McGraw·Hill Companies

are under investigation. It is interesting to note that the nanoparticles like silica are used to increase the photosynthesis process of the plants and also as a nutrious food. The nano silica particles along with a suitable solvent are sprinkled on the leaf of the plant and it cleans the leaf surfaces, which facilitate more photosynthesis. Similarly, the nano silica particles along with fertilizers are used as a nutrious food to plant to increase its growth and yields.

We know that the applications of the nanoparticles such as coatings, electronics industries, fuel storage cells, sensors, drug delivery, implant applications, etc., are due to the novel physical and chemical properties. Nanoparticles, nanotubes, nanorods, nanofilters, etc., are manufactured in industry in a highly protected environment which prevents the nanoparticles from polluting the environment (or) harming the workers.

It is proven that some of the nanoparticle production processes, like low temperature synthesis route, will themselves reduce the pollution and environmental related issues. In automobiles, the nanomaterial is used as a hydrogen storage or efficient oil filters which reduce the pollution from the vehicles. The development of new systems or protocols based on the nanostructured materials leads to the shrinkage of size and hence, the prices. This will help in bringing this technology product to the poor man also.

Most of the nano sensors are more sensitive than their counterpart bulk sensors and hence, they are known as *smart sensors*. These sensors are used to detect and rectify the problems. For example, water purification filters, detection of toxic ions and gas, robotic machines, etc., utilize nano sensors for the improved efficiency and remedial effect.

26.10 PROPERTIES AND POSSIBLE APPLICATIONS

In view of the available different techniques for the synthesis and fabrication of nanosized particles and nanostructured materials, there is an increasing demand for potential applications. The nanosized particles at the nanoscale length exhibit exotic physio-chemical properties as compared to the bulk materials. The properties of the nanostructured materials mainly depends on the nanostructured induced effects, namely, size dependent effect and the surface or interface induced effect. The physical properties mainly depend on the size. Similarly, the surface or interface induced effect plays a dominant role in chemical processing. Different experimental techniques and measurement parameters are available to explore the above two effects. Therefore, the tailoring of the nanosized particles and nanostructured materials is quite important and needed for the potential applications.

26.10.1 Chemical Properties

Chemistry plays an important role in creating the as-required nanoscale structures. The chemical reactions at the nanoscale structures explore the optical and electronic properties of the nanoscale materials. The chemical reactions are normally governed by electrons, ionization potentials and electron orbital densities. When a particle size is reduced from the bulk, the electronic band in the metals becomes narrow,

The McGraw·Hill Companies

leading to the transformation of the delocalized electronic states into more localized molecular bonds. As a result, the ionization potential increases. A catalyst will have much influence on the variations of geometry and electronic structure of the nanophase materials. For example, the characterisation of nonconducting oxides supported by metal catalysts leads to the formation of clusters of platinum, iridium or osmium with a size less than 1 nm, which are supported by alumina or silica, and they show electronic properties similar to that found for large crystallites of the metal (bulk electronic properties). The nanoscale particles exhibit some unusual behaviours, such as equilibrium vapour pressure, higher chemical potential and solubilities than some material when they are expressed as large particles. The above information reveals that the high surface-to-volume ratio, and the change in the geometry and electronic structure changes the optical and electronic properties of the nanoscale materials due to chemical reactivity.

26.10.2 Mechanical Properties

The nanostructured materials offer a wide range of applications for many challenging problems, in view of their fabrication of a variety of new materials with the required and fine-tuned strength and ductility. The application of the nanomaterials may be on both low and high temperatures. Therefore, the mechanical properties of the nanophase materials both at low and high temperatures are given in brief.

Low Temperature Properties We know that the grain refinement leads to an improvement in the properties of the metals and alloys. For example, a reduction in grain size lowers the transition temperature in steel from ductile to brittle. A major interest in the nanophase materials is the change in mechanical properties with the reduction in grain size. The average grain size and yield stress (σ_y) are simply related by the Hall petch relation given below,

$$\sigma_y = \sigma_0 + \frac{k}{\sqrt{d}} \tag{26.5}$$

where σ_0 is the friction stress, k the constant and d the average grain size.

Similarly, one can write the relation between hardness and grain size as,

$$H = H_i + \frac{k}{\sqrt{d}} \tag{26.6}$$

where H_i is the Vickers hardness and k a constant.

The hardness measurements that are carried out in the nanophase materials as a function of grain size, follow the traditional way and are represented in the form of Hall petch plots. The observed results show an increase in hardening with continuous grain refinement, down to the finest grain sizes.

It is interesting to note that, in most of the cases, the rate of hardening with the increase in the values of $1/\sqrt{d}$ is less at the nanometer scale than that which occurs at the conventional grain sizes. Further, the hardness approaches zero with continuous reduction in the $1/\sqrt{d}$ values. A typical observation made in nanocrystalline Cu is shown in Fig. 26.30.

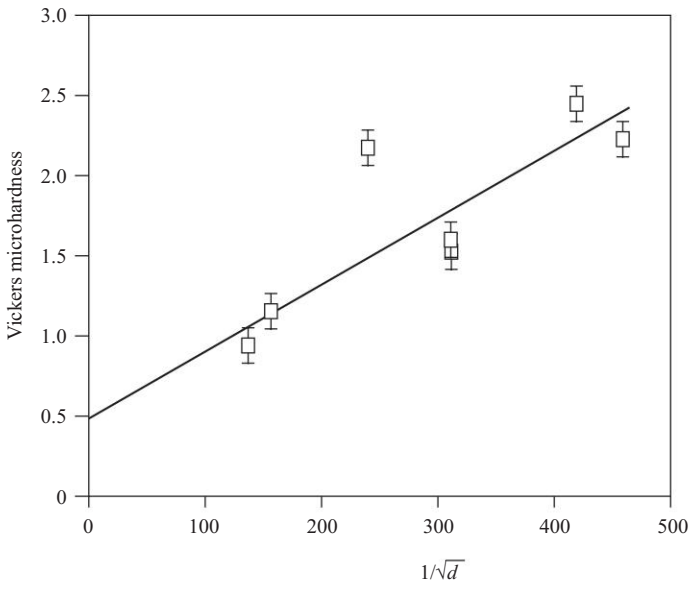


Fig. 26.30 *Vickers hardness versus* $1/\sqrt{d}$

The observed hardness measurements on the nanophase materials reveal positive, zero and negative Hall petch slope due to the change in the structures such as densification, stress relief, phase transformations, grain sizes with respect to the preparation and annealing of the samples. A typical observation made in nanocrystalline Fe is shown in Fig. 26.31.

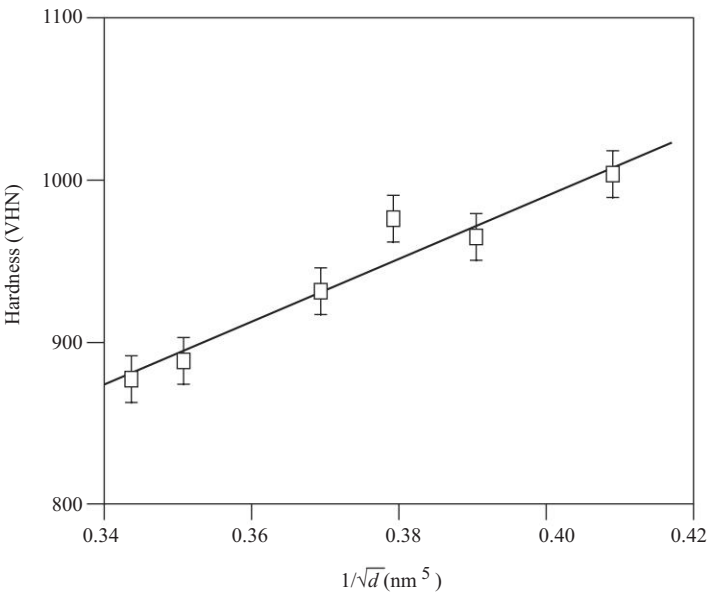


Fig. 26.31 *Hall petch plot of hardness for nanocrytalline Fe*

The McGraw·Hill Companies

High-Temperature Properties The nanophase materials are also termed as super plastic materials, since they exhibit all the common microstructural features of super plastic materials, such as extensive tensile deformation without cracking or fracture. Some of the characteristics of super plastic materials are smaller grain size (less than 5 mm), equiaxed grains, high energy grain boundaries and the presence of second phase. In case of nanophase materials, the occurrence of super plastic temperature is decreased due to the decrease in grain size, which results in an increase in the strain rate.

26.10.3 Applications

The chemical, physical, electrical, magnetic and mechanical properties of the nanophase materials show exotic behaviour in the reduced grain size than the bulk materials. The observed new properties of the nanophase materials lead to several interesting applications, given as follows.

Memories and Electronic Devices The magnetic nanophase materials show a variety of unusual magnetic behaviours when compared to the bulk materials, which are mainly due to the surface/interface effect and which include symmetry breaking, electronic environment charge transfer and magnetic interactions. The magnetisation and coercivity of the nanophase materials are higher. The Fe, Co and Fe(Co)-B based particles with a particle size of 7–20 nm possess an effective anisotropy and coercivity about two orders of magnitude than the bulk materials.

As a result, the nanophase materials are the potential candidates for soft and permanent magnetic materials which lead to several applications. The special applications of these materials are nanoelectronic devices such as nanotransistors, memory devices such as recording heads, magnetic storages, etc. Nanophase materials are used for the manufacturing of size and weight reduced microstrip patch antennas. The important properties of these miniaturised antennas are the large bandwidth, tunability and mechanical flexibility.

Nanophase materials are used for the fabrication of the signal processing elements such as filters, delay lines, switches, etc. The advantage of using nanomaterials in the above devices is the reduction in problems such as cross-talk and interference due to the intrinsic losses in the materials. It is used for the development of robot electromagnetic sensors to operate in harsh environments.

The typical properties of the nanostructured materials and possible applications are given in Table 26.3.

Table 26.3 *Typical properties of nano-structured materials and possibilities*

Property	*Application*
Bulk	
Single magnetic domain	Magnetic recording
Small mean free path of electrons in a solid	Special conductors
Size smaller than wavelength	Light or heat absorption, scattering
High and selective optical absorption or metal particles	Colours, filters, solar absorbers, photovoltaics, photographic material, phototropic material

The McGraw·Hill Companies

Property	*Application*
Formation of ultra fine pores due to superfine agglomeration of particles	Molecular filters
Uniform mixture of different kinds of superfine particles	R&D of new materials
Grain size too small for stable dislocation	High strength and hardness of metallic materials
Surface / interface	
Large specific surface area	Catalysis, sensors
Large surface area, small heat capacity	Heat- exchange materials combustion catalysts
Lower sintering temperature	Sintering accelerators
Specific interface area, large boundary area	Nanostructured materials
Superplastic behaviour of ceramics	Ductile ceramics
Cluster coating and metallization	Special resistors, temperature sensors
Multi-shell particles	Chemical activity of catalysts Tailored optical elements

26.11 STORAGE

Due to the tiny nature of the nanoparticle, the storage and handling is important. Normally, nanoparticles are stored in a cooled condition. This will help to avoid the aggregation of nanoparticles. At room temperature, the mobility and collision of nanoparticles in a liquid phase is very high and hence, it induces the aggregation of nanoparticles. Metal nanoparticles are normally stored in a tightly closed vessel because nanoparticles are highly reactive with atmospheric oxygen. When the metal nanoparticles are exposed to atmospheric conditions, they convert to their corresponding metal oxides. For example, iron nanoparticles are normally stored in ethanol, otherwise they react with oxygen to form iron oxides.

Semiconductor nanoparticles are stored in a dark atmosphere to prevent photo luminescence. Metal oxide nanoparticles are stored in a closely packed dry vessel to prevent hydration. For example, CaO nanoparticles react with moisture to form Ca (OH). Generally, all nanopowders have to be stored in a very compact vessel and a stress-free condition to avoid surface charges. Similarly, the template nanomaterials can be stored in a closed container to avoid the peel of the material.

The following are the important points to be considered for the nanomaterials storage:

- Keep container tightly sealed.
- Store the container which contains nanoparticles in a cool and dry place.
- Ensure good ventilation at the workplace.
- Protect against explosions and fires.
- Do not store together with acids.
- Store away from oxidizing agents.
- Store away from halogens.
- Protect from the humidity and water.
- Nanoparticle containing container can be stored with the above protection in a clean room with high standard.

