The McGraw-Hill Companies

Engineering Physics

(CC 110011)

Common for all Engineering Branches

About the Author

Dattuprasad Ramanlal Joshi has been associated with The M S University of Baroda since 1978 when he first enlisted himself as a student. After completing his education, he has held various positions like Lecturer, Reader and Associate Professor. He has a teaching experience of 23 years and a research experience of 27 years.

His education includes an MSc (Applied Physics), obtained from the Faculty of Technology and Engineering, The M S University of Baroda, in the year 1983. He was then awarded the University Research Fellowship and subsequently UGC (JRF and SRF) Merit Research Fellowship for four years. Thereafter, he obtained a PhD (Applied Physics) from the Faculty of Technology and Engineering, The M S University of Baroda, in the year 1991. During his doctoral research period, he visited, worked and presented research papers at many national laboratories like Crystal Growth Centre (Anna University, Chennai), Centre for Device Applications (IIT Madras, Chennai), BARC (Mumbai), TIFR (Mumbai), (IIT Bombay, Mumbai) NPL (New Delhi) where he was trained in the field of crystal growth and characterisation which was the field of his research.

Dr Joshi has attended and participated in several international conferences/seminars/symposia/workshops within India and presented more than 15 research papers at national conferences/seminars, all of which were published in the international and national journals.

Beside these achievements, he has worked as an expert member for the selection committee as a Lecturer and Reader in Physics, as a Secretary and Chairman of the IPA Baroda chapter, and as a chairman of the Syllabus Committee for the paper on Applied Physics. Dr Joshi is a Member of the Board of Studies in Applied Physics and a Member of the Board of Studies, Faculty of Technology and Engineering, The MS University of Baroda.

Yashesh H Gandhi has been associated with the Maharaja Sayajirao University of Baroda since 1987. He obtained his MSc and PhD degrees from this university. Presently, he is working as Associate Professor in the same university. He has a teaching experience of about 18 years and a research experience of about 20 years.

His education includes MSc (Physics) with specialization in Electronics and Radio Physics, obtained in the year 1989. He was then awarded the university research fellowship. Dr Gandhi obtained his PhD(Applied Physics) in the year 1996. He did pioneer work in thermo-luminescence studies on synthetic quartz material. During and after his doctoral research work, he visited, worked and presented a number of research papers in the research institutions and universities like BARC (Mumbai), PRL(Ahmedabad), IIT Mumbai, NPL Delhi, and Crystal Growth Centre, Chennai.

Dr Gandhi has participated in several international/national conferences/seminars/workshops/symposia, delivered many invitee lectures and presented more than 25 research papers to his credit that have been published in international/national journals or proceedings. He was the first person to initiate Optically Stimulated Luminescence (OSL) studies in synthetic quartz in India and has guided a research student in the field of optically stimulated luminescence.

Besides these achievements, He has worked as a treasurer of the IPA Baroda chapter. He was instrumental in framing the revision of the entire course work of PG diploma in solid-state electronics and in revising the syllabus of Applied Physics for the first semester of the BE course.

Yogesh D Kale has been associated with the L C Institute of Technology, Bhandu since 2009. Prior to 2009, he worked as a physics teacher for two years in a higher secondary school. He has also worked in Cosmo Films Limited as a Production Officer from 1998 to 2009. He joined research since 1999 after completion of The M S University minor research project in the field of luminescence. He continued his study of luminescence for his doctoral thesis and obtained a PhD in the field of Optically Stimulated Luminescence (OSL) of synthetic quartz from The M S University of Baroda in 2006.

He became the first candidate who started work on OSL of synthetic quartz in the university and became the third such candidate in the country. Along with his research work, he has participated in several international and national conferences and published more than 15 research papers in international and national journals and proceedings. He is a member of the Luminescence Society of India and the Indian Physics Association.

The McGraw-Hill Companies

Engineering Physics

(CC 110011)

Common for all Engineering Branches

Dattu Joshi
Head, Applied Physics Department
Faculty of Technology and Engineering
The MS University of Baroda
Vadodara, Gujarat

Yashesh Gandhi
Associate Professor
Applied Physics Department
The MS University of Baroda
Vadodara, Gujarat

Yogesh Kale
Lecturer in Physics
LC Institute of Technology
Bhandu, Mahesana
Gujarat

Tata McGraw Hill Education Private 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

The McGraw·Hill Companies

Tata McGraw-Hill

Published by Tata McGraw Hill Education Private Limited,
7 West Patel Nagar, New Delhi 110 008.

Engineering Physics (GTU)

Copyright © 2011 by Tata McGraw Hill Education Private 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 Education Private Limited.

ISBN (13): 978-0-07-107282-3
ISBN (10): 0-07-107282-9

Vice President and Managing Director—McGraw-Hill Education, Asia Pacific Region: *Ajay Shukla*

Head—Higher Education Publishing and Marketing: *Vibha Mahajan*
Manager—Sponsoring: SEM & Tech. Ed.: *Shalini Jha*
Editorial Executive: *Tina Jajoriya*
Executive—Editorial Services: *Sohini Mukherjee*
Sr Production Manager: *Satinder S Baveja*
Production Executive: *Anuj Srivastava*

Dy Marketing Manager: SEM & Tech Ed.: *Biju Ganesan*
Senior Product Specialist: *John Mathews*

General Manager—Production: *Rajender P Ghansela*
Asst. 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 Bharati Composers, D-6/159, Sector-VI, Rohini, Delhi - 110 085, and printed at Nice Printing Press, 3 Rashid Market Extn. Delhi - 110 051

Cover: Rashtriya Printers

The McGraw-Hill Companies

The McGraw-Hill Companies

Dedicated

to

Our Parents

and

Students

The McGraw·Hill Companies

Preface

Brief Introduction to the Subject

The developments in scientific research are so astonishing that modern technology and engineering, which is built on the basic laws of nature, has also changed shape in tune with it. The strong foundation of engineering and technology is laid by the basic fundamentals of physics. In view of this, for quick use of new discoveries and thorough understanding of current-day revolutionized developments in technology and engineering, a perfect knowledge of basic physics and its principles is absolutely necessary. Keeping this fact under consideration, **Gujarat Technological University** (**GTU**) has included basic physics related to new advanced developments in their curriculum under the title of **Physics** (CC 110011) for the courses of all the branches of engineering and technology.

Inspiration Behind this Book

There are plenty of books available in the market showing shortcuts to success in physics. The approach is based on question–answer type guidelike textbooks (!) which loses its real sense of learning, assimilating and applying the ideas / laws in the respective fields. Students of Gujarat are deprived of a physics book which does not create the fear of exams. Thorough understanding is an essential aspect for understanding the simplicity of natural laws of physics. It was felt by most of the teachers and students of Gujarat that there is no single book available that can cater to all the needs. The lack of availability of good textbooks in this area leads to underdeveloped proficiency among the students of this subject, which in turn, affects their professional expertise later on. The problems faced by students and the questions raised regarding the concepts they have grasped during their study stimulated us to write a textbook that aims to develop a thorough understanding of the concepts of physics and analytical skills through our teaching experience of more than two decades. The book will not only serve the purpose of imparting knowledge but will also inculcate curiosity, and analytical quantifying power through engineering applications among the students, and a healthy respect for the scientific community and its ideals. An engineer with a thorough knowledge of his/her subject can work towards developing products to make self-sufficient societies.

Scope/Target Audience

The book is strictly intended for the students of first-year engineering of **Gujarat Technological University (GTU)**. It covers the prescribed syllabus of GTU from the academic year 2006–2007. Solved and unsolved university papers are included at the end of the book with a complete solution of the latest paper. Questions from the past exams are also included at the end of each chapter. The content is presented in such a way so as to enable even an average student to comprehend the concept without much difficulty. A knowledge level of higher secondary grade is enough to understand the topics explained in the book.

Salient Features

- The book covers the prescribed syllabus of physics for first-year BE of Gujarat Technological University of Gujarat.
- Each chapter opens with a puzzle to make the reader develop an interest and look forward to the contents of the chapter**.**
- Beginning each topic with its basics, the content is developed gradually so that students can develop a clear idea of the concepts.
- The book emphasizes on its engineering applications to enhance practical knowledge of the subject.
- Ample conceptual, numerical solved and unsolved problems, MCQ's, reasoning-type short questions and long questions are provided for improving concepts in physics.
- Wherever necessary, neat and clear diagrams with explanations are provided to enhance better grasp over the subject.