Keypoints to Remember

- Richard Feynman is considered the father of Nanoscience and technology.
- Nano is equal to one billionth of a metre (1 nm = 1×10^{-9} m).
- Classification of nanomaterials are zero, one, two and three dimensions.
- Nanostructured materials are prepared by bottom-up and top-down approach.
- Lithography means carving or writing on a stone.
- The material which is sensitive to radiation is known as resist.
- Lithography which is classified based on photon radiations are UV light, laser and X-ray lithography.
- The classification of lithography based on particle beam radiations are electron beam, ion beam and neutral beam lithography.
- The de Broglie wavelength of a particle is equal to $\lambda = h/mv$, where h is Planck's constant, m is the mass of particle and v, the velocity of particle.
- In particle beam lithography, the resolution depends on the mass and velocity of particles.
- Electron beam lithography is operated in two modes, namely, vector scan and raster scan methods.
- In the vector scan mode, the electron beam is made to write on a specified region and during scanning of other regions, the electron beam is put off.
- In raster scanning, the material is scanned by line, but the portions of the sample are moved at right angles to the beam according to the pattern to be obtained.
- In dip-pen lithography, the AFM tip is used as a pen and the molecules are used as ink to obtain the required pattern on the substrate.
- The different classifications of vapour phase deposition (VPD) are physical vapour deposition (PVD), chemical vapour deposition (CVD) and plasma enhanced chemical vapour deposition (PECVD).
- Substrates are used to deposit atoms on its surface during the physical vapour deposition method.
- Erosion is the process of removal of atoms from the target.
- Two different types of sputtering methods are DC voltage and RF voltage.
- The formation of colloidal suspension during the process of formation of nanomaterials is known as sol.
- The formation of an inorganic network containing a liquid phase is known as gel.
- Electroplating is the process of coating the metal on the template through electrochemical deposition.
- Self-assembly is the process of arrangements of atoms or molecules at the nanoscale lengths on the surface of a template.

The McGraw·Hill Companies

- The coherent process method is used to obtain self-annealed inorganic nanomaterials through the deposition of highly structured process.
- Self-organisation is the method used for the fabrication of nanomaterials for applications like sensors and membranes. In this process, the nanohole arrays are obtained by self-organisation.

One Mark Question

26.1 The value of one nanometre is equal to_____.
26.2 Example for zero dimension nanomaterial is _____.
26.3 _____ process is used to assemble atoms or molecules to form nanomaterials.
26.4 Nanoparticles are obtained by breaking the bulk solids employing _____ method.
26.5 Mechanical grinding is an example for nanomaterial synthesis of _____ method.
26.6 In mechanical grinding _____ is used to increase the brittleness of the powders.
26.7 In mechanical grinding, hard materials are used to synthesis _____ materials.
26.8 An increase in microhardness increases the _____ resistance.
26.9 The carving or writing on a stone is known as _____.
26.10 _____ material is sensitive to radiation.
26.11 Lithography is classified based on the type of radiations, namely, _____ and _____.
26.12 Expand AFM.
26.13 Acronym for SPL
26.14 Expand STM.
26.15 The wavelength of laser beam from KrF is equal to _____.
26.16 _____ is the wavelength emitted by ArF laser source.
26.17 _____ and _____ laser beams are used to obtain the lithography pattern with a size of 150 nm.
26.18 In particle beam lithography, the resolution depends on larger _____ and _____.
26.19 The different modes of SPL mechanical method are _____, _____ and _____.
26.20 The AFM tip is treated as _____ and the molecules are treated as _____ to write the pattern on substrate.
26.21 Expand PVD, CVD and PECVD.
26.22 The sputtering is achieved by _____ and _____ method.
26.23 In CVD, the inert gas like_____, is used as a carrier gas.
26.24 Expand MBE, MOVPE and LPE.
26.25 Molecular beam epitaxy method requires ultra high vacuum. (True/False)

The McGraw·Hill Companies

26.26 The high vacuum required in the reaction chamber is not less than _____ m bar.
26.27 Expand CAR.
26.28 Expand BFM.
26.29 Expand RHEED and LEED.
26.30 Expand AES and MBES.
26.31 Gold nanoparticles are synthesised by the reaction of _____ acid.
26.32 The chemical reaction taking place in the sol gel process is _____ and _____.
26.33 The process of coating a metal on template is known as _____.
26.34 The Hall petch relation is equal to

a) $\sigma_y = \sigma_0 + \frac{k^2}{\sqrt{d}}$ b) $\sigma_y = \sigma_0 + \frac{k}{\sqrt{d}}$

c) $\sigma_y = \sigma_0 + \frac{k^2}{\sqrt{2d}}$ d) $\sigma_y = \sigma_0 + \frac{k}{\sqrt{2d}}$

26.35 The relation between hardness and grain size is

a) $H = H_i - \frac{k}{\sqrt{d}}$ b) $H = H_i + \frac{k}{\sqrt{2d}}$

c) $H = H_i + \frac{k}{\sqrt{d}}$ d) $H = H_i + \frac{k}{\sqrt{3d}}$

KEY

26.1 10^{-9} m.
26.2 Quantum dots
26.3 Bottom up
26.4 Top down
26.5 Top down
26.6 Cryogenic liquids
26.7 Softer
26.8 Grinding
26.9 Lithography
26.10 Resist
26.11 Photons, particle beams
26.12 Atomic force microscopy
26.13 Scanning probe lithography
26.14 Scanning tunneling microscope
26.15 248 nm
26.16 193 nm
26.17 KrF, ArF
26.18 Mass, particle velocity
26.19 Scratching, pick-up, pick-down
26.20 Pen, ink
26.21 Physical vapour deposition, chemical vapour deposition, plasma enhanced chemical vapour deposition
26.22 dc voltage, RF voltage
26.23 Ar or Ne
26.24 Molecular beam epitaxy, Metal organic vapour phase epitaxy and Liquid phase epitaxy
26.25 True
26.26 10^{-11}
26.27 Continuous azimutal rotation
26.28 Beam flux monitor
26.29 Reflection high energy electron diffraction, Low energy electron diffraction
26.30 Auger electron spectroscopy and modulated beam mass spectroscopy

The McGraw·Hill Companies

26.31 Chloroauric
26.32 Hydrolysis, Condensation
26.33 Electroplating
26.34 b)
26.35 c)

Short Questions

26.1 What is meant by nano?
26.2 Explain Feynman,s statement.
26.3 What are the classifications of nanomaterials?
26.4 What is the dimension of quanter dot
26.5 What is the grain size range of nanostructure materials?
26.6 How the nano materials are clarrified?
26.7 Explain the difference between top-down and bottom-up approach needed for nano synthesis.
26.8 Compare the relative merits of chemical, physical, biological and hybrid methods for the preparation of nanomaterials.
26.9 What is the principle behind mechanical milling?
26.10 Explain the principle behind lithography.
26.11 Mention the different types of lithography.
26.12 What is meant by photolithography?
26.13 How to pattern in dip-pin lithography?
26.14 Compare the relative merits of the usage of photons and particles in lithography.
26.15 Explain the principle behind vapour phase deposition.
26.16 What is meant by chemical vapour deposition?
26.17 What is meant by sputtering?
26.18 Explain the difference between glow discharge and RF sputtering.
26.19 What is meant by plasma enhanced CVD?
26.20 What is the difference between MBE and MOVPE?
26.21 Explain the principle behind MOVPE.
26.22 What is meant by bubblers?
26.23 Explain why MOVPE mehot does not require vacce is sample chapter.
26.24 Mention the applications of MOVPE meet over the methods.
26.25 How is LPE used to obtain nanowire or nanorods?
26.26 What is meant by colloids?
26.27 How is the template used to obtain nanowire or nanorods?
26.28 Explain the difference between self-assembly and self-organizations.
26.29 What is meant by nanosafety?
26.30 How can we store the nanoparticles?
26.31 Why do nanostructured particles find potential applications?
26.32 What is meant by surface induced effect?
26.33 How are nanostructured particles used for health applications?
26.34 What is the relation between properties and applications of nanoparticles?

Descriptive Questions

26.1 Describe with neat sketch how the nanoparticles are prepared employing the top-down methods, namely, milling and lithography.

26.2 Explain with suitable diagram, the synthesis of nanostructured materials using PVD and also reveal the applications by comparing with PECVD.

26.3 What is meant by MBE? Explain with neat sketch the principle, working and synthesis of nanoparticles using MBE, MOVPE and LPE.

26.4 How is the chemical method different from physical method of synthesis of nanoparticles? Explain the synthesis of nanoparticles using colloids and sol gel method.

26.5 Explain with suitable example, the synthesis of nanostructured materials employing self-assembly, self-organisation and template-based methods.

26.6 Discuss in detail how the mechanical and magnetic properties of nanomaterials vary with particle size.

26.7 Describe a technique to synthesize nanophase materials. Discuss their applications in various fields.

Chapter

Nanodevices

27

OBJECTIVES

- To discuss the importance of nanostructured materials such as nanomagnets, nano semiconducting and carbon nanotubes
- To understand the structure of nanomagnets and its properties such as GMR, CMR, TMR, etc.
- To explore the possible applications of nanomagnetic materials such as data storage devices
- To study the features of nanostructured semiconducting materials and their device applications
- To discuss the role of the organic semiconducting materials and their applications
- To study the preparation, methods, properties and applications of carbon nanotubes

27.1 INTRODUCTION

Even though magnetism is one of the oldest and fascinating fields of science, nanomagnetism is the forefront field in the nanotechnology revolution. The invention of nanoscale phenomenon and the magnetic structures have opened up new materials and technologies. The modern technological devices like electrical power generators, transformers, electrical motors, computers, sound and video reproduction systems, MEMS and NEMS, etc., depend on the magnetism and magnetic materials. The microstructural and magnetic properties are the important features to be considered for a nanomagnetic materials. The above features depend on the particle size distributions, chemical compositions, crystal defects, etc. Nanomagnetism is the study of the ferromagnetic behaviour, when their domains are geometrically restricted in atleast one dimension. The nanomagnetic materials find potential applications in technology and industries like the generation and distribution of electrical power, biomedical applications, sensors and computers.

In this section, the principles of nanomagnetism, synthesis of nanomagnetic materials, properties and their applications are given in brief.

27.2 NANOMAGNETS

We know that the magnetic materials are different categories, namely, diamagnetic, paramagnetic, ferromagnetic, anti-ferromagnetic and ferrimagnetic materials. The diamagnetism is the fundamental property of atoms or molecules. There is no permanent dipole moment for diamagnetic materials. The dipoles are oriented such that the resultant dipole moment tends to be zero. When an external field is applied, the individual dipoles are related and produce an induced dipole moment. This induced dipole opposes the applied field. Due to the magnetic interaction between any two dipoles, they try to align themselves. However, even at small thermal agitation at room temperature, it disturbs the alignment. Thus, the dipoles are randomly oriented in such a way that the resultant dipole moment is zero. In case of paramagnetic materials, the magnetic dipoles are randomly oriented. On the other hand, in ferromagnetic materials, the magnetic dipoles are equal and aligned parallel to each other. When a small amount of external magnetic field is applied, a large value of magnetisation is produced. The ferromagnetic material exhibits permanent magnetic dipole and hence, it produces a spontaneous magnetisation. The details of anti- ferromagnetic and ferric magnetic materials are given in detail in Chapter 20.

Generally, the ferromagnetic materials exhibit hysteresis curve as shown in Fig. 27.1. We know that remanence or *retentivity* is the remaining magnetisation when the field is reduced to zero from the saturation field. The coercivity is the field required to bring magnetisation to zero from resonance.

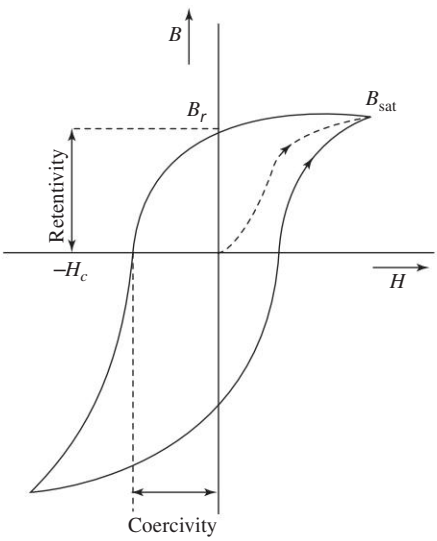


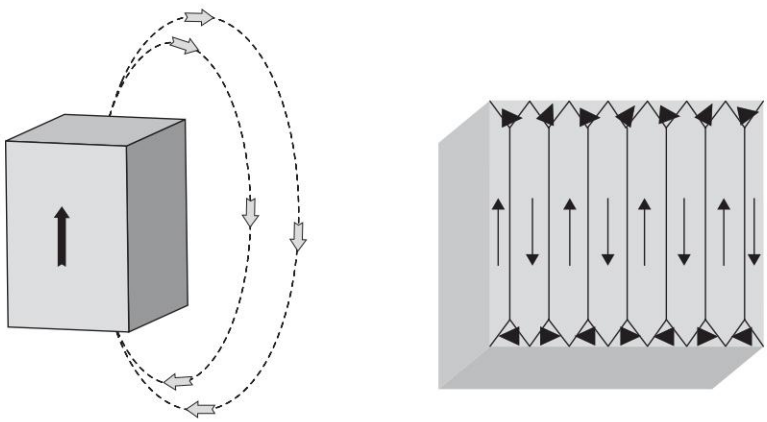
Fig. 27.1 *Hysteresis loop*

We know that a ferromagnetic material consists of tiny individual magnetic domains. A schematic representation of the single and multi-domain ferromagnetic particles is shown in Fig. 27.2. The magnetic moments of all the components of atoms or molecules are pointing in the same directions. When an external magnetic

The McGraw·Hill Companies

field is applied, the magnetic domains align themselves and the net resultant magnetic moment is zero. Depending on the direction of the magnetisation by an applied field, there are two types of magnetic materials, namely, soft and hard magnetic materials. The soft magnetic materials exhibit low coercivity and high retentivity, while the hard magnetic materials exhibit high coercivity and low retentivity.

The study of the behaviour of the ferromagnetic materials when the domains are aligned in one dimension is known as *nanomagnetism*. The magnetic properties of the materials are explained based on the magnetic dipole, which is at the nanoscale length. The nanomagnetism is not a new invention, rather it is the production of magnetic materials under controlled condition. Nanomagnets have been synthesised in different sizes, shapes and materials. The different forms of nanomagnets are zero, 1D and 2D dimension, while the shapes are dots, pillars, disks, rods, chains, etc.



(a) Single-domain (b) Multi-domain

Fig. 27.2 *Ferromagnetic particles- magnetic domain*

27.3 CLASSIFICATIONS OF NANOMAGNETIC MATERIALS

Nanomagnetic materials are classified in the following ways based on the process method:

(i) Particulate
(ii) Geometrical, and
(iii) Layered

Let us discuss the above nanomagnets in the following sections in detail:

27.3.1 Particulate Nanomagnets

Particulate nanomagnets are granular solids with one or more phases having magnetic properties. Particulate nanomagnets are generally prepared employing the rapid

solidification process technique. In rapid solidification, the materials which are in the molten state (melt) are rapidly transferred into the cold block. The rate of cooling is in the order of 10 K s^{-1} and hence, nanophase materials are produced. Generally, one can produce the nanophase materials employing two different methods using the rapid solid technique. In the first method, the amorphous phase material is obtained using the rapid solidification. Then, the material is subjected into necessary heat treatment to obtain the nanostructure. However, in the second method, the molten melt is first transferred into the block and then it is cooled slowly to obtain the nanostructures. During the above process, the nanomagnetic materials properties can be changed from softest (i.e., low coercivity and anisotropy) to the hard magnetic state (i.e., high coercively and anisotropy).

The properties of the nanostructured materials depend on the interplay between the exchange and anisotropy energies. The existence of the exchange forces causes the magnetic moments of the neighbouring grains to lie parallel to each other by overcoming the intrinsic properties of the individual grains. A schematic representation of the nanostructured soft ferromagnet is shown in Fig. 27.3. The small magnetic grains are represented by a square while the direction of magnetisation in each grain is shown by a arrows. The direction of magnetisation depends on the exchange length and is given by,

$$L_{ex} = \frac{A}{K_1} \tag{27.1}$$

where A is the exchange stiffness and K_1 the anisotropy constant.

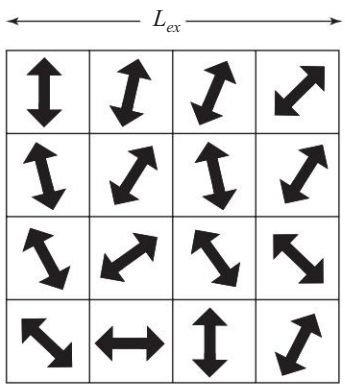


Fig. 27.3 *Nanostructured soft ferromagnet*

It is clear from the Fig. 27.3. that the direction of magnetisation changes from nano grains to nano grains. When the local anisotropy is very strong, the magnetisation is well aligned, while for weak anisotropy it leads to a change in magnetisation with respect to the direction. The spin alignment of the nanomagnetic materials for weak and strong local anisotropy energy in interaction with the exchange energy is shown in Fig. 27.4. Thus, the direction of magnetisation changes gradually by overcoming the local anisotropy. The resultant exchange length which exists in nanomagnets reveals the soft nanophase magnetic nature.

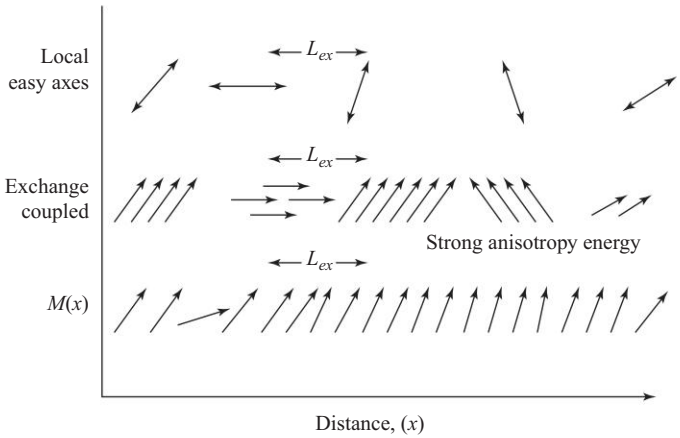


Fig. 27.4 *Nanomagnets—exchange length*

Some of the examples for the particulate nanomagnets are $Fe_{73.5}$ $B_9Si_{13.5}$ Nb_3Cu, Nd_2 $Fe_{14}B$, CaCrTa, etc.

Advanced nanomagnetic materials are used for manufacturing the devices like cordless power tools and miniaturised earphones on portable audio devices. Similarly, high-efficiency electromotors are under progress employing nanomagnets. In order to store more energy, a high energy product $(BH)_{max}$ nanomagnet is required, which facilitates to meet the required applications. The microstructure and the hysteresis loop of the $Nd_2Fe_{14}B$ and FeCu or CoCu doped $Nd_2F_{14}B$ nanomagmetic is shown in Fig. 27.5. The highest value of energy product required for NdFeB type nanomagnet is 450 KJ m^{-3}.

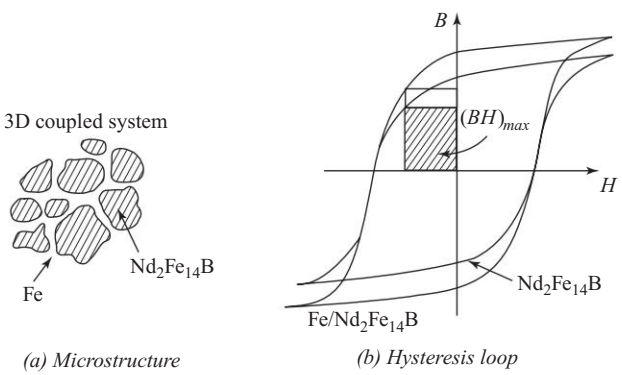


Fig. 27.5 *Nano granular solidas—$Nd_2Fe_{14}B$*

A nanoparticulate magnetic film is grown on a suitable seeding layer as shown in Fig. 27.6. The seeding layer with close pakced direction is first grown with small grain size using CoCr. The functional layer i.e., CoCrTa is grown over the textured seed layer. The functionalised layer is grown with nanostructured phase. These films are used for recording media.

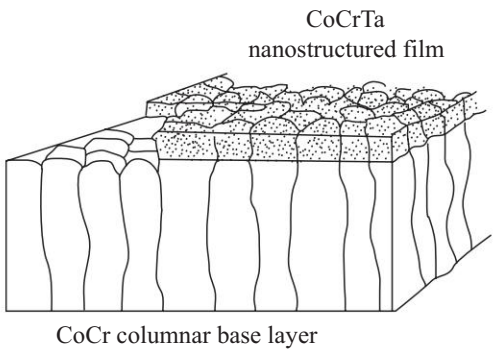


Fig. 27.6 *CoCrTa nanostructures on CoCr base layer*

27.3.2 Geometrical Nanomagnets

In nanotechnology, the self-assembly is one of the techniques used to obtain the nanostructured materials on a base layer. For example, in lithography, a mask is formed on a polymer resist. A pattern of magnetic islands is retained after etching or lift off. Similarly, nanoparticles can be self-assembled on a template which consists of base materials. When nanoparticles are self-assembled through surfactant and ligand structures, it covers larger area. The ligand lengths control the spacing between the nanoparticles. The energy products of self assembled nanoparticles are very high. The self assembled nano permanent magnets find potential applications in micro and nano electro mechanical systems (MEMS and NEMS) and data storages. In view of the existence of excellent nano scale magnetic properties, these are geometrical permanent nanomagnetic materials.

In geometrical permanent nanomagnetic materials, the shape anisotropy is such that it produces the hard magnetic properties. The magnetic particles are elongated perpendicular to the direction of the applied field and a normal structure along the direction of the field. Due to the shape anisotropy, the geometrical nanomagnets like Alnico are not important as that of NdFeB nanomagnets. Recently, the nanomagnetic materials are coated with biological macromolecules by nano encapsulation for drug delivery and cancer treatments. Thus, the nanomagnetic materials are used for smart diagnostic systems.

27.4 MAGNETO RESISTANCES

Magneto resistance is the property of a material to change the value of its electrical resistances under the influence of an external magnetic field. This effect is called *Ordinary Magneto Resistance (OMR)*. It was first discovered by William Thomson in 1856. He failed to lower the electrical resistance beyond 5%. In recent years, the electrical resistance has been reduced beyond 5% employing new materials, namely, *Giant Magneto Resistance (GMR)*, Colossal Magneto Resistance (CMR), and by changing the sample magnetisation in magnitude or direction known as

Tunneling Magneto Resistance (TMR). With the development of nanomagnetic materials with granular solids or multilayer, the electrical resistivity has been reduced considerably. In the following section, a brief description about the above three effects are given.

27.4.1 Ordinary Magneto Resistance

The effects of Ordinary Magneto Resistance (OMR) in a material can be explained by considering the corbino disc. The corbino disc consists of a conducting annual with perfectly conducting rims as shown in Fig. 27.7. The battery drives a current between the rims in the absence of an external magnetic field. When a magnetic field is applied, the Lorenz force drives a circular component of current. Thus, it increases the resistance between the inner and outer rims. The increase in resistance due to the magnetic field is known as magneto resistance.

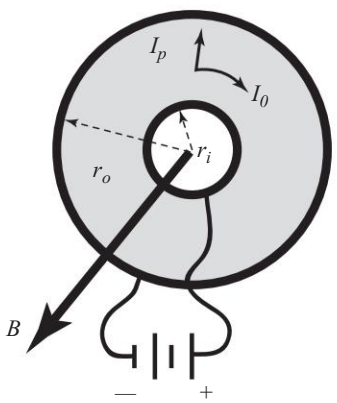
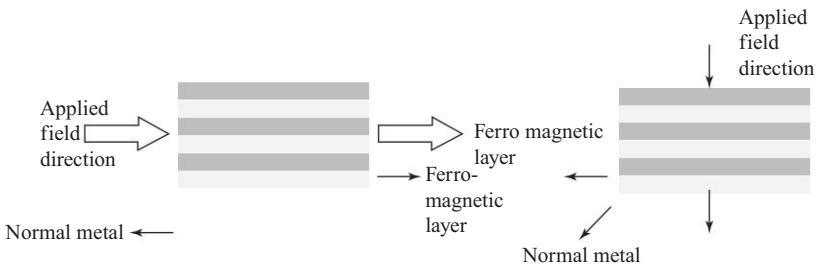


Fig. 27.7 *Magneto resistance*

27.4.2 Giant Magneto Resistance

Giant Magneto Resistance (GMR) is the phenomena in which a significant change (>5%) in electrical resistance occurs due to the change in magnetisation direction or magnitude. The new class of materials which exhibits GMR consists of normal metal and ferromagnetic layers. The general structure of the GMR is shown in Fig. 27.8. The electrical conductivity depends on the orientation of the magnetic field applied to the GMR materials. One can apply the external current in two different ways. i.e., current is parallel to the plane of layers (CIP) or current is perpendicular to the plane of layers (CPP) as shown in Fig. 27.8. In CIP mode, the materials will have high resistivity, while in CPP mode, they have low resistivity. Hence, CIP geometry of the GMR offers higher resistance than CPP geometry. Some of the GMR materials which exhibit this property are transition metals (example, iron) and noble metals (example, silver).



(a) CIP in plane *(b) CPP current in perpendicular plane*

Fig. 27.8 *Giant magneto resistance*

27.4.3 Tunneling Magneto Resistance

Tunneling Magneto Resistance (TMR) is an another form of magneto resistance device which relies on electron tunneling through oxide barrier between two ferromagnets. The schematic structure of a spin valve is shown in Fig. 27.9. The layer 1 is the antiferromagnetic layer, i.e., FeMn, while the layer 2 is the pinned soft magnetic layer, i.e., NiFe. The third layer, a copper spacer layer separates the pinned soft magnetic layer 2 and the free soft magnetic layer 4 (i.e., NiFe). Thus, the two ferromagnetic layers 2 and 4 are separated by an insulating tunnel barrier. When the spin flip scattering length is greater than the barrier width, the electrons will tunnel across the barrier. Therefore, when a large density of states is available on both sides of the barrier, it gives a low resistance across the barrier.