Organization

The book includes 10 chapters and the sequence is arranged according to the prescribed syllabus of GTU Physics to help the students. Each chapter does not assume any prior requirement so that individual study of the chapter is possible without any difficulty of coherence in matter due to interdependency of topics. The book is written for a one-semester physics course assuming a total of 50 lecture hours. Problems, MCQ's, conceptual questions/problems can be treated separately in the tutorial lectures.

The book contains the main topics like Architectural Acoustics (Chapter 1), Ultrasonics (Chapter 2), Crystal Physics (Chapter 3), Band Theory of solids (Chapter 4), Lasers (Chapter 5), Optical-Fibre Communication (Chapter 6), Conducting materials (Chapter 7), Superconducting Materials (Chapter 8), New Engineering Materials (Chapter 9), Non Destructive Testing (Chapter 10).

The book format is more suitable for a textbook rather than a reference book. The selection of problems and questions are on the basis of increasing difficulty, keeping in mind its pedagogic character and increasing inquisitions.

Acknowledgements

We feel highly obliged to numerous authors and scholars on the subject whose esteemed works have been freely consulted in the course of preparation of this book. Mentioning of the individual names in particular will be rather too long to be enlisted within such a short span of space. However, special mention ought to be made of our guides Prof. J R Pandya, Prof. R V Joshi and Prof. T R Joshi who were Ex. Head of Physics Department and Applied Physics Department of Faculty of Science as well as Faculty of Technology of The M S University of Baroda due to whom we started our careers as researchers, teachers and thinkers in physics, and whose able guidance made us worthy of writing this book. We take this opportunity to express our deep sense of gratitude to our beloved friend and philosopher Dr R T Shah who was a pioneer in developing CO_2 Laser in India and constantly inspired us to do something useful in the field of physics. The authors are grateful to Dr K V R Murthy and Prof Narayan Murthy, colleagues Dr Chetan Panchal, Dr M S Desai, Dr B S Chakraborthy, Prof V N Potbhare and Prof Arun Pratap, Ex. Heads, Applied Physics Department, Dr R U Purohit, Joint Director, Higher Education, Gujarat State, friends and well wishers whose constant encouragement and cooperation made us capable of completing this work.

Finally, we unreservedly acknowledge our indebtedness to the publishing team of Tata McGraw Hill Education, New Delhi—Ms Vibha Mahajan, Ms Shalini Jha, Ms Tina Jajoriya, Ms Sohini Mukhergee, Mr Anuj Srivastava and Mr Maunish Pandya, who were pioneers in making this book emerge from our dreams and take shape into reality!

Feedback

We sincerely hope that the present book will inspire students and young teachers to think about the mysteries of physics and discuss the ideas behind its basic principles. This book will surely make some contribution in developing a conceptual foundation for the upcoming generation of engineers.

We have used our rich experience, in the presentation and the development of the subject matter in the simplest possible manner, keeping in view the difficulties and limitations of the average students. In spite of our best efforts, some misprints and omissions might have crept in here and there. We shall be grateful if readers bring them to our notice. Any genuine and authentic suggestion for the substantial improvement of the book shall also be heartily welcomed.

The readers are encouraged to provide feedback to the book, either at the author's own email id or at the web site.

Dattu Joshi
Yashesh Gandhi
Yogesh Kale

Publisher's Note

Do you have a feature request? A suggestion? We are always open to new ideas (the best ideas come from you!). You may send your comments to *tmh.corefeedback@gmail.com* (kindly mention the title and author name in the subject line).

Piracy-related issues may also be reported.

The McGraw·Hill Companies

The McGraw·Hill Companies

Contents

The McGraw·Hill Companies

The McGraw-Hill Companies

Roadmap to the Syllabus

Physics

Unit 1: Architectural Acoustics

Classification of Sound : Loudness – Weber – Fechner law Decibel – Absorption Coefficient Reverberation – Saline’s formula – Factors affecting acoustics of buildings and their remedies

GO TO **Chapter 1 Architectural Acoustics**

Unit 2: Ultrasonic

Introduction, production, properties and detection of ultrasonics. Determination of velocity and application of ultrasonic in Engineering

GO TO **Chapter 2 Ultrasonic**

Unit 3: Crystal Physics

Introduction and classification of solids-crystal structure – The crystal systems and Bravias Lattice – Space Lattices of cubic systems – Miller Indices – Relation between Interplanner Distance and cubic Edge and Laws Formula.

GO TO **Chapter 3 Crystal Physics**

Unit 4: Band Theory of Solids

Based theory of Solids – Classification of solids – Energy band structure of conductors, insulator and semi conductions types of diodes (simple diode, Zener diode, varactor diode, LED Solar cells, photovoltaic cell, Photo Conductivity, Hall effects.

GO TO **Chapter 4 Band Theory of Solids**

Unit 5: Lasers

Introduction and properties of Lasers, Stimulated and instantaneous emersion – Relation between Eienstein's '*A*' and '*B*' Coefficients-Population Inversion – Optical – Pumping – Nd-Yag Laser and CO_2 Laser – Application of Laser in Material Processing – Holography – Application of Lasers 6) Optical – Fibre Communication Introduction – Fibre – Optic System – advantages of Fibre optics – Basic principle – Acceptance angle and Numerical Aperture – Types of optics preparation through optical fibre

GO TO **Chapter 5 Lasers**

Unit 6: Optical Fibre Communication

Introduction – Fibre – Optic System – advantages of Fibre optics – Basic principle – Acceptance angle and Numerical Aperture – Types of optics preparation through optical fibre

GO TO **Chapter 6 Optical Fibre Communication**

Unit 7: Conducting Materials

Introduction – conduction in Metals, Electron theory Q.M. treatment – Free electrode theory of metals – Electrical Conductivity – Thermal Conductivity – Wildemann – Franz law – Drawbacks of classical free electrode theory

GO TO **Chapter 7 Conducting Materials**

Unit 8: Super Conducting Materials

Introduction to super conductor – properties of super conductor Type I and Type II super conductor – Comparision between I and II – High T conductors – Application

GO TO **Chapter 8 Super Conducting Materials**

Unit 9: New Engineering Materials

Introduction – Metallic glasses, types, properties, preparation and its application–Introduction to nano technology – method of producing, properties and its application – shape memory alloys – types – shape Memory effect – Pseudo elasticity – properties – application – Bio-materials – General information –Biomedical compatibly of Ti-Al-Nb alloys for implant application

GO TO **Chapter 9 New Engineering Materials**

Unit 10: Non-Destructive Testing

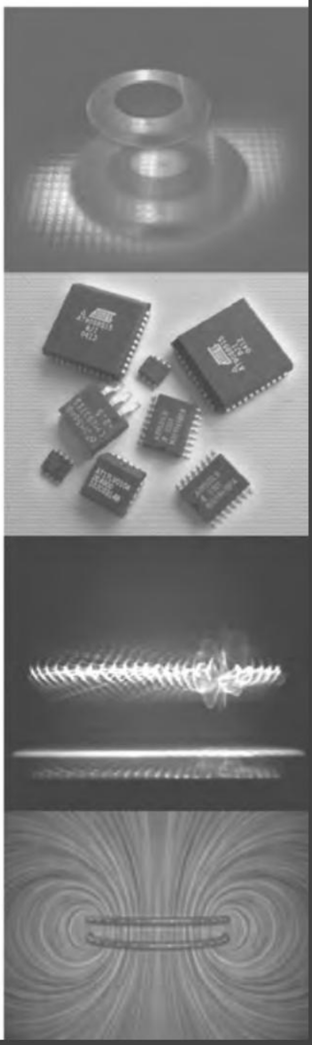
Introduction – The objective of NDT – Types of Defects – Methods of NDT (Liquid Penetrate – Dye Penetrate Radiographic) x X-ray Radiography – X-ray Fluoroscopy – Ultrasonic Inspection method – Pulse Echo System – Visual Display units.

GO TO **Chapter 10 Non-Destructive Testing**

Architectural Acoustics

1

If you have ever attended a music concert in the open air or in a big hall which is generally used for social gatherings like marriages, you may have noticed that the effect of the music is totally absent even if the instruments are the latest or the singer is talented. If a stage drama is performed is the same hall, it seems boring even if the artists and the play are very good. Improving the quality of this type of event is a problem that an engineer solves with his knowledge of acoustics, a branch that deals with the study of the physical properties of sound. In this chapter, you will learn about the fundamental rules that tackle such problems.