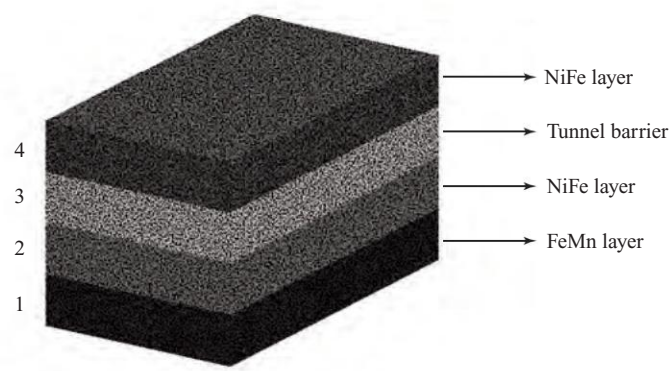


Fig. 27.9 *Structure of a spin value*

Recently, very thin insulating layer has been used as a tunnel barrier between the two ferromagnets. Generally, the insulating layers are fabricated by growing an Al layer first and thus oxidising the same into Al_2O_3. Therefore, by growing a suitable oxide layer which is antiferromagnetic, one can destroy the spin polarisation. The transistion metal oxides are used for the formation of tunnel functions, which gives the high transistion temperatures. The overall resistance of the TMR device is much higher

The McGraw·Hill Companies

than GMR. Further, the two magnetic materials in TMR are not coupled magnetically and hence, the junction remains in one configuration. The TMR materials are good candidate for Magnetic Random Access Memory (MRAM).

27.5 PROBING NANOMAGNETIC MATERIALS

We know that scanning electron microscopes (SEM) are used to obtain the topology of the solid samples. Nanomagnetic materials play an important role in scanning probe microscope. In atomic force microscope (AFM), a tip is used either in contact or noncontact mode to obtain the topographical information of the sample. In non contact mode, the AFM tip is not in contact with the sample, while in contact mode, the tip is in contact with the sample. A magnetic force microscope (MFM), which is an analog to AFM, consists of the tip coated with soft or hard (i.e., FeBBic or CoCr) magnetic materials.

In contact mode MFM, the tip scan takes a line scan to obtain topographical information and thus, rescan the sample. Thus, it gives a complete magnetic contrast of the sample. When the magnetic top is in contact with the sample, the Van der Waals forces dominate over the magnetic forces, and this results in the patterning of the topography of the sample. When the tip is at far distance, a magnetic interaction between the tip dipole moment and the stray field from the sample are obtained. The principle of working of MFM is shown in Fig. 27.10.

An external magnetic field is used to magnetise the sample which gives rise to stray fields above the sample plane in the vertically upward direction. The tip coated with the magnetic materials is magnetised in the downward direction. As a result, the tip will notice an attraction at *A* and repulsion at *B* leading to a contrast magnetic topographical image of the sample. The MFM images are used to study the magnetic tracks and data bits of the sample.

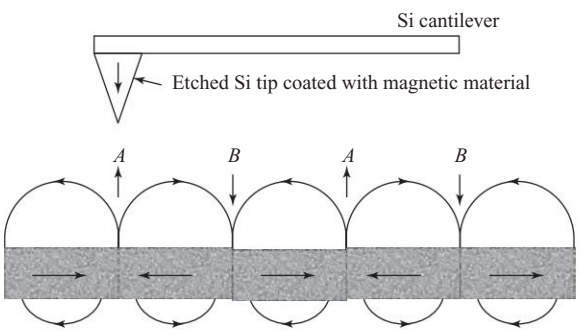


Fig. 27.10 *Magnetic force microscope—principle*

27.6 NANOMAGNETISM IN TECHNOLOGY

Some of the application of nanomagnetism in technology are given below:

(a) Storage The most important application of nanomagnetism is in memory technology. The advent of nanomagnetism in memory technology leads to cost reduction with the development of ever faster, more compact and less power consuming memory systems with greater storage capacity. The benefit of nanotechnology results in decrease

in physical size of each bit leading to an increase in storage density. One can generate a new generation magnetic hard disk employing the nanomagnets wherein the size of data bit has been reduced to less than 300 nm × 15 nm from the original 10 μm × 0.7 μm. A schematic representation of the nanomagnetic materials which demonstrate the storage capacity on a nano bit of different surface area is shown in Fig. 27.11. Very recently, it was demonstrated by IBM computers that the nanomagnetic compact disc (nano CD) can store data equivalent to 1000 CD's capacity (i.e, 700 MB × 1000 = 700 GB).

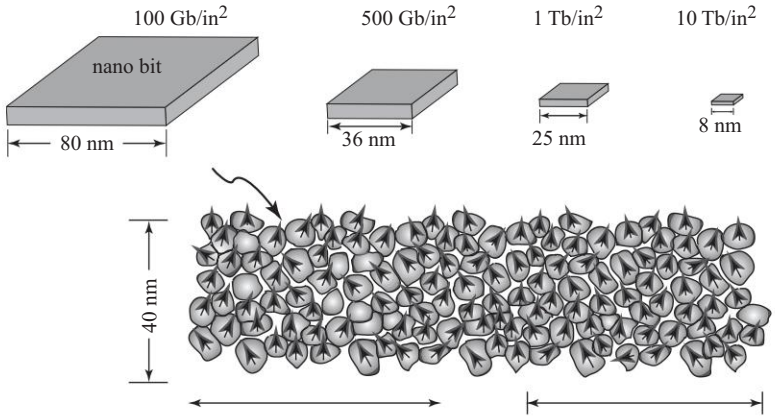


Fig. 27.11 *Storage capacity—nanomagnets*

(b) Recording Head During the magnetic recording, a change in the direction of magnetisation is stored in the storage medium. The data stored on the magnetic bit is sensed by a magnetic read/ write head based on the magnetoresistive effect. Thus, the storage media requires a very high coercive force and saturation magnetisation, while the read/ write head requires high data rates with high permeability. These constraints are overcomed by nanomagnetic materials particularly GMR materials.

We know that GMR nanomagnets are used as recording head. In GMR, the material size required per bit has been dramatically reduced in addition to the reduction in the memory access time from milliseconds to nanoseconds. Due to the shrinkage in the dimension of the GMR materials, the signal to noise ratio is reduced considerably.

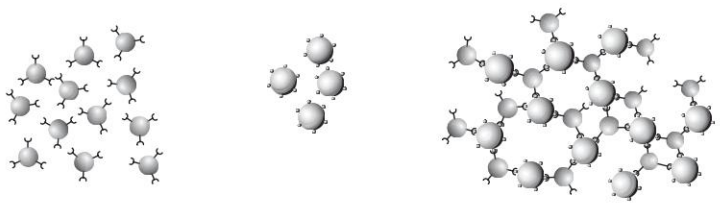
(c) High Power Magnets We know that a decrease in grain size and an increase in surface area of the grains leads to an increase in coercivity and saturation magnetisation reduces. This in turn increases the magnetic strength of the materials. Hence, nanomagnetic materials possess fascinating magnetic properties giving rise to potential applications like quiter submarines, automobile alternators, power generators, ultra sensitive analytical instruments, etc.,

(d) High Sensitive Sensors Sensors are used to measure the change in parameters like electrical resistivity, chemical activity, magnetic permeability, thermal conductivity, capacitance, etc. All the parameters depend on the microstructure, i.e., grain size of the materials used for sensors. Thus, the sensors are very sensitive to the change in the above parameters. For example, when a sensor notices a change in environment either in chemical or physical or mechanical parameters, it is detected by the sensors. These characteristic changes on the sensors are exploited suitably and measured. For example, sensor which is made up of zirconium oxide is used to detect the presence of carbon monoxide, known

The McGraw·Hill Companies

as *carbon dioxide sensor*. When the sensor environment detects the carbon monoxide, the oxygen atoms in zirconium oxide react with the carbon in carbon monoxide and reduce it to zirconium oxide. The above chemical reaction changes to sensors characteristics, i.e., conductance/resistance and capacitance of the sensors. The change in the conductance or resistance is used to explore the presence of carbon monoxide employing suitable electric devices. Thus, the sensitivity of the sensors depend on the nano grain size and the surface area of the nano grains. The potential applications of sensors are used to detect smoke, the performance of automobile engine, and ice on aircraft wings, etc.

(e) Biomedical Nanomagnetic materials find wide application in biomedicine. Nanomagnetic particles with size ranges from 10 to 500 nm have been used as contrast agents to obtain the magnetic resonance imaging (MRI). These nanomagnetic materials have been coated along with suitable chemically neutral materials to prevent the reaction with body fluids. Other important applications of nanomagnetic materials are drug-delivery. The nanomagnetic materials are first laced with drug molecules. The drug molecules coated nanoparticles travel into different organs. The nanoparticles are steered by means of external magnetic field gradients to reach the desired parts of the human body. This technique is known as *targeted drug delivery technique* which finds applications like cancer treatment. The schematic representation of the creation of a viral nanosensor is shown in Fig. 27.12. The magnetic nanoparticles are covered with antibodies. These particles are injected into the body, which in turn interact with the body fluid or tissue. The nanoparticles coated with antibodies bind to the Virus and create the clusters as shown in Fig. 27.12. These clusters are identified when the body is scanned through MRI or NMR scans.



(a) Magnetic nanoparticles Covered with antibodies

(b) Viruses in body fluid on tissue

(c) Nanoparticle bind to viruses, creating clusters visible on a MRI or NMR scan

Fig. 27.12 *Viral nanosensor*

In view of the excellent magnetic properties, the nanomagnetic particles show potential application in different areas like storage, sound record / reading heads, biomedical, sensors and electronic industries. Thus, the opportunities as well as the challenges in nanomagnetic materials are every increasing.

27.7 APPLICATIONS OF SEMICONDUCTOR NANOSTRUCTURES AND DEVICES

The applications of nanostructured materials and particles are ever increasing in almost all areas of technology and industries. The semiconductor nanostructured materials are useful for the development of organic semiconductor products. The nanostructured semiconductors are used to produce the devices like injection lasers, quantum cascade

lasers, optical memories, electronic applications and coulomb blockade devices. Similarly, the other products which are developed are organic field effect transistors, organic light emitting devices, organic photovoltaics, organic photoconductors and laser printers.

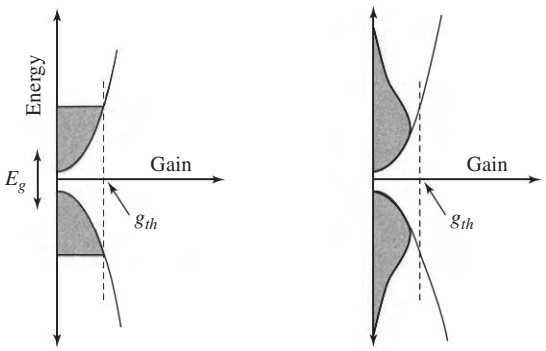
In this chapter, the future applications of semiconductor nanostructured materials and devices are discussed in brief.

27.8 APPLICATIONS OF SEMICONDUCTOR NANOSTRUCTURE

The practical applications of semiconductor nanostructures are discussed and demonstrated briefly.

27.8.1 Injection Lasers

Semiconductor injection laser system is used to convert electrical energy into light energy with high efficiency. It has a long period of operating lifetimes. Generally, injection lasers are used in data transmission, storage, printing and medicine. A laser is an optical oscillator and its working is based on the principle of population inversion. The population inversion is achieved in a forward biased *p-i-n* structure where large number of electrons and holes are injected into the intrinsic region. Thus, it creates a sufficient density of electrons and holes. The required gain is achieved by the injection of current. The gain is proportional to the density of the occupied states and it will increase with increase in carrier density. When a nanostructured material is used in the emitting region of a laser, it leads to significant improvements in its performance. The limitation of the semiconductor lasers, at are overcome employing a nanostructured semiconductor by the modification of the density of the states with a nanostructure. The electron and hole distribution in the conduction and valence bands of a bulk semiconductor laser at absolute zero and non-zero temperature is shown in Fig. 27.13.



(a) Absolute temperature (b) Non-zero temperature

Fig. 27.13 *Bulk semiconductor laser*

The McGraw·Hill Companies

When the energies are low, the density of the state increases with respect to the higher energies. Thus, it reduces the threshold current and its sensitivity with increase in temperature. The energy of the carrier remains same at all temperatures in quantum dot as it consists of both only electron and hole states. Therefore, there is no state available for excitation. In such a case, it emits laser with very low and temperature insensitive threshold current.

The potential applications of nanostructures have been undertaken after the theoretical analysis of ideal systems. The dependence of current density with variation in the gain for different nanostructured dimensionalities and the temperature dependence of threshold current are given in Fig. 27.14.

The nanostructured materials with different dimensionalities reveal an increase in current as shown in Fig. 27.14a. It is also evident from Fig. 27.14b, that the threshold current of different nanostructure dimensionalities depends on temperature. The temperature dependence on the threshold current density of a semiconductor laser is given as,

$$J_{th} = J^{\circ}_{th} \exp (T/T_o), \tag{27.2}$$

where J°_{th} is the threshold current density at 0°C, and T_o a parameter which determines the temperature sensitivity.

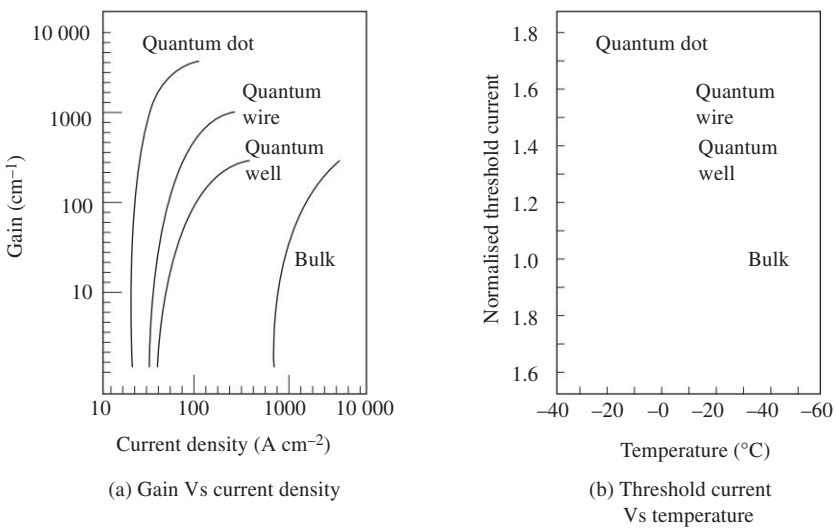


Fig. 27.14 *Variation of gain and threshold current—injection laser*

From Fig. 27.14b, it is clear that as the dimensionality decreases from bulk to quantum dots, the threshold current density decreases. For example, the threshold current densities of 1050, 380, 140 and 45 A cm^{-2} have been obtained for bulk, quantum wells, quantum wire and quantum dot respectively. The quantum dot laser has been made based on self assembled dots. These lasers offer number of advantages like low threshold current density, high temperature stability, high maximum modulation frequency, etc. Similarly, the quantum wire laser has been fabricated with limited

performance. A typical InAs quantum dot laser operates at the telecommunication band of 1.3 μm with a prospect for device operation at 1.55 μm band. In recent years, the data storage devices like CD/DVD and fiber optic data transmission are produced employing semiconductor laser band quantum wells. The self assembled quantum dots are used for the laser productions.

27.8.2 Quantum Cascade Laser

Conventional semiconductor lasers are based on interband transition which emits electromagnetic radiation through the recombination of electron hole pairs and are called as *bipolar*. In bulk semiconductor crystals, the two energy bands are separated by an energy band gap in which there are no permitted states available for electrons to occupy. Interband transitions emit a single photon during the recombination of an electron and hole respectively from the conduction and the valence band. In this laser, the spectral range changes from the near infrared to the violet, i.e., from 2.0 to 0.4 mm. It is more difficult to obtain a laser in the operating range of region between 2 to 100 mm due to the small band gap energy.

Quantum cascade laser, a new type of semiconductor laser has been developed, which emits in the infrared region of the electromagnetic spectrum. Quantum cascade lasers are intraband unipolar devices. The laser emission is achieved employing intraband transition in a repeated stack of semiconductor superlattices. In quantum cascade structures, electrons undergo intraband transition between confined quantum well conduction band states and hence, photons are emitted. A typical band structure of a quantum cascade laser is shown in Fig. 27.15. Quantum cascade lasers are typically based upon a three level system. However, in a unipolar quantum cascade laser, the electron which undergoes an intraband transition emits photon in one period of the super lattice. Then, the emitted photons can tunnel into the next period of the structure which results in the emission of another photon. Thus, the process of a single electron causes the emission of multiple photons when it travels through the quantum cascade laser structure. Hence, this laser is named as *cascade laser*. The gain provided by a single stage is not sufficient to overcome the losses. Therefore, the electrons are cascaded down with multiple stages of the cascade structure to over come the losses.

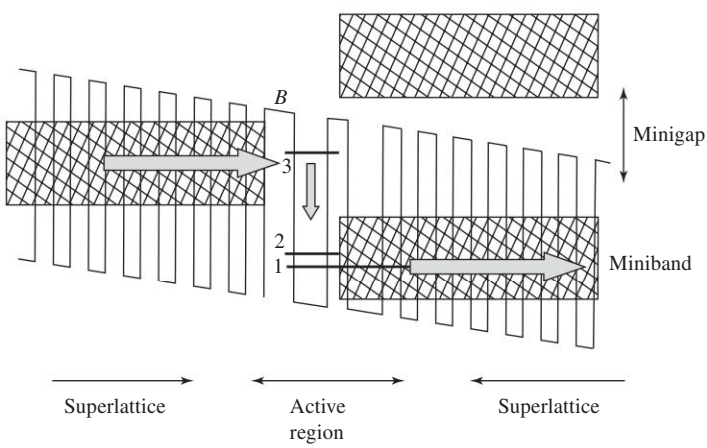


Fig. 27.15 *Quantum cascade laser-band structure*

The McGraw·Hill Companies

The wavelength of laser emission ranges from 3.5 to 106 μm has been obtained employing number of quantum cascade lasers. A continuous room temperature operation has been demonstrated at 6 and 9.1 μm, while the laser operations are under pulsed mode. When the laser is operated continuously at room temperature, it generates heat to the large threshold current densities. Due to this in-efficiency, the electrons are being wasted and get excited out from the upper lasing state.

The high optical power output, tuning range and room temperature operation makes quantum cascade lasers as a useful one for spectroscopic applications such as gas monitoring pollutants monitoring in atmosphere, free space communication systems and medical diagnostics such as breath analyses.

27.8.3 Optical Memories

In recent years, quantum dots are used as a high density memory system to store the charges. One can use both optical and electrical reading/writing to store and restore the charges. Thus, it provides a solid evidence to replace the electronic random access memory (RAM) used in the present day computers. The multiple dot self-assembled quantum dots layer increases the storage density area to tenfold as that of the self-assembled quantum dots storage area of 1×10^{11} cm^{-2}. For memory applications, the homogeneous broadening of the quantum dots facilitate to store very high energy and also turned to the laser light for emission energy.

One can use an alternate scheme of type II quantum dots, i.e., GaSb dots grown on GaAS as shown in Fig. 27.16. It consist of one carrier, i.e., positive charge, which repels the fraction of hole from a two dimensional hole gas (2DHG) which is parallel to the quantum dot layer.

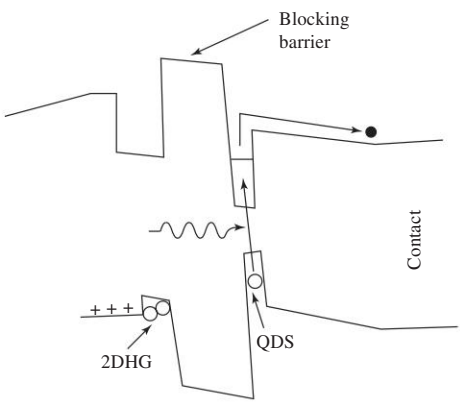


Fig. 27.16 *Quantum dot memory device-band structure*

The band structure of InAs self-assembled quantum dots memory device is shown in Fig. 27.16. When an electric field is applied, the electrons escape from the dots while the holes are prevented from the escape due to the large barrier placed to the left of the dot as shown in Fig. 27.16. It is possible to store charge in the dots for more than 8 h at a temperature of 145 K. At higher temperature, the storage time decreases, where the holes may be thermally excited out of the dots. The homogeneous broadening of the dot transition appears to prevent the optical writing to individual dots due to the thermal excitation and temperature influence.

27.8.4 Impact of a Nanotechnology on Conventional Electronics

According to Moore's law, which states that the power calculation has been increased in the revolution of information, the development of transistors in an integrated circuit or microprocessor has been increased tremendously. When an electron and a hole are trapped in a quantum dot, a recombination radiation is possible within a time scale of less than 1 ns. This is too short when compared with any practical memory systems. The necessary electrons and holes are created in quantum dots by optical excitation.

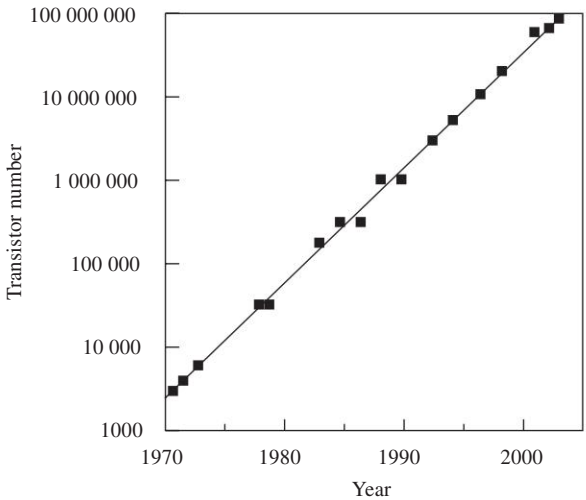


Fig. 27.17 *Moore's law- Growth of transistors*

The doubling in the number of transistors in a microprocessor for a span of every 26 month is shown in Fig. 27.17. At the same time, it also demonstrates the reduction in component size. It is clear from Fig. 27.17 that the increase in the transistors number increases the operating frequency of the microprocessor from 1 MHz to 3 GHz. It is evident from the Moore's law that the transistors are soon entering into the nano scale. At present, the fabrication and operation of the electronic components will be scaled down from larger sizes to nano scale size. Moore's law will continue for the next two decades with the inclusion of application of nanotechnology concepts.

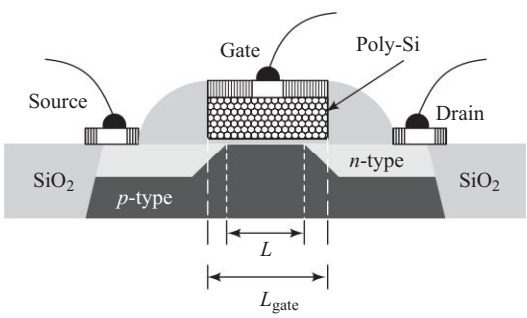


Fig. 27.18 *Metal oxide semiconductor field effect transistor*

The McGraw·Hill Companies

The metal oxide semiconductor field effect transistor (MOSFET) is used in microprocessor as shown in Fig. 27.18. The current is carried out between the drain and source contacts by the free carriers, either electrons or holes respectively, in the *n*-type (NMOS) or *p*-type (PMOS) device in the channel. A gate which is placed above the channel is the third contact. The flow of current between the gate and the channel is prevented by a thin SiO_2 insulating layer. Current conduction occurs between the source and the drain. One can obtained the basic switching action of the transistor by changing the gate voltage. The insulating SiO_2 is used to isolate from the neighboring devices. It is also used to reduce the size of the switching voltage. The metal gate of a MOSFET is replaced by polysilicon. A reduction in the gate length reduces the MOSFET size and hence, one can reduce the gate length if Moore's law holds until 2018. One can use the lithographic process to reduce the size of a given wavelength. The immersion lithography technique is used to decrease the size of the MOSFET. According to International Road map for semiconductors 2003, a microprocessor will contain approximately 9 billion transistors with 10 nm gate lengths at an operating frequency of 28 GHz.

27.8.5 Coulomb Blockade Devices

The schematic diagram of the structure of Coulomb blockade oscillations for a small bias voltage applied between the left and right reservoir is shown in Fig. 27.19. The applied gate voltage controls the current flow of the reservoir as indicated in Fig. 27.19.

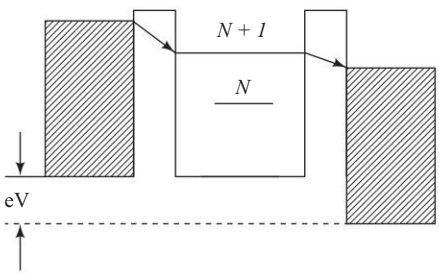


Fig. 27.19 *Coulomb blockade oscillations – Quantum dot*

This is similar to the transistor action. The transfer of current between the reservoir takes place due to the transport of one electron and the quantum dot. Therefore, the *Coulomb blockade device* is known as *single electron transistor device*. It is used to increase the resistance at a small bias voltage which comprises of at least one low capacitance tunnel junction. Further, it provides the possibility to be used as a single electron memory cell. A simple way of writing is the transfer of an additional electron. Then, it leads to a modification in the gate voltage due to the presence of additional electrons.

Coulomb blockade effects can be observed in any conducting system of suitably small size. However, silicon-based devices are required for good compatibility with existing electronics. Once can notice the Coulomb blockade when the charging energy is greater than the thermal energy. The silicon based devices with dot size of 10 nm operate at room temperature. The devices which are fabricated well below the size using the lithography can be used to operate at low temperatures. The devices fabricated using the above techniques have been successfully applied for memory operations.

27.9 ORGANIC SEMICONDUCTOR MATERIALS DEVICES

Some of the synthetic metals which are used as organic semiconductor materials are polyacetylene (PA), polydiocatylene (PDA), polyaniline (PANi) and poly (3, 4–ethylene dioxythiophene (PEDOT). These materials are organic semiconductors in their pristine state. The above synthetic metals become quasi-metallic when a dopant is added. In order to obtain a highly conductive PEDOT/ PSS complex, PEDOT is acid doped with polystyrene sulfuric acid (PSS).

27.9.1 Organic Field Effect Transistors

The organic field effect transistors (FET) are used to produce low performance integrated circuits based on organic materials. One can fabricate low cost and better quality devices, by employing the organic materials.