INTRODUCTION

1.1 The science of architectural acoustics deals with the design and planning of a building or a hall which will be useful for musical concerts, stage dramas, etc., so as to secure the most advantageous flow of the properly diffused sound for the entire audience without affecting the speech intelligibility and the tonal qualities of the music. Therefore, it is essential that the size and shape of each room and the arrangement of absorptive and selective materials in it should be such as to provide optimum conditions for the growth, decay and steady distribution of sound in the room which will lead to satisfactory hearing and appreciation of the performance of players and speakers. We shall begin the study by a basic understanding of sound, its characteristics and the terms involved in the measurement of the strength and quality of sound.

SOUND, ITS MEANING, HOW IT WORKS

1.2

The word *sound* conveys a double meaning: (i) the mental sensation perceived by the ears, and (ii) the cause responsible for that perception, namely, the physical phenomenon external to the ear—the wave motion which excites the auditory nerves.

Sound occurs due to the vibratory movement of some body, although the movement may be so slight and rapid that it is not visible. A bell, a tuning fork, the strings of a violin, the air inside an organ pipe, etc., are all in a state of tremor when producing sound.

It is evident from the fact that sound stops when vibrations are stopped. If the sounding body is lying in air, the repeated vibratory motion sets up waves which propagate in all directions. On falling upon the tympanic membrane of the ear, they cause it to vibrate and thus produce the sensation of sound. Sound waves cannot be propagated in vacuum as can be easily verified by suspending a ringing electric bell inside a gradually evacuated vessel. The sound would become fainter and fainter, the better the vacuum becomes. It is easy to infer from this that an elastic medium is necessary in order that sound waves may be propagated and conveyed to the ear. Sound is transmitted not only by air but by liquids and solids as well.

Sound is a longitudinal wave motion in material media—solids, liquids and gases—and the velocity with which it is propagated, depends upon the two characteristics of the medium: (i) its elasticity—a property which indicates that when the particles of a medium suffer relative displacements, a force is set up which tends to bring them back to their original configuration, and (ii) its inertia, which depends upon its mass. The actual wave motion depends, therefore, on the kind of initial impulse given from outside, the elasticity and the mass contents of the medium.

How Sound Waves Work

Sound waves exist as variations of pressure in a medium such as air. They are created by the vibration of an object, which causes the air surrounding it to vibrate. The vibrating air then causes the human eardrum to vibrate, which the brain interprets as sound.

In Fig. 1.1, the black dots represent air molecules. As the loudspeaker vibrates, it causes the surrounding molecules to vibrate in a particular pattern represented by the waveform. The vibrating air then causes the listener's eardrum to vibrate in the same pattern.

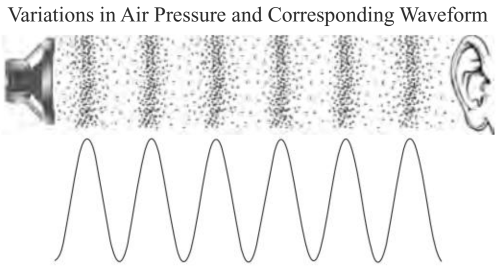


Fig. 1.1 *Sound is produced in air medium by a speaker*

Note that air molecules do not actually travel from the loudspeaker to the ear (that would be wind). Each individual molecule only moves a small distance as it vibrates, but it causes the adjacent molecules to vibrate in a rippling effect all the way to the ear.

Some Important Characteristics of Sound

(i) The propagation of sound waves requires the presence of an elastic medium.
(ii) Sound waves cannot travel through vacuum.
(iii) The velocity of sound waves depends on the nature and temperature of the medium.

CLASSIFICATION OF SOUND

1.3

Sound waves are classified on the basis of the frequency range:

(i) Sound waves having frequencies of less than 20 Hz are known as *infrasonic waves*.
(ii) Sound waves with a frequency range of 20 Hz to 20 kHz are the *audible sound waves*.
(iii) Sound waves with frequencies greater than 20 kHz are the *ultrasonic waves*.

Only audible sound can be perceived with human ears and the other two cannot. Again these audible sound waves are classified into two groups:

(i) Musical sound
(ii) Noise

The waveforms of musical sound and noise are shown in Fig. 1.2.

The waveform of musical sounds is smooth as shown in Fig. 1.2 (a). Musical sounds can be produced using a flute, veena, sitar, harmonium, etc. Periodicity, regularity and continuity are their hallmarks. The waveform of noise, on the other hand, is irregular as shown in Fig. 1.2 (b).

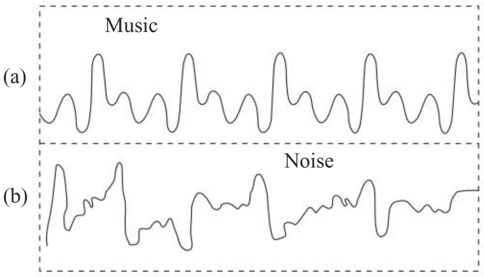


Fig. 1.2 *Waveform of musical sound and noise*

CHARACTERISTICS OF MUSICAL SOUNDS

1.4

The three characteristics which differentiate musical sounds from one another are

- Loudness—intensity of sound
- Pitch—frequency of sound
- Quality, or timbre

(i) Loudness

It is that characteristic of a musical sound that determines the degree of sensation that the sound can produce in the ear. It depends on the intensity of the sound which is objective in nature and the sensitivity of the ear which is subjective in nature.

(ii) Pitch

It is that characteristic of a musical sound by which a shrill sound can be distinguished from a grave one, even though the two sounds may be of the same intensity. It is also defined as that characteristic of sound by which the ear assigns it a place on a musical scale. When a stretched string is plucked, a sound of a certain pitch sensation is produced. If the tension in the string is increased, the pitch (the shrillness) becomes higher. Increasing the tension also increases the frequency of vibration. Therefore, pitch is intimately related to frequency. But frequency alone does not determine the pitch. Below 1000 Hz, the pitch is slightly higher than the frequency and above 1000 Hz, the position is reversed. The loudness of sound also affects the pitch up to 1000 Hz. An increase in loudness causes a decrease in pitch. From about 1000 to 3000 Hz, the pitch is independent of loudness, while above 3000 Hz an increase in loudness causes an increase in pitch.

The voice of a woman or a child is shrill but that of a man is hoarse (i.e., flat or grave). The sound emitted by a cuckoo is shrill while that by a crow or a lion is hoarse. The buzzing of a mosquito, though low in intensity, has a high pitch. The pitch of the sound changes when the source or the observer or both are in relative motion.

(iii) Quality, or Timbre

It is that characteristic of a musical sound which enables us to distinguish between the sounds produced by two different musical instruments or two different persons although their pitch and loudness may be same. It is because of this characteristic that we are able to recognize the voice of a known person over the telephone or to distinguish between the sounds produced by different musical instruments in an orchestra.

WEBER – FECHNER LAW

1.5 The Weber – Fechner law states that the loudness of sound sensed by the ear is directly proportional to the logarithm of the intensity of sound. If L is the loudness and I is the intensity of sound then the mathematical form of the Weber-Fechner law is represented by Eq.1.1 below.

$$L \propto \log_{10} I$$

$$\therefore \quad L = K \log_{10} I \tag{1.1}$$

If D represents the difference in the loudness then

$$D = L_1 - L_2$$

$$= K \log_{10} I_1 - K \log_{10} I_2$$

$$D = K \log_{10}\left(\frac{I_1}{I_2}\right) \tag{1.2}$$

Now the sensitivity of the human ear is defined by $\frac{dL}{dI}$.

On differentiating Eq. (1.1) with respect to I we get the sensitivity of the ear as given by Eq. (1.3).

$$\frac{dL}{dI} = \frac{K}{I} \tag{1.3}$$

The above equation indicates that the sensitivity of the ear decreases as the intensity of the sound increases.

The intensity of sound I is defined as the amount of sound energy S_E flowing per unit area A in unit time t i.e.,

$$I = \frac{S_E}{A}$$

The unit of energy is joule, the unit of area is square metre and the unit of time is second. So intensity is measured in J/ms or Wm^{-2}. Another term related to intensity is threshold intensity.