A schematic representation of a FET with bottom-gate architecture is shown in Fig. 27.20. A drain or source-drain voltage V_D is applied between the source (S) and the drain (D). The applied voltage drives a current through the semiconducting transistor channel. When a mobile charge carrier present in the channel gives the drain current I_D, the carriers are extracted from the source by a gate voltage V_G into the semiconducting channel. The above carriers do not reach the metal gate since the gate dielectric is an insulator, i.e., dielectric material. Hence, it forms an accumulation layer at the channel-insulator surface. When a potential V_G is applied, the semiconductor channel gets doped and it is reversed quickly when V_G is switched off. Hence, FET is known as an electronic switch. One can switch on the OFET by applying V_D and V_G of equal polarity opposite to the sign of the mobile carriers. Therefore, OFETs can easily determine the type of carrier in a material. When a positive V_G is applied, the transistor is switched on by V_D and hence, the carriers are negative, i.e., electrons. Similarly, when a negative V_G, is applied, the carriers are positive, i.e. holes. The high work function metals such as PEDOT/ PSS are used to minimise the injection barriers from source to the semiconductor for hole channel semiconductors. Similarly, the low work function metals, such as Ca, are used for the electron channel semiconductor.

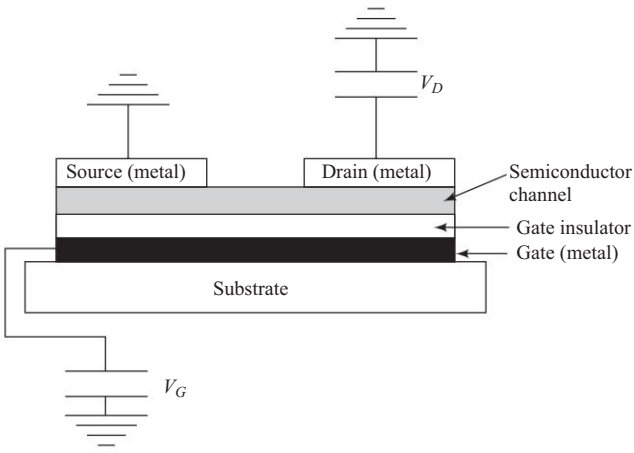


Fig. 27.20 *FET with bottom gate architecture*

The McGraw·Hill Companies

The logic circuits require large number of transistors which are obtained based on the principle of inverter. For example, a simple inverter is formed by using two FETs.

27.9.2 Organic Light Emitting Devices

We know that CRT are bulky and consume large amount of energy. On the other hand, LCDs are passive devices which require backlighting and also suffer from poor viewing angle. All the drawbacks of LCDs and CRTs are overcome employing the organic light emitting devices (OLED).

During fluorescence, excitons are created by the absorption of light before fluorescing, while in electroluminescence, excitons are created by electron and hole polaron capture. The injected polarons migrate towards each other to form excitons through the electrodes. These excitons decay to emit the light source. The basic architecture of an organic light emitting device is shown in Fig. 27.21.

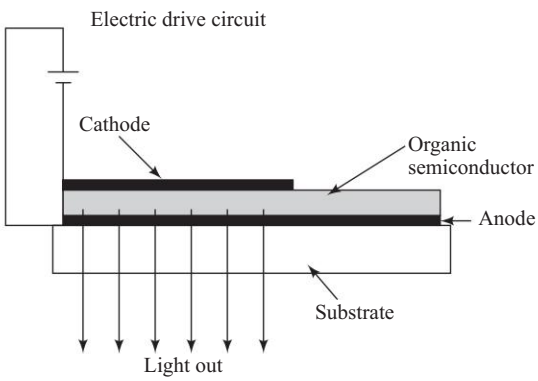


Fig. 27.21 *OLED architecture*

The bipolar carrier injection can induce the excitons in a semiconductor. The electrons and holes are injected into semiconductor regime through the electrodes which are connected to an electric circuit. The negative electrode known as *cathode* induces electrons while the positive electrode known as anode induce holes into semiconductors. The combination of anode and cathode produce electrons hole, i.e, polarons. In OLEDS, high work function anode and a low work function cathode are used as polar injector electrodes. Thus, different metal electrodes are used as source and drain electrodes. One can use the forward and a reverse bias for OLED operations. For example, electrodes like ITO anode and magnesium cathode are used as injectors for OLED. It is very difficult to achieve ohmic, i.e., (barrier free) injection at both electrodes, which is the breakthrough towards the efficient organic energy level devices.

Organic were successfully used to generate barrier free injection which consists of hole transporting layer (HTL) and an electron-transporting layer (ETL). This double transporting layer has the necessary band gap (2.5 eV) for the free conduction of electrons and holes from electrodes to organic semiconductor and it can reduce the barrier resistance and band gab energy in organic semiconductors. The energy level diagram of the double layer device using an ITO anode and a magnesium cathode is shown in Fig. 27.22.

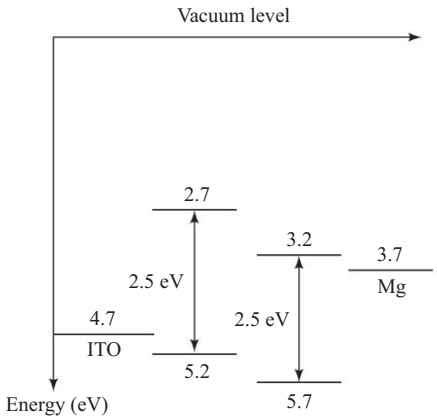


Fig. 27.22 *Energy level diagram—double layer device*

Exciton Formations We know that electron and hole polarons injected into the device combine into excitons which emits light. In order to form the exciton in a multilayer, it has to overcome an internal barrier at the HTL-ETL interface. Thus, to form the excitons it requires certain amount of energy. The exciton binding energy (E_b) lies between 0.1 – 0.3 eV. The formation of exciton requires certain amount of energy in the form of binding energy which is given by,

$$h_{EL} = \frac{(\sigma_s / \sigma_T)}{\sigma_s / \sigma_T + 3} \eta_{PL} \tag{27.3}$$

where σ_s and σ_T are the polaron capture cross-sections for singlet and triplet exciton formulation, respectively. Thus, one can maximise the amount of light generated from the excitons by an appropriate selection of materials with high luminescence quantum yield.

In organic semiconductor, these radioactive decays are under in chain process. Energy (E_f) is required to transfer the polarons and excitons though the hole of semiconductor regime. When the binding energy E_b is less than polarons transfer energy, the semiconductor emits light weakly. On the other hand, if $E_f > E_b$, the semiconductor forms excitons and hence, it emits the light simultaneously. In order to remove the unwanted non-emissive excitons in OLED and hence to improve the efficiency, there are several methods used. Some of the most ideal methods are singlet excitons formation, use of dendrons with a conjugated core surrounded by non-conjugated dendrimers, etc.

27.9.3 Organic Photovoltaic

Reverse electro luminescence is the basic mechanism for photovoltaics. When a semiconductor absorbs photons, it leads to the production of a exciton. Thus, the exciton is separated into electron and hole polarons. The separated electrons and hole polarons are migrated into opposite electrodes through an external load circuit. In order to achieve exciton separation, the internal field available in a semiconductor should overcome the electron hole coulomb interaction.

A schematic representation of the formation of a conventional inorganic photovoltaic semiconductor device is shown in Fig. 27.23. It consist of a thick

The McGraw·Hill Companies

n-doped layer above which there is a thin semi-transparent *p*-doped inorganic semiconductor. Thus, a rectifying p-n junction is formed between p- and n- layer. As a result, the depletion layer is formed with a few charge carriers resulting in a high resistivity and internal field. When a light is incident on the photovoltaic device, an exciton is created due to the absorption of photons at the p-n junction. Due to the existence of strong internal field, the carriers will recombine through the external load R, when it is smaller than the depletion layer.

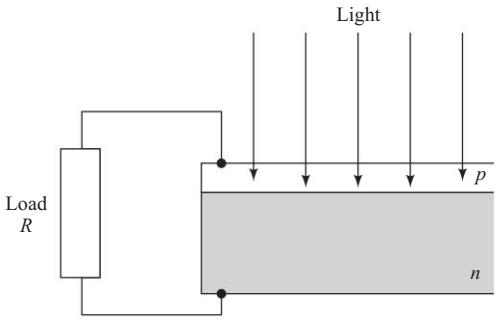


Fig. 27.23 *Inorganic semiconductor photovoltaic diode*

The I/V characteristic of the photovoltaic diode under dark and light illumination is shown in Fig. 27.24. *V* is positive during the forward bias and current flows from *p* to *n*, and when it is negative, then a current flows from *n* to *p*. When *p* and *n* terminals are shortened ($R = 0$), and the diode is exposed to illumination, a current will flow due to the photo generated carriers. The above short circuit current will be moved downwards when compared with current under forward bias.

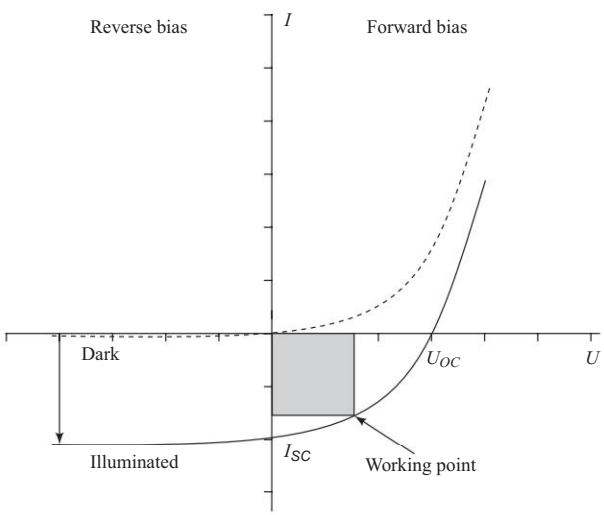


Fig. 27.24 *Characteristics of p-n junctions with and without illuminations*

One can obtain the power of the photovoltaic cell from the area of the shaded portion. For a particular voltage, the power delivered by the cell is maximum, which

can be obtained from the fill factor. The fill factor is an important characteristic of a photovoltaic cell. The efficiency of photovoltaic cell is calculated from the ratio between number of carriers generated and the number of photons absorbed.

$$\eta_p = \frac{P_{out}}{P_{in}} \tag{27.4}$$

The main problem in case of organic semiconductor diodes is the short diffusion lengths. The excitons are separated into carriers only when they are created near the *p-n* interface, i.e., at a diffusion radius of nearly 5 nm. The excitons produced further away from the interface will recombine only under fluorescence but will not provide any current. This is the only problem in inorganic semiconductors. One can overcome the problem of these short diffusion lengths by using organic semiconductors.

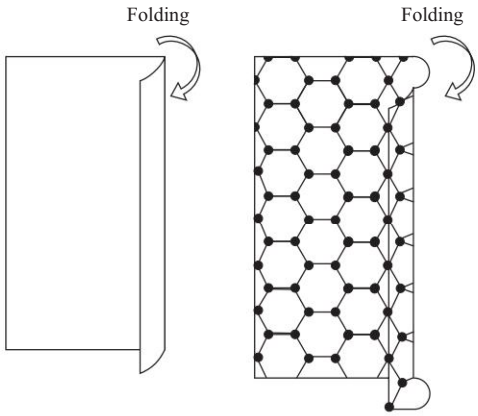
27.10 CARBON NANOTUBES

The unique nature of the carbon bond in organic molecules of life leads to interesting nanostructures particularly, carbon nanotubes (CNT). One of the more interesting nanostructures with huge potential applications is the carbon nanotube. The main interest in CNT is due to the existence of exotic properties like very low specific resistivities in metallic carbon nanotubes and high hole mobilities for semiconducting nanotubes. The carbon nanotubes are formed by rolling the graphite (or graphene) sheet into tubes with the bonds at the end of the sheet. These bonds are used to close the tube. Generally, the carbon nanotubes are formed in the range of 2 to 10 nm in diameter and a length of 100 μm.

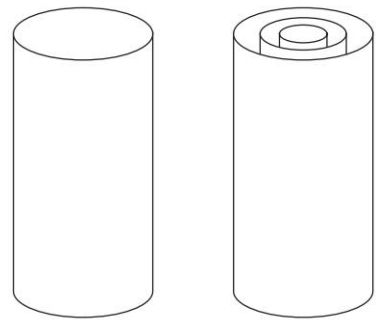
Let us discuss briefly the different types, synthesis, properties and applications of CNT in detail in the following sections.

27.11 TYPES OF CARBON NANOTUBES

One can obtain the carbon nanotube by folding the graphite sheet as shown in Fig. 27.25. During the folding of the graphite sheet, concentric cylinders can be formed as a nanotube. These concentric nanotubes are known as *multi-wall carbon nanotubes* (MWCNT). The MWCNT walls are separated by a small distance of 0.3 nm. The MWCNT is the most common and easily formed. One can obtain the *single wall carbon nanotubes* (SWCNT) under specified conditions.



(a) Rolling of carbon sheet



(b) Carbon nanotubes

Fig. 27.25 *Schematic representation of carbon nanotubes*

Even though the carbon nanotubes are not formed by rolling the graphite sheets, one can explain the different structures of CNT by considering the way in which the graphite sheet might be rolled into tubes. In order to explain the three different structures of CNT, let us consider the structure of graphite sheet as shown in Fig. 27.26.

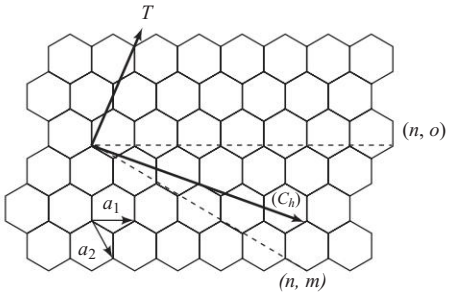
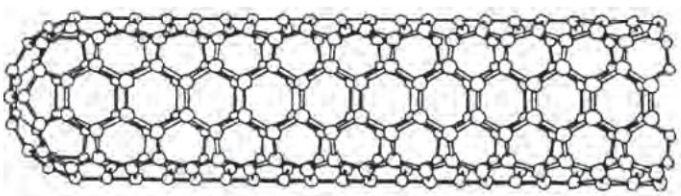
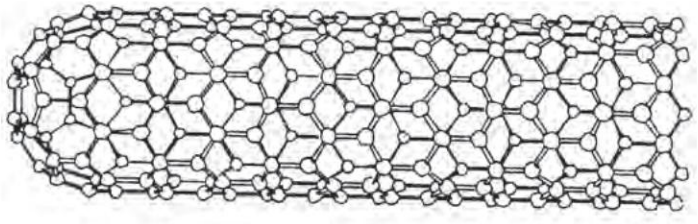


Fig. 27.26 *Graphite sheet with various directions and cells*

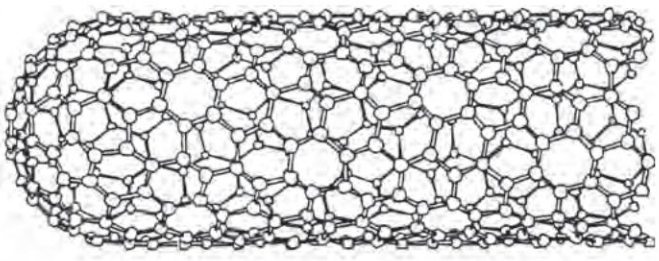
When the graphite sheet shown in Fig. 27.26, is rolled up about an axis *T*, the CNT is formed. The circumferential vector is at right angles to *T*. By rolling the graphite sheet about the *T* vector at different orientations, three different CNT structures can be obtained. When the vector *T* is parallel to the C—C bonds of the carbon hexagons, a structure known as *armchair structure* is obtained. By rolling the graphite sheet at different orientation about a *T* vector, the *zigzag* and *chiral* structures are obtained. In both cases, the *T* vector is not parallel to C—C bonds. The arrangement of carbon atoms in three different structures are different and are shown in Fig. 27.27.



(a) Arm chair structure



(b) Zigzag structure



(c) Chiral structure

Fig. 27.27 *Structure of CNT*

During the formation of CNT, it gets capped at both ends with hemispheres of fullerenes. One end of the SWCNT consists of metal particles which reveal the catalytic role of the metal particles for the formation of CNT. In addition to the above three types, there are many other CNT shapes like ropes, springs, conical shapes, bamboo structures, etc.

27.12 SYNTHESIS OF CARBON NANOTUBES

Carbon nanotubes are synthesized by three different methods, namely, laser evaporation, carbon arc and chemical vapour deposition. Even though different methods are available for the synthesis of CNT, the laser ablation and chemical vapour deposition methods have been optimised for the good quality of CNT and an increased yield. A brief discussion on the synthesis of CNT employing the electric arc discharge, chemical vapour deposition and laser ablation are given in the following section.

(i) Electric Arc Discharge The electric arc discharge experimental arrangements used to produce the carbon nanotubes are shown in Fig. 27.28. The graphite rods with a diameter of 5 to 20 mm have been used as electrodes. The two electrodes are fixed in the discharge chamber by separating them at a distance of 1 mm. The discharge chamber is evacuated and filled with helium gas at a pressure of 500 torr, which helps in forming the CNT. A potential of 20–25 volts is applied across the electrodes. The central region of the chamber reaches temperature in the order of 3273 K. The necessary cooling arrangements have been made to control the temperature inside the chamber.

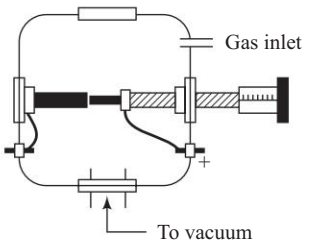


Fig. 27.28 *Electric arc discharge CNT*

When the potential is applied to the electrodes, the carbon atoms are ejected from the positive electrodes while the nanotubes are formed on the negative electrodes. Thus, the nanotubes are aligned in the correct direction between the two electrodes. Due to the formation of the nanotubes, the length of the positive electrode decreases and hence, carbon dioxide is formed on the electrode. By incorporating a small amount of cobalt or nickel or iron, as a catalyst in the central regions of the positive electrode, one can produce single wall carbon nanotubes. On the other hand, if no catalysts are used, it produces multiwall carbon nanotubes. The electric arc method can produce the nanotubes of diameter 0.7 to 5 nm with a length of few micrometers.

(ii) Chemical Vapour Deposition The chemical vapour deposition (CVD) is used to produce large quantity of nanotubes. The experimental set-up used for synthesising the CNT by CVD is shown in Fig. 27.29. During this process, the hydrocarbon gas is decomposed at certain conditions. A high temperature vacuum furnace with a provision to maintain the inert atmosphere inside the furnace is used for producing the nanotubes. The solid substrate which contains catalyst like cobalt, nickel and iron is kept inside the furnace. The gas like methane is passed into the furnace which is kept at 1373 K and hence, the decomposition of the methane gas takes place. The decomposition of the gas produces carbon atoms and is allowed to condense on a cooler substrate which contains the catalyst. The catalyst plays a dominant role in forming the nanotubes. The above process is continuous and hence, it produces the nanotubes continuously. The purity of the nanotubes produced by this method is very high as compared to other methods. As the method is a continuous one, the same can be scaled up for mass productions of nanotubes.

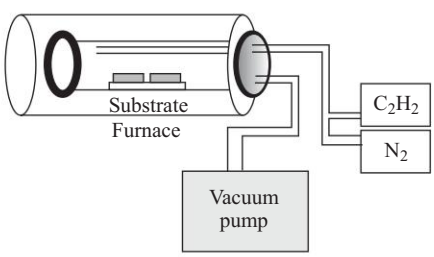


Fig. 27.29 *Chemical vapour deposition- CNT*

(iii) Laser Ablation The experimental arrangement used for the synthesis of CNT employing the laser ablation is shown in Fig. 27.30. A quartz tube which contains the graphite target is kept inside the high temperature muffle furnace. The quartz tube is filled with argon gas and heated to 1473 K. A water cooled copper collector is fitted at the other end of the tube as shown in Fig. 27.30. The target materials graphite

contains small amount of nickel and cobalt as a catalyst to nucleate the formation of nanotubes. When an intense pulse of laser beam is incident on the target, it evaporates the carbon from the graphite. The evaporated carbon atoms are *sweped* from the higher temperature argon gas to the colder copper collector. When the carbon atoms reach the colder copper, it condenses into nanotubes. The nanotubes of 10–20 nm diameter and 100 μm long can be produced by this method.

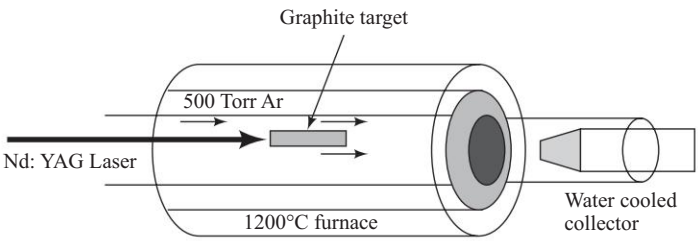


Fig. 27.30 *Laser ablation- CNT*

27.13 PROPERTIES OF CNT

Generally, the CNTs are produced both in metallic and semiconducting nature during the process. The final product consists of two-thirds with semiconducting and one-third with metallic property. The semiconducting or metallic property of the CNT depends on the diameter and chirality, i.e., the nature of the rolling of the tube with respect to the vector *T*. The metallic CNT will have the armchair structure, while the semiconducting tubes will have the chiral structure. The dependence of the energy gap of the semiconducting chiral CNT on the reciprocal of the diameter is shown in Fig. 27.31. It is evident that there is a decrease in the band gap with increase in diameter of CNT.

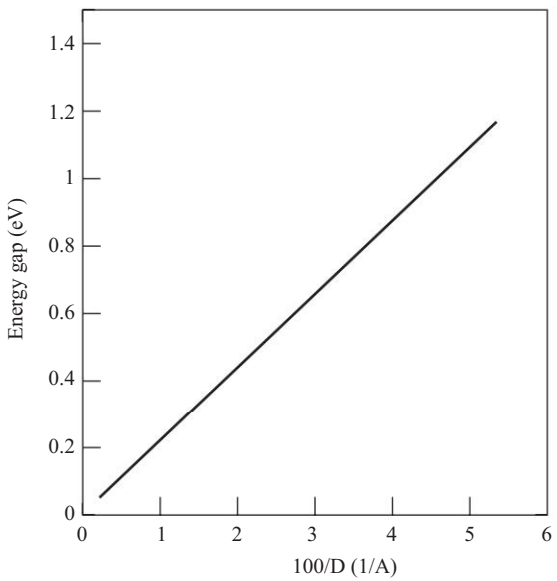


Fig. 27.31 *Energy gap versus reciprocal of diameter of chiral CNT*

The McGraw·Hill Companies

(i) Electrical Properties The electronic structure of CNT is obtained using STM. During this measurement, the STM tip is fixed just above the CNT. The tunneling current I is measured during the change in the applied voltage V between the tip and the sample. From the above measurements, the conductance of the nanotubes is

$$G = I/V \tag{27.5}$$

The above measured conductance G reveals the local electronic density of states. One can obtain information about the energy gap in a semiconducting material by drawing a plot between (dI/dV) and (I/V).

(ii) Magnetic Property CNT exhibits the magneto resistance phenomenon similar to that of bulk materials. The CNT display the magneto resistive effects at low temperature, i.e., the resistance of the CNT is changed by the application of dc magnetic field. The magnetic field dependence of the charge in resistance DR to their resistance in zero magnetic fields at 2.3 and 0.35 K of nanotubes are shown in Fig. 27.32. The observed magneto resistance effect is negative. This is due to the decrease in resistance with increase in magnetic field. The decrease in resistance results in an increase in conductance G ($\simeq 1/R$). The above effect can be explained by considering the applications of dc field to the nanotubes. When a = magnetic field is applied to the nanotubes, the conduction electron acquires a new energy level known as Landan level. This energy level lies very close to the Fermi level. As a result, more states are available for electrons to increase their energy and hence, the conducting property.

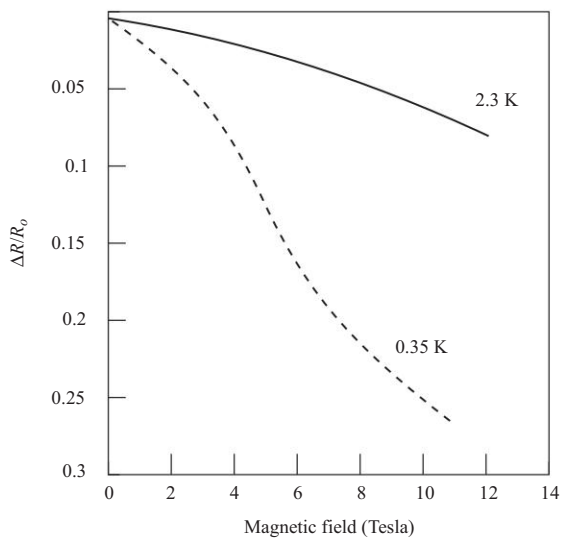


Fig. 27.32 *CNT-Magneto resistance effect*

(iii) Vibration Properties Similar to carbon dioxide, the atoms in a molecules (or) nanoparticles can vibrate back and forth. Each molecules can have a set of vibrational motions known as *normal mode of vibrations*. The different modes of vibrational motion are determined from the symmetry of the molecule. There are two normal modes of vibration which exist in CNT, as shown in Fig. 27.33. In the first

mode marked as A_{1g}, the diameter of the CNT moves in and out at a frequency of 165 cm^{-1}. In the second mode, known as E_{2g}, the squashing of the tube takes place in one direction while the expansion takes place in the perpendicular direction. Thus, it oscillates between a sphere and an ellipse at a frequency of 17 cm^{-1}. The frequency of the vibration modes depends on the diameter of the nanotube. Thus, one can determine the radius of the nanotube from the frequency dependent radius of the tube.

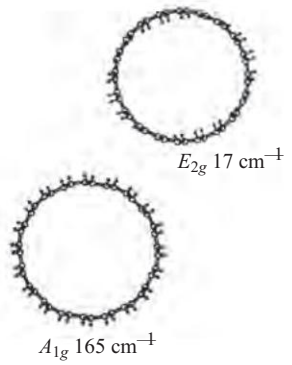


Fig. 27.33 *CNT—Vibrational properties*

(iv) Mechanical Properties The CNT are very strong, that is, they are about 10 times stronger than steel. Let us consider that one end of a thin wire is nailed to the roof of a room while a weight W is attached to the other end. The stress acting on the wire is given by

$$S = W / A \tag{27.6}$$

where A is the cross-sectional area of the wire.

The strain of the wire is given by the amount of stretch ΔL of the wire of length L,

$$S = \Delta L / L \tag{27.7}$$

where L is the length of the wire before attaching the weight.