Threshold Intensity is defined as the minimum sound intensity which a human ear can easily sense. It is 10^{-12} Wm^{-2}. It is also known as standard or zero intensity. It is an intensity that can just be heard by the human ear at standard frequency of 1000 Hz. The intensity of sound is a relative term and measured with reference to this threshold intensity.

DECIBEL

1.6 The relative intensity or the intensity level of a sound is defined as the logarithmic ratio of intensity I of a sound to the standard intensity I_0. Or it is the difference in the two loudnesses.

If the loudness of the sound waves are L and L_0 given by the expression from the Weber – Fechner law,

$$L = K \log_{10} I$$

$$L_0 = K \log_{10} I_0$$

Difference in the loudness is given by

$$\Delta L = L - L_0$$

$$= K (\log_{10} I - \log_{10} I_0)$$

$$\Delta L = K \log_{10} \frac{I}{I_0}$$

Relative intensity or *intensity level* or *difference in loudness* are the same and written by Eq. (1.4).

$$I_L = K \log_{10} \frac{I}{I_0} \tag{1.4}$$

In Eq. (1.4), if the constant $K = 1$ then the relative intensity or intensity level I_L is expressed in a unit of ***bel***. It is defined as the intensity level of a sound whose intensity is 10 times the standard intensity. As we can see from Eq. (1.4), by putting $K = 1$, $I = 10\, I_0$.

$$I_L = (1) \log_{10} \frac{10\, I_0}{I_0} = 1 \text{ bel}$$

Similarly, 2 bels and 3 bels correspond to the increase in intensity of sound by 100 and 1000 times I_0.

Bel is a large unit, and hence a smaller unit is more practical. This is known as ***decibel*** or dB which is given by

$$1 \text{ dB} = \frac{1}{10} \text{ bel} \tag{1.5}$$

This gives,

$$I_L = 10 \log_{10} \left(\frac{1}{I_0}\right) \text{dB} \tag{1.6}$$

Physical meaning of 1 dB

The importance of 1 dB is that it is the smallest change in intensity level I_L that the human ear can detect. Therefore, let us find the corresponding change in intensity or loudness, i.e., I_L = 1dB,

$$I_L = 10 \log_{10} \frac{I}{I_0} \text{ dB}$$

$$1 \text{ dB} = 10 \log_{10} \frac{I}{I_0} \text{ dB}$$

$$\frac{1}{10} = \log_{10} \frac{I}{I_0}$$

$$0.1 = \log_{10}\frac{I}{I_0}$$

$$\frac{I}{I_0} = 10^{0.1} = 1.26$$

$$I = 1.26\, I_0$$

It means that a change in intensity level of 1 dB changes the intensity by 26%. Further, we can verify that various values of $I = 100\, I_0$, $1000\, I_0$, $100\, I_0$, $1000\, I_0$, give 20 dB, 30 dB, 40 dB, and so on. Thus, when two sounds differ by 20 dB, the louder of them is 100 times more intense than when they differ by 40 dB. The loudest sound that can be heard without pain is about 120 dB known as the *threshold of feeling* or *pain threshold*.

Phons

Two different 60-decibel sounds will not, in general, have the same loudness. Saying that two sounds have equal intensity is not the same thing as saying that they have equal loudness. Since the human hearing sensitivity varies with frequency, it is useful to plot equal loudness curves which show that variation for the average human ear. If 1000 Hz is chosen as a standard frequency then each equal loudness curve can be referenced to the decibel level at 1000 Hz. This is the basis for the measurement of loudness in phons. If a given sound is perceived to be as loud as a 60 dB sound at 1000 Hz then it is said to have a loudness of 60 phons.

60 phons means “as loud as a 60 dB, 1000 Hz tone”

The loudness of complex sounds can be measured by a comparison to 1000 Hz test tones, and this type of measurement is useful for research, but for practical sound-level measurement , the use of filter contours has been commonly adopted to approximate the variations of the human ear.

Sones

The use of the phon as a unit of loudness is an improvement over just quoting the level in decibels, but it is still not a measurement which is directly proportional to loudness. Using the rule of thumb for loudness, the sone scale was created to provide such a linear scale of loudness. It is usually presumed that the standard range for orchestral music is about 40 to 100 phons. If the lower end of that range is arbitrarily assigned a loudness of one sone, then 50 phons would have a loudness of 2 sones, 60 phons would be 4 sones, etc.

SOUND ABSORPTION AND REVERBERATION

1.7

When sound propagates in air or a material medium from its source, we assume that spherical sound waves spread out in an infinite homogeneous medium without loss of energy in accordance with the law of inverse squares. This is only partially justified because of *dissipation* or *absorption of sound* energy on account of viscous forces, heat conduction and heat radiation, etc. Whenever the layers of a medium are in relative motion, the viscous forces opposing the relative motion must be overcome and this involves the transformation of the mechanical energy of the waves into heat. During the propagation of sound waves, compressions and rarefactions are formed. At a compression there is rise of temperature and in a neighbouring layer in a state of rarefaction, the temperature is correspondingly lowered. These changes are not perfectly adiabatic, hence there is a tendency for the heat to be conducted from the compression to the rarefaction and this passage of heat down a temperature gradient means an increase of entropy and so dissipation of energy occurs. Besides these two causes, there is radiation of heat from a compression to a rarefaction.

A further cause of dissipation is to be found in intermolecular exchanges of energy, either between molecules of different gases or between different degrees of freedom of the same gas.

It is clear that the absorption of sound is essentially the dissipation of sound energy into other forms of energy and ultimately into heat. Two causes are mainly responsible: (i) porosity, and (ii) flexural vibrations. The sound waves are converted into heat by propagation through the interstices of the material and also by vibration of the small fibres of the material. The other reason is when sound waves fall on flexible materials that are not rigidly mounted, the material is set into vibration and the damping forces called into play dissipate the incident sound energy into heat.

It is the porosity which accounts for the high absorption coefficient of felt, rugs, carpets, etc., and a number of such products are now available. It is observed that the absorption coefficient of any given material increases with the thickness up to a certain limiting value. A marked difference is observed in the coefficients of painted and unpainted bricks that shows the importance of porosity. At a given frequency, the absorption coefficient is a function of the angle of incidence. Therefore, average value of absorption coefficient is considered for acoustical designing.

The coefficient of absorption of a material is defined as the ratio of the sound energy absorbed by the surface to that of the total incident sound energy on the surface, i.e.,

$$\text{Absorption coefficient } (\alpha) = \frac{\text{Sound energy absorbed by the surface}}{\text{Total sound energy incident on the surface}}$$

The sound waves falling on the open windows can easily pass through it without any reflection, we assume that the open window is a perfect absorber of sound and the standard absorption is taken as the unit area of an open window as a standard unit of absorption.

We know that when a source of sound emits sound waves inside a closed building, the waves are generally reflected repeatedly by walls, the ceiling and objects in the building. The intensity of the sound waves decreases at every reflection and finally, it becomes too low to be audible. This repeated reflection and persistence of sound even after the source of sound has stopped is called *reverberation*. The duration for which the sound persists is called the *reverberation time*. It is the time taken for the sound to fall below the minimum audibility, measured from the instant when the source stops emitting sound.

When sound waves cross any given plane, the average rate at which energy flows along with the wave across unit area of the plane normally is called the *intensity* of the sound wave. The intensity is related to the loudness of the sound heard by the ear placed at that plane. The Weber–Fechner law states that the loudness is proportional to the logarithm of the intensity. If S is the loudness and I is the intensity then

$$S = K \log I \tag{1.7}$$

where K is a constant depending on the sensitiveness of the ear, quality of the sound, and other factors.

Now the average rate of flow of energy per second or power W transmitted by a wave is compared to a standard power W_0 by using a unit called the *bel*.

$$\text{Power } W = \log_{10}\left(\frac{W}{W_0}\right)\text{bel} \tag{1.8}$$

If a power W is 10 times a standard power W_0, then it is equivalent to 1 bel. A commonly used unit of power ratio is the *decibel*, which is equal to 1/10 of a bel.

$$\therefore \quad 1 \text{ bel} = 10 \text{ decibels} = 10 \text{ dB}$$

$$\therefore \quad \text{power } W = 10\log_{10}\left(\frac{W}{W_{10}}\right)\text{dB} \tag{1.9}$$

SABINE'S FORMULA

1.8 Sabine used an organ pipe to define reverberation time. He took an organ pipe of frequency 512 Hz and found that sound became inaudible when its intensity fall to 10^{-6} times its intensity just before the source is cut off. He noticed that the time of reverberation depends upon the size of the hall, loudness of the sound and upon the kind of the music or sound for which hall is to be used. The derivation of the formula for reverberation time is as follows.

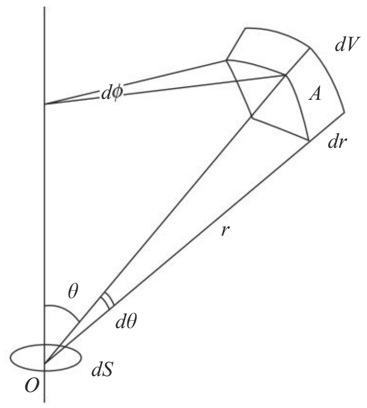


Fig. 1.3 *Showing the small volume element dV and surface area element dS*

Consider a room of volume V filled with sound energy produced by a source. Assume that there is a uniform distribution of sound energy in the given volume of the room. At some instant of time let the energy density be E. Consider a volume element dV around a point A as shown in Fig. 1.3, whose spherical polar coordinates are r, q, f.

The volume element is given by

$$dV = r^2 \sin\theta\, dr\, d\theta\, df \qquad (1.10)$$

So the energy contained within this volume element is EdV. This energy is radiated equally in all directions into a solid angle 4π.

The fraction of energy radiated into a unit solid angle is therefore given by $EdV/4\pi$.

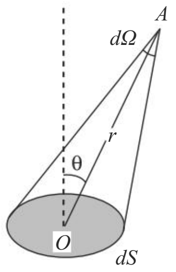


Fig. 1.4 *Showing the solid angled dΩ, and surface element dS*

As shown in Fig. 1.4, the solid angle $d\Omega$ subtended at the point A by a surface element dS around O is given by

$$d\Omega = \frac{dS \cos\theta}{r^2} \qquad (1.11)$$

Therefore, the energy falling on the surface element dS is

$$\frac{E\,dV}{4\pi}\, d\Omega = \frac{E\,dV}{4\pi}\,\frac{dS \cos\theta}{r^2} \qquad (1.12)$$

Now the energy will reach the surface element dS from all directions above it from $\phi = 0$ to 2π, $\theta = 0$ to $\pi/2$. If we now imagine the concentric hemispheres around O with radii ranging from 0 to v, where v is the velocity of sound then the energy contained within this hemisphere ($r = 0$ to v) will have reached dS in one second. Therefore, the total energy incident on the surface element dS in one second is

$$\begin{aligned}
\varepsilon &= \int \frac{E\,dV}{4\pi}\,\frac{dS\cos\theta}{r^2} \\
&= \iiint \frac{E\,dS}{4\pi}\, r^2 \sin\theta\, d\theta\, d\phi\, \frac{\cos\theta}{r^2} \\
&= \int_0^v dr \int_0^{\frac{\pi}{2}} \sin\theta\cos\theta\, d\theta \int_0^{2\pi} d\phi\, \frac{E\,dS}{4\pi} \\
&= \frac{E\,dS}{4\pi} \int_0^v dr \int_0^{\frac{\pi}{2}} \sin\theta\cos\theta\, d\theta\, [2\pi]
\end{aligned}$$

$$= \frac{E\,dS}{4\pi}[2\pi]\int_0^v dr\left[\frac{1}{2}\right]$$

$$= \frac{E\,dS}{4\pi}[2\pi]\left[\frac{1}{2}\right]v$$

$$\varepsilon = \frac{Ev\,dS}{4} \tag{1.13}$$

If we consider the absorption coefficient of the surface element dS to be a then the rate at which the energy is absorbed by it is ($a\,Ev\,dS/4$) Summing over all absorbing surfaces in the room we get the

$$\text{Rate of absorption of energy} = \frac{Ev}{4}\sum a\,dS \tag{1.14}$$

Let the source of sound be supplying energy at a steady rate of P. Since the energy density in the room is E, the total energy in the room is EV. Therefore, the rate at which this total energy is changing is given by

$$\frac{d}{dt}(EV) = V\frac{dE}{dt} \tag{1.15}$$

Therefore, we have at any instant of time

Rate of change of energy in the room (1.16)

= Rate at which the energy is supplied by the source

– Rate at which energy is absorbed by the absorbing surfaces

$$\therefore \qquad V\frac{dE}{dt} = P - \frac{Ev}{4}\sum a\,dS$$

$$\therefore \qquad V\frac{dE}{dt} = -\frac{v\sum a\,dS}{4}\left[E - \frac{4P}{v\sum a\,dS}\right] \tag{1.17}$$

Now let

$$\left[\frac{E - 4P}{4\sum a\,dS}\right] = E' \tag{1.18}$$

Then, $\frac{dE}{dt} = \frac{dE'}{dt}$, since $\frac{4P}{v\sum a\,dS}$ = constant

$$\therefore \qquad V\frac{dE'}{dt} = -\frac{\mathrm{v}\sum a\,dS}{4}E'$$

or

$$\frac{dE'}{dt} = -\frac{v\sum a\,dS}{4V}E'$$

or

$$\frac{dE'}{E'} = -\frac{v\sum a\,dS}{4V}dt$$

On integrating we get,

$$\log E' = -\frac{v\sum a\,dS}{4V}t + K \tag{1.19}$$

where K is constant of integration.

Let
$$\frac{v \sum a\, dS}{4\, V} = \sigma \tag{1.20}$$

Then
$$\log E' = -\sigma t + K$$

∴
$$E' = e^{-\sigma t + K} = e^{-\sigma t} \,.\, e^{K}$$

Substituting for E', we get

$$\left[E - \frac{4\, P}{v \sum a\, dS}\right] = e^{-\sigma t}.e^{K}$$

or

$$E - \frac{P}{\sigma V} = e^{-\sigma t}.e^{K} \tag{1.21}$$

Let $E = 0$ at $t = 0$. Then,

$$-\frac{P}{\sigma V} = e^{K}$$

∴
$$E - \frac{P}{\sigma V} = e^{-\sigma t}.\left(-\frac{P}{\sigma V}\right)$$

or

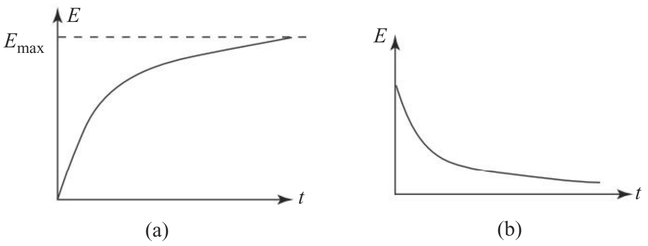
$$E = \frac{P}{\sigma\, V}(1 - e^{-\sigma t}) \tag{1.22}$$

The maximum value of E is reached when $t = \infty$. We get,

$$E_{\max} = \frac{P}{\sigma V} \tag{1.23}$$

∴
$$E = E_{\max}\,(1 - e^{-\sigma t}) \tag{1.24}$$

Thus, sound energy grows exponentially as shown in Fig. 1.5(a) below.



(a) (b)

Fig. 1.5 *(a) Sound energy grows; (b) Decay of sound energy*

Now let us assume that the maximum energy density has been attained in the room. Again let the source of sound be switched off at that instant. We shall now put the new initial condition, namely, $P = 0$, $E = E_{\max}$ at $t = 0$, the equation $a = -w^2 y$. We get

$$E_{\max} = e^{K}$$

Since $P = 0$, for $t > 0$ also, we get

$$E = e^{-\sigma t} E_{\max} \tag{1.25}$$

Thus, the energy density decreases exponentially with time, as shown in the Fig. 1.5(b).

Now reverberation time (T) is defined as the time (measured from the instant when the source is switched off) at which $E = 10^{-6} E_{\max}$. Substituting $t = T$, $E = 10^{-6} E_{\max}$ in the above equation, we get

$$10^{-6} E_{\max} = e^{-\sigma T} E_{\max}$$

or

$$e^{-\sigma T} = 10^{-6}$$

Taking logarithms to the base e on both the sides, we get

$$\log (e^{-\sigma t}) = \log (10^{-6})$$

$$-\sigma T = -6 \times 2.303$$

$$T = \frac{6 \times 2.303}{\sigma} = \frac{13.818 \times 4\, V}{v \sum a\, dS}$$

Assuming the velocity to be 330 m/s, we get

$$T = \frac{6 \times 2.303}{\sigma} = \frac{13.818 \times 4\, V}{330 \sum a\, dS} = \frac{0.1675\, V}{\sum a\, dS}$$

So the reverberation time (T) is

$$T = \frac{0.1675\, V}{\sum a\, dS} \tag{1.26}$$

where V is in m^3, and dS is in m^2. If V is in ft^3 and dS is in ft^2, the reverberation time is

$$T = \frac{0.05\, V}{\sum a\, dS} \text{ (approx.)} \tag{1.27}$$

ACOUSTICS OF BUILDINGS

1.9

If the reverberation time of a room is too short then the sounds emitted in that room appear to be 'dead' or 'flat'. If the reverberation time is too long, the words become unintelligible, since each syllable continues to be heard even after the next syllable has been uttered. For good acoustics of a room, the reverberation time should be neither too short nor too long. The actual values depend on the nature of the sounds. For speech the reverberation time should be about 1–2 seconds. For music, it should be a little longer.

The desired value of reverberation time can be obtained by the suitable arrangement of absorbing materials like upholstery, curtains, etc., and by using special 'acoustic plasters', bricks, tiles, etc.

In addition to a suitable reverberation time, the distribution of sound in a proper manner is also necessary for good acoustical quality of a room. For this purpose, a glass or plastic model of the room is made. Strong sound pulses are produced in different parts of the model, and 'spark photographs' of the model, showing the distribution of sound, are obtained. After studying the distribution of sound, the model is redesigned, if necessary.

Conditions for Good Acoustical Design of Rooms or Factors Affecting Acoustics of Buildings and their Remedies

(1) Control of Reverberation As reverberation is produced due to repeated reflections from the walls of the room, the remedy lies in increased absorption which may be brought about by the following:

(i) Keeping a few windows open will let the entire sound energy out and serve as perfect absorbers.

(ii) Walls covered with absorbent materials like celotex, perforated cardboards, coarse cloth, maps and pictures, felt, asbestos, heavy curtains, etc., or walls engraved and roughened with decorative materials to diffuse the sound. Porous bodies having small capillaries absorb a good deal of sound energy incident on them and serve very well as sound suppressors.

(iii) Upholstered seats are responsible for absorption and avoid approximately the effects of the presence or absence of the audience.

(iv) A large audience (one listener is equivalent to 4.7 sq ft of an open window (women being rather better absorbers than men owing to the nature of their clothing).

(v) Reduction of reverberation is also achieved by covering the floor with carpets and using acoustic tiles.

(2) Shape of Walls and Ceiling There should be no curved walls or corners bounded by two walls for two reasons—first of all, to avoid focusing and concentration of sound at one place and secondly to avoid 'dead' spaces. Concave rear walls are responsible for troublesome echoes and delayed reflection in many cinema halls and auditoria. These detrimental reflections can be converted into useful ones by introducing a ceiling splay between the ceiling and the rear wall.

(3) Concave Surfaces and Balconies Concave surfaces like domes, curved arches and barreled ceiling should be avoided unless given careful acoustical design in order to eliminate focusing effects, nondistribution of sound or echoes. The balconies by rule should have a shallow depth and high opening. The depth should not exceed twice the height of the opening. If these conditions are achieved, the sound flows readily into the space under the balcony.

(4) Floor Plan with Diverging Side Walls The reflections from the ceiling and the walls can be utilised for the establishment of a higher sound level at the rear of a hall. For the proper angle for the ceiling and the side wall reflecting surfaces, the geometrical law of reflection can be used so that they will guide the sound to those seats where the sound level is not adequate.

Flutter echoes frequently occur in rooms having smooth, hard and parallel walls. They can be reduced by the use of diffusing elements on the walls or by the use of diverging or tilted walls or splays.

(5) Seats To secure good distribution of sound, the general plan of the hall with seats should be such that the speaker should be at the focus of a paraboloid reflecting surface and a plane beam of sound can be sent down the hall. Moreover, from the standpoint of hearing as well as seeing, the seats should be gradually elevated in order to promote a free flow of a direct sound from the source to the listeners.

In case of a room having bad acoustical properties, the person should speak slowly so as to produce as little volume of sound as is just sufficient to fill the room.

We should note that by decreasing the time of reverberation, the steady value of sound also decreases. Hence one would not like the sound energy to be too low to be audible, i.e., to make the room feel 'dead'. So the time of reverberation should not be decreased beyond a certain optimum value as decided by experience.

Concluding Remark Unless we study the basic concepts of the above topics like SHM, forced and damped vibrations, the type of the sound waves and the equation of motion of sound waves, we cannot understand how a good auditorium or a cinema hall may be designed. Even our classrooms or halls for indoor games are based on the principles stated above. The reverberation time is a key factor in designing a good acoustical structure. This knowledge is the backbone of civil engineering, structural and architectural engineering.

Solved Problems

1 *The sound from a drill gives a noise level of 90 decibels at a point a few feet away from it. What is the noise level at this point when four such drills are working at the same distance away?*

Solution Let the power transmitted by the sound waves from one drill be W. Let W_0 be a standard power. Then the power ratio is

$$P = 10\ \log_{10}\left(\frac{W}{W_0}\right) \text{ decibels} = 90 \text{ decibles}$$

When four drills are working, the power transmitted will be 4 W. The power ratio is

$$P' = \log_{10}\left(\frac{4W}{W_0}\right) \text{ decibels}$$

$$= 10\log_{10} 4 + 10\log_{10}\left(\frac{W}{W_0}\right)$$

$$= (10 \times 0.6) + 90$$

$$= 96 \text{ dB}$$

This gives the noise level when four drills are working simultaneously.

2 *A lecture hall with a volume of 45000 ft^3 is found to have a reverberation time of 1.5 seconds. What is the total absorbing power of all surfaces in the hall?*

Solution The reverberation time is given as

$$T = \frac{0.05\ \text{V}}{\Sigma adS} \text{ (approx.)}$$

where T is the reverberation time, V is the volume of hall in ft^3, and ΣadS is the total absorbing power of all the surfaces. Given, $V = 45000$ ft^3, $T = 1.5$ s.

The total absorbing power of all surfaces in the hall

$$\Sigma adS = \frac{0.05\ V}{T} = \frac{0.05 \times 45000}{1.5} = 1500 \text{ FPS units}$$

3 *A hall with a volume of 56000 cu. ft is found to have a reverberation time of 2 seconds. If the area of the sound-absorbing surface be 7000 sq. ft, calculate the average absorption coefficient.*

Solution By Sabine's formula, the reverberation time is given as

$$T = \frac{0.05\ V}{A}$$

where

$$A = \sum a\,dS = a \sum dS = a\,S$$

where T is the reverberation time, V is the volume of the hall in ft^3, and $\sum a\,dS$ is the total absorbing power of all surfaces. Given, $S = 7000$ sq. ft, $V = 56000$ ft^3, $T = 2$ s.

$$T = \frac{0.05\ V}{A} = \frac{0.05\ V}{a\,S}$$

$$\therefore \quad a = \frac{0.05\ V}{TS} = \frac{0.05 \times 56000}{2 \times 7000} = 0.2$$

4 *A hall has a volume of 80000 cu. ft. Its total absorption is equivalent to 1000 sq. ft. of open window. What will be the effect on the reverberation time if an audience fills the hall and thereby increases the absorption by another 1000 sq. ft.?*

Solution $V = 80000$ cu. ft., $S = 1000$ sq. ft.,

From Sabine's formula, we have the reverberation time T given by,

$$T = \frac{0.05\ V}{A} = \frac{0.05\ V}{a\,S}$$

$$\therefore \quad T_1 = \frac{0.05\ V}{a\,S} = \frac{0.05 \times 80000}{a \times 1000} = \frac{4}{a}$$

When the audience fills the hall, the total absorption will become equal to (1000 + 1000) = 2000 sq.ft. of open window.

$$T_2 = \frac{0.05\ V}{a\,S} = \frac{0.05 \times 80000}{a \times 2000} = \frac{2}{a}$$

$$\therefore \quad \frac{T_2}{T_1} = \frac{2/a}{4/a} = \frac{1}{2}$$

Thus, the reverberation time is reduced to one-half of its initial value.

5 *A hall has a volume of 80000 ft^3 and its total absorption is equivalent to 1000 ft^3 of open window. How many persons should be seated in the hall so that the reverberation time becomes 2 seconds, given that the absorption area of one person is equivalent to 2 sq. ft. of open window? Calculate the reverberation rime of the empty hall also.*

Solution Let n be the number of persons seated in the hall.

Therefore, total absorptions due to n persons is equivalent to $2n$ sq. ft.

Total absorption of the hall = A = $1000 + 2n$ sq.ft.

$$T = \frac{0.05\ V}{A}$$

$$2 = \frac{0.05 \times 80000}{1000 + 2n}$$

$$2(1000 + 2n) = 0.05 \times 80000$$

$$n = 500 \text{ persons}$$

When the hall is empty, $A = 1000$ sq. ft., so

$$T = \frac{0.05 \times 80000}{1000} = 4 \text{ s}$$

6 *The intensity of sound produced by thunder is 0.1 Wm^{-2}. What is the relative intensity in dB?*

Solution $I_0 = 10^{-12}\ \text{Wm}^{-2}, I = 0.1\ \text{Wm}^{-2}, I_L = ?$

$$I_L = 10 \log_{10}\left(\frac{0.1}{10^{-12}}\right) \text{dB}$$

$$= 10 \log_{10}(10^{11})$$

$$= 10 \times 11$$

$$= 110 \text{ dB}$$

7 *Calculate the change in intensity level when the intensity of sound increases 100 times its original intensity.*

Solution Given:

Initial intensity = I_0

Final intensity = I

$$\frac{I}{I_0} = 100$$

Increase in intensity level = L

$$\therefore \quad L = 10 \log_{10}\left(\frac{I}{I_0}\right) \text{(in dB)}$$

$$\therefore \quad L = 10 \log_{10} 100 = 20 \text{ dB}$$

8 *Calculate the change in intensity level when the intensity of sound increases 1000 times its original intensity.*

Solution Initial intensity = I_0

Final intensity = I

Ratio of $I/I_0 = 1000$

Increase in intensity level =

$$I_L = 10 \log_{10} \frac{I}{I_0} \text{ dB}$$

$$I_L = 10 \log_{10} 10^3 \text{ dB}$$
$$= 10 \times 3 = 30 \text{ dB}$$

Multiple-Choice Questions

1. Which of the following gives the range of audible sound?
(a) 0 Hz to 20 Hz (b) 20 Hz to 20 kHz
(c) 20 kHz to 20 MHz (d) 0 Hz to 200 kHz

2. Which of the statements is true?
(a) Sound waves move faster in air than in solids.
(b) Sound waves move with equal speed in all media.
(c) Sound waves move faster in solids than in air.
(d) It cannot be said with certainty, in fact it depends on the wavelength and frequency.

3. The velocity of sound
(a) does not depend on the temperature of the medium
(b) does not depends on the nature of the medium
(c) depends on both the nature and temperature of the medium
(d) both (a) and (b) are correct.

4. The following waves have frequencies greater than those of audible waves:
(a) Supersonic
(b) Infrasonic
(c) Ultrasonic
(d) Both (a) and (c) are correct

5. The following waves have frequencies less than those of audible waves:
(a) Supersonic (b) Infrasonic
(c) Ultrasonic (d) none of the above

6. The velocity of sound in air
(a) increases with increase in temperature
(b) decreases with increase in temperature
(c) does not depend on the temperature
(d) decreases with decrease in temperature

7. The intensity level of threshold of hearing is
(a) 10^{-12} Wm^{-2} (b) 1 dB
(c) 10 Wm^{-2} (d) 100 dB

8. The sensitiveness of the ear is given by
(a) dL/dI (b) $L = K \log 10\, I$
(c) $I_L = 10 \log 10(I/I_0)$ (d) pitch

9. Reverberation of sound in a hall is
(a) a desirable effect in a controlled way
(b) undesirable effect
(c) sometimes desirable and sometimes not desirable
(d) essential evil

10. Reverberation time is given by
(a) Sabine's formula
(b) Weber – Fechner law
(c) Weber's law
(d) Sabine – Fechner law

ANSWERS TO MCQ'S

1. (b) 2. (c) 3. (c) 4. (d) 5. (b) 6. (d) 7. (a) 8. (a)
9. (a) 10. (a)

Practice Problems

1. Calculate the intensity level in dB if the intensity of sound in a street during heavy traffic is 10^{-6} Wm^{-2}. [Ans: 60 dB]

2. Find the intensity level in phons if 3000 Hz with intensity level of 70 dB produces the same loudness as a standard source of frequency 1000 Hz at an intensity level of 67 dB. [Ans: 67 Phons]

3. A hall of volume 1.2×10^5 m^3 has a reverberation time of 1.55 s. If the total absorbing surface is 2.65×10^4 m^2, calculate the average absorbing power of the surface. [Ans: 0.488 sabine]

4. An auditorium has a volume = 7.5×10^3 m^3. If reverberation time is 1.5 s, calculate the total absorption in the hall. [Ans: 8.35×10^2 sabine-m^2]

5. For a given hall having absorption coefficient 0.2 and volume 10^3 m^3 and absorbing surface area 4×10^2 m^2, calculate the reverberation time of the hall. [Ans: 2.088 s]

Short-answer Questions

1. Explain the term ‘decibel’.

2. How is the loudness of sound related to the intensity of the sound wave?

3. State the Weber–Fechner law in sound.

4. What is reverberation?

5. State Sabine’s formula.

6. Can you hear your own words if you are standing in a perfect vacuum? Can you hear your friend in the same conditions?

7. Two loudspeakers are arranged facing each other at some distance. Will a person standing behind one of the loudspeakers clearly hear the sound of the other loudspeaker or the clarity will be seriously damaged because of the ‘collision’ of the two sounds in between?

8. If you are walking on the moon, can you hear the sound of stones cracking beneath your feet? Can you hear the sound of your own footsteps?

9. Why is a given sound louder in a hall than in the open?

10. Draperies and furniture often improve the acoustics of a room; why?

Long-answer Questions

1. Explain clearly what causes reverberation in a hall and how it can be minimised. Derive Sabine’s expression for the reverberation time.

2. Write an essay on ‘The Acoustics of Buildings’.

3. What factors determine good acoustics of a room?

4. Derive an expression for the reverberation time.

5. Discuss the condition for good acoustical design of rooms.

Questions from Previous Exams

Architectural Acoustics

(i) What is standard intensity? Give its value. **(GTU- Jan' 2009)**

The minimum sound intensity which a human ear can sense is called the standard intensity. Its value is 10^{-12} Wm^{-2}.

(ii) Define reverberation time. **(GTU- Jan' 2009)**

The time taken by a sound intensity to fall below the minimum audibility level after the source stops sounding.

(iii) The volume of a room is 600 m^3. The wall area of the room is 220 m^2, the floor area is 120 m^2 and ceiling area is 120 m^2. The average sound absorption coefficient for the wall is 0.03, for the ceiling is 0.8 and for the floor it is 0.06. Calculate reverberation time. **(GTU- Jan' 2009)**

Given

$V = 600$ m^3

Area of the room $S_1 = 220$ m^2, $a_1 = 0.03$

Area of the floor $S_2 = 120$ m^2, $a_2 = 0.06$

Area of the ceiling $S_3 = 120$ m^2, $a_3 = 0.8$

Average absorption coefficient

$$= a_1S_1 + a_2S_2 + a_3S_3/S_1 + S_2 + S_3$$

$$6.6 + 7.2 + 96/460 = 0.2374 \text{ O.W.U.}$$

$$\bar{a}\,\Sigma S = 0.2374 \times 460 = 109.8 \text{ O.W.U.-m}^2$$

$$T = 0.167\text{V}/\bar{a}\,\Sigma S = 0.91 \text{ s}$$

(iv) What is meant by time of reverberation? Discuss Sabine's formula. **(GTU- Jan' 2009)**

(Refer Section: 1.7)

(v) State any five factors affecting the acoustics of the building and give at least two remedies. **(GTU- Jan' 2009)**

(Refer Section: 1.9)

(vi) What is the resultant sound level when a 70-dB sound is added to an 80-dB sound? **(GTU- Jan' 2009)**

Given $I_{L1} = 70$ dB $I_{L1} = 80$ dB

$$I_{L1} = 10 \log_{10} (I_1/I_o) = 70 \text{ dB}$$

$$I_{L2} = 10 \log_{10} (I_2/I_o) = 80 \text{ dB}$$

$$I_1/I_o = 10^7$$

$$I_2/I_o = 10^8$$

$$I = I_1 + I_2 = 10^7\, I_o + 10 \times 10^7\, I_o = 11 \times 10^7\, I_o$$

$$I_L = 10 \log_{10} (I/I_o)$$

$$I_L = 80.41 \text{ dB}$$

(vii) Explain the various factors affecting the acoustics of a building and give their remedies. **(GTU- Jan' 2010)**

Reverberation time, loudness, focusing, echelon Effect, resonance

(viii) What is absorption coefficient? Explain how you can determine the absorption coefficient. **(GTU- Jan' 2010)**

The sound-absorbing coefficient of a material is defined as the ratio of sound energy absorbed by it to the total sound energy incident on it.

Suppose T_1 is reverberation time an for empty hall, so $T_1 = 0.167\ \mathrm{V}/\Sigma aS$

Suppose T_2 is reverberation time of hall with absorbing material, so $T_2 = 0.167\ \mathrm{V}//\Sigma aS + a_1S_1$

Therefore, $1/T_2 - 1/T_1 = a_1S_1/0.167\ \mathrm{V}$ and hence $a_1 = 0.167\ \mathrm{V}/S_1\ [1/T_2 - 1/T_1]$ Sabine.

(ix) A cinema hall has a volume of 7500 m^3. What should be the total absorption in the hall if the reverberation time of 1.5 seconds is to be maintained? **(GTU- Jan' 2010)**

$T = 0.167\ \mathrm{V}/\Sigma as$

$T = 1.5$ seconds

$V = 7500\ \mathrm{m}^3$

$\Sigma\ as = ?$

$\Sigma as = 835$ Sabin-m^2

(x) What is the resultant sound level when a 70 dB sound is added to an 85-dB sound?

Given $I_{L1} = 70$ dB $I_{L1} = 85$ dB

$I_{L1} = 10 \log_{10} (I_1/I_o) = 70$ dB

$I_{L2} = 10 \log_{10} (I_2/I_o) = 85$ dB

$I_1/I_o = 10^7$

$I_2/I_o = 10^{0.5}10^8 = 3.16\ (10^8)$

$I = I_1 + I_2 = 10^7\ I_o + 3.16 \times 10^8\ I_o = [1 + (3.16)\ 10] \times 10^7\ I_o$

$I = (32.6)\ 10^7\ I_o$

$I_L = 10 \log_{10} (I/I_o)$

$I_L = 85.13$ dB

(xi) Derive Sabine's formula for reverberation time. **(GTU- June' 2010)**

(Refer Section: 1.8)

(xii) What are the characteristics of musical sound? Explain them in details. **(GTU- June' 2010)**

(Refer Section: 1.4)

The McGraw-Hill Companies

Ultrasonics

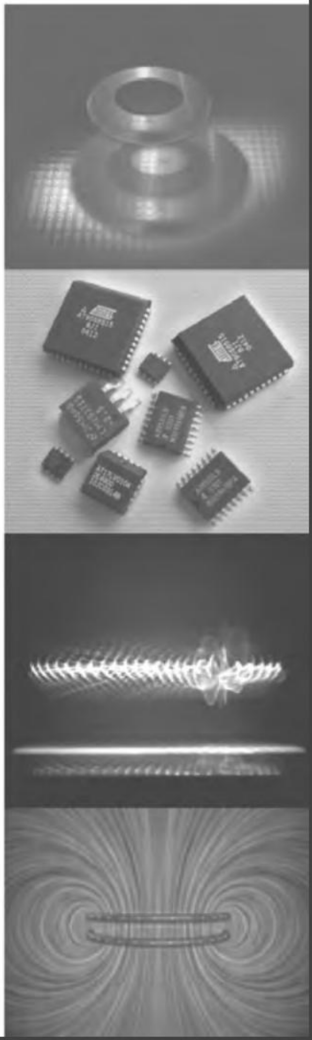
2

If you are on the ground and want to know the speed and direction of an aircraft in the sky or a steamer in the ocean, is it possible? A wall of a dam is leaking. Can you detect the position of leakage and the amount of cavity in the wall of the dam? The following chapter will provide the techniques to solve such problems.

ULTRASONIC WAVES

2.1

Sound is the mechanical vibrations of particles of the material medium. The human ear is sensitised to hear the sound waves between 20 Hz to 20 kHz, which is known as the *audible range.* Sound waves with frequencies above 20 kHz (above the audible range) are known as *ultrasonic waves.* The energy of ultrasonic waves is large on account of high frequency; so they can propagate through larger distance. The propagation also depends on the elastic properties and the density of the medium. The is highest in solids and lowest in gaseous media but it cannot propagate in vacuum. Due to their smaller wavelength as compared to audible sound, a large number of ultrasonic applications are possible in almost all branches of technology, engineering and medicine. There are three types of ultrasonic waves depending on the mode of vibration of the particles of the medium with the propagation direction of the wave. They are (i) longitudinal or compressional type of waves, (ii) transverse or shear waves, and (iii) surface or Rayleigh waves. The very first ultrasonic devices were developed early in the 20th century when the piezoelectric effect was discovered by Jacques and Pierre Curie. Our aim is how to generate ultrasonic waves of appropriate frequency and power, how to detect their presence in the given medium and how to manipulate them for various applications in different fields of science, technology and medicine.



PRODUCTION OF ULTRASONIC WAVES

2.2 To produce ultrasonic waves of frequencies greater than 20 kHz, the usual method is not helpful. One reason is that in the method of a diaphragm loudspeaker fed with alternating current at very high frequencies (greater than 20 kHz), the inductive effect of the loudspeaker coil is so large that practically no current passes through it. The second reason is that the diaphragm of a loudspeaker cannot vibrate at such high frequencies. So other appropriate methods are to be used for the production of ultrasonic waves.

There are two important methods based on two principles, namely, magnetostriction and piezoelectric effect. The magnetostriction method is commonly used for frequencies up to 300 kHz and the piezoelectric generators are used for even higher frequencies, i.e., for frequencies up to few hundred MHz. First, we shall discuss these two principles along with their production methods.

2.2.1 Magnetostriction Principle

When a rod of ferromagnetic material, like nickel or iron, is placed in a magnetic field parallel to its length (Fig. 2.1), a small change in the length, of the order of 1 ppm, occurs. This change of length is independent of the direction of the field and only depends upon the magnitude of the field and nature of the material. This is known as the *magnetostriction phenomenon.* Out of many existing ferromagnetic material, large magnetostriction effect is exhibited by nickel. So it is preferred most.

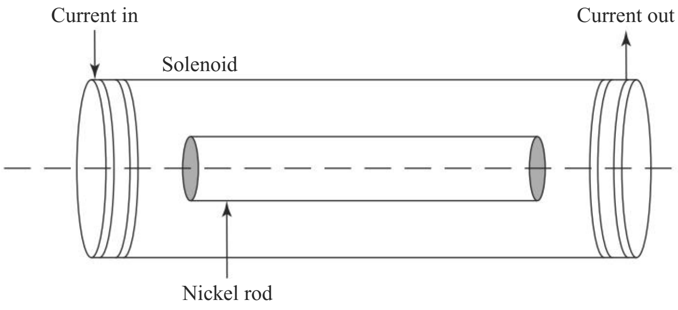


Fig. 2.1 *Magnetic field is parallel to the axis of solenoid and length of nickel rod*

Longitudinal ultrasonic waves with frequencies greater than 20 kHz, can be produced with this method. For this purpose, we apply an alternating magnetic field parallel to the axis of the ferromagnetic rod. For producing an ac magnetic field, a rod is kept along the axis of a solenoid and alternating current is fed to the solenoid which produces an alternating magnetic field parallel to its axis. The frequency of the alternating current and hence ac magnetic field is same. In each half cycle of ac current or ac magnetic field, the length of the rod changes, hence longitudinal vibrations in the rod are set up with a frequency twice that of the applied frequency. If the rod is not magnetised initially then the resulting changes like elongation or contraction depends on the type of the material and are independent of the direction of the magnetic field. If the natural frequency of vibration of the rod is equal to the frequency of ac magnetic field, resonance occurs and produces maximum amplitude of vibrations in the rod. If the frequency f of ac field is in the ultrasonic range, the medium surrounding the rod (e.g., air) will also vibrate with a frequency $2f$, hence ultrasonic waves are produced and propagate in air medium.

The frequency of desired value can be generated by knowing the following data of the material of the rod. If Y is the Young's modulus