We known that stress is proportional to strain

$$S = Ee \tag{27.8}$$

where E is the proportionality constant and is equal to $LW/ A \Delta L$, known as Young's modulus.

The Young's modulus of the material is used to characterise the elastic flexibility. For example, the larger the value of Young's modulus, the lesser is the flexibility. The Young's modulus of the carbon nanotube is in the range from 1.28 to 1.8 Tpa (1Tpa~ 10^7 atm), while for steel is 0.21 Tpa. It means that the CNT is 10 times stronger than that of steel. The mechanical property of SWCNT and MWCNT makes the nanotubes a different material from other conventional materials.

27.14 APPLICATIONS

The potential applications of CNTs are increasing more than the other materials due to their unusual properties like electrical, magnetic and mechanical. The applications of nanotubes include battery electrodes, electronic devices, etc. Even though CNT has diversified potential applications, few applications are given in detail in the following sections.

The McGraw·Hill Companies

(i) Field Emission and Shielding Consider that a small amount of electric field is applied parallel to the axis of a nanotube. As a result, it emits electrons at a very high rate from the ends of the tube. This emission is known as *field emission*. The electron emission of carbon nanotubes are used in electronic industries like flat panel display. The nanotubes in the form of thin film are placed over control electronics with a phosphorous coated glass plate at the top. The nanotubes are used for electron emission effect in vacuum lamps. The resultant lights are much brighter than the conventional light. The nanotubes used electron emission products are long level and more efficient in terms of performance.

(ii) Computers Due to the miniaturisation of the computers, there is an increasing demand in increasing the number of switches in a chip. The metal wires with small diameters are used to do the interconnections between the computers and switches. A decrease in the diameter of a wire in turn increases the resistance and hence, the wire generates heat due to the flow of current. The carbon nanotubes with a diameter 2 nm can carry large current without any heating. Thus, the CNT are used as connectors in computer industries. The electrical current made using the CNT in field effect transistor (FET) is shown in Fig. 27.34.

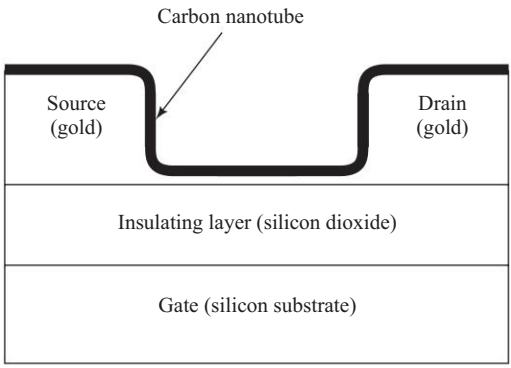


Fig. 27.34 *FET made from carbon nanotubes*

Similarly, the CNT is used as switching device in computers. The schematic representation of the concept of switching device using CNT is shown in Fig. 27.35.

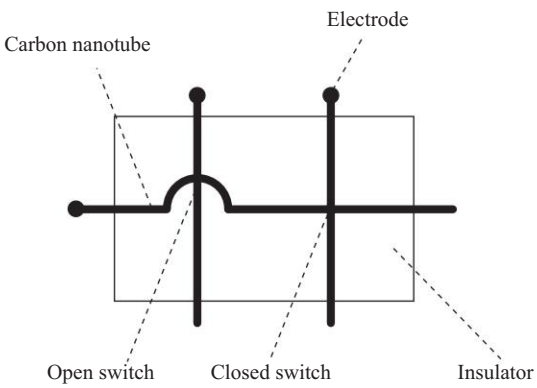


Fig. 27.35 *Computer switching device using CNT*

(iii) Fuel Cells Carbon nanotubes have been used as a storage device in battery, fuel cells, etc. In case of batteries, lithium is used as a charge carrier. The lithium can be stored in a carbon nanotube. One electron atom requires six carbons in the nanotube. Similarly, one can also store hydrogen in the carbon nanotube. The schematic representation of the electro chemical experimental cell is used to inject the hydrogen atom into the carbon nanotube is shown in Fig. 27.36. The electro chemical cell consists of an electrolytic solution of KOH. The negative electrode is made up of single wall carbon nanotubes (SWCNT) paper, while the positive electrode consist of $Ni(OH)_2$. When a potential is applied between the electrodes, the water of the electrode is decomposed into positive hydrogen ions (H^+). The hydrogen ion is attracted towards the negative MWCNT electrode. The existence of hydrogen on the CNT can be studied by measuring the intensity of the Raman active vibration before and after the electrochemical process. The development of fuel cells by storing hydrogen on SWCNT can be used as an alternative source of energy for future automobile.

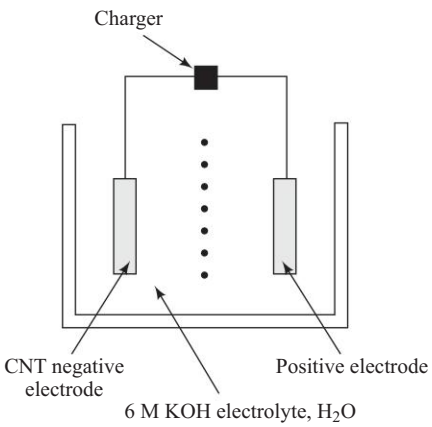


Fig. 27.36 *Fuel cells-CNT*

(iv) Other Applications In addition to the above applications, the CNT finds wide potential in industries. The chemical sensors are made using CNT to detect various gases like NO_2. CNT is used as a catalyst for enhancing the chemical reactions. For example, cadmium sulphide (CdS) crystals are formed inside the CNT due to the reaction of hydrogen sulfide gas (H_2S) at a temperature of 673 K.

Keypoints to Remember

- In diamagnetic materials, there is no permanent dipole.
- Retentivity is the remaining magnetisation when the magnetic filed is reduced to zero from saturation field.
- Coercivity is the field required to bring magnetisation to zero from resonance.

The McGraw·Hill Companies

- Nanomagnetism is the study of the behaviour of ferromagnetic material when the domains are aligned in one dimension.
- Particulate nanomagnets are granular solids with one or more phases having magnetic properties.
- Ordinary Magneto Resistance (OMR) is the property of a material to change the value of its electrical resistance under the influence of an external magnetic field.
- Tunneling magneto resistance (TMR) is the process of changing the sample magnetism in magnitude or direction.
- GMR is the phenomena in which a significant change (>5%) in electrical resistance occurs due to change in magnetisation direction or magnitude.
- TMR is a magneto resistance device which relies on electron tunneling through oxide barrier between two ferromagnets.
- Targeted drug delivery technique is the process of sterring of nanoparticles by means of external magnetic field gradients to reach the desired parts of the human body.
- Injection laser is a semiconductor laser used to convert electrical energy into light energy with high efficiency.
- Quantum cascade laser is a new semiconductor laser which emits light in the infrared region of electromagnetic spectrum.
- Carbon nanotubes (CNT) are formed by rolling the graphite sheet into tubes with bonds at the end of sheet.
- CNT are two types, namely Single Wall Carbon NanoTube (SWCNT) and Multi Wall Carbon NanoTube (MWCNT).
- Structure of CNT are three types, namely, chair, zigzag and chiral structures.
- Electron arc discharge, chemical vapour deposition and laser ablation are the methods used to synthesise CNT.
- Semiconducting CNT takes chiral structures, while metallic CNT takes armchair structures.
- Field emission is the process of emitting electrons at a very high rate from the ends of CNT when a small amount of electric field is applied to parallel to the axis of nanotube.

Objectives Questions

27.1 There is no permanent dipole in diamagnetic materials. (True / False)
27.2 The different forms of nanomagnetes are _____, _____, and _____.
27.3 Particulate nanomagnetes are prepared employing _____ method.
27.4 Examples for particulate nanomagnets are _____, _____ and _____.
27.5 Expand MEMS and NEMS.
27.6 Expand OMR and TMR.

The McGraw·Hill Companies

27.7 Expand GMR and CMR.
27.8 In GMR materials, the change in the electrical resistance is _____.
27.9 Expand CIP and CPP.
27.10 Expand MRAM.
27.11 Expand MFM.
27.12 Acronym for MRI.
27.13 Nanomagnetic materials is used as a _____.
27.14 Expand NMR.
27.15 Injection laser convert the electrical energy into _____ energy.
27.16 The threshold current density of a semiconductor laser is _____.

a) $J_{th} = J_{th} \exp\left(T_1 / T_0\right)$ b) $J = J_{th}^{0} \exp\left(T_1 / T_0\right)$

c) $J_{th} = J^{0} \exp\left(T_1 / T_0\right)$ d) $J_{th} = J_{th}^{0} \exp\left(T_1 / T_0\right)$

27.17 Quantum cascade laser emits light in the region_____.
a) ultraviolet b) visible
c) infrared d) far a infrared
27.18 Expand RAM
27.19 According to Moore's law, the transistors are entering into nanoscale very soon. (True / False)
27.20 Expand MOSFET.
27.21 Expand OFET.
27.22 Expand CRT and LCD.
27.23 Acronym for OLED.
27.24 Expand HTL and ETL.
27.25 The efficiency of photovoltaic cell is equal to _____.
27.26 Expand CNT.
27.27 Expand SWCNT and MWCNT.
27.28 The following are the structure of CNT.
a) Chiral, zigzag b) chiral and armchair
c) chiral, zigzag, SWCNT d) chiral, zigzag and armchair.
27.29 Catalyst are required to produce multiwall carbon nanotube. (True/False).
27.30 The catalyst used for the MWCNT is _____ or _____.
27.31 Metallic CNT takes _____ structures.
27.32 Semiconductors CNT takes _____ structures.
27.33 The conductance of the nanotube is equal to G = _____.
27.34 CNT are _____ time stronger than steel.
27.35 CNT's are used for electron emission effect in vacuum lamps. (True/False)

KEY

27.1 True
27.2 zero, 1D, 2D
27.3 rapid solidification
27.4 $Fe_{73.5}BaSi_{13.5}$ Nb_3Cu, $Nd_2Fe_{14}B$, CaCrTa
27.5 Micro Electro Mechanical Systems, Nano Electro Mechanical Systems.
27.6 Ordinary Magneto Resistance, Tunneling Magneto Resistance.
27.7 Giant Magneto Resistance, Colossal Magneto Resistance
27.8 > 5%

The McGraw·Hill Companies

27.9 Current is Parallel to Plane, Current is Perpendicular to plane
27.10 Magnetic Random Access Memory
27.11 Magnetic Force Microscope
27.12 Magnetic Resource Imaging.
27.13 contrast
27.14 Nuclear Magnetic Resonance
27.15 light
27.16 d)
27.17 c)
27.18 Random Access Memory
27.19 True
27.20 Metal Oxide Semiconductor Field Effect Transition
27.21 Organic Field Effect Transition
27.22 Cathode Ray Tube, Liquid Crystal Display
27.23 Organic Light Emitting Diode
27.24 Hole Transporting Layer, Electron Transporting Layer
27.25 P_{out} / P_{in}
27.26 Carbon Nano Tube
27.27 Single Wall Carbon Nano Tube, Multi Wall Carbon Nano Tube
27.28 d)
27.29 True
27.30 cobalt, nickel
27.31 armchair
27.32 chiral
27.33 I /V
27.34 10
27.35 True

Short Questions

27.1. Explain ordinary magneto resistance effect.
27.2. What is GMR?
27.3. Distinguish between GMR and CMR.
27.4. Mention the applications of GMR materials.
27.5. What is meant by TMR?
27.6. What is meant by probing nanomagnetic materials?
27.7. What is meant by sensor?
27.8. Mention few applications of nanomagnetic materials.
27.9. What is meant by carbon nanotubes?
27.10. Differentiate multiwalled and single walled carbon nanotubes.
27.11. Explain the different structure of CNT.
27.12. What is meant by electric arc discharge?
27.13. How the CNT is produced employing CVD?
27.14. Define laser ablation.
27.15. What is meant by magnetosensitive effects?
27.16. Mention the various application of CNT.
27.17. What is meant by injection laser?
27.18. Explain Quantum dot laser.
27.19. Explain the principle behind the quantum cascade laser.
27.20. What is meant by optical memory?

27.21. How do you correlate the Moore's law with nanotechnology?
27.22. Explain single electron memory cell.
27.23. What is meant by Field effect transistor?
27.24. How the organic light emitting diodes overcomes the drawback of LCDs and CRTs?
27.25. Explain the principle of organic photovoltaic.

Descriptive Questions

27.1. What are the classifications of nanomagnetic materials? Explain the different nanomagnetic materials with suitable illustrations.
27.2. What is the relationship between bulk magnetic and nano magnetic materials? Explain the various applications of nanomagnetic materials with suitable examples.
27.3. Describe with neat sketch how the CNT are synthesized by different methods. Explain the important properties of CNT.
27.4. Explain the different structures of CNT with suitable diagram. How the CNT are used in industries for potential applications?
27.5. Describe with neat sketch the applications of semiconductor nano-structured devices.
27.6. Explain with neat sketch the principle, working and applications of OFET, OLED and organic photovoltaic.
27.7. Discuss the construction and working of organic solar cell.
27.8. What is electro chemical deposition? Explain how this technique is used in templating growth with necessary diagram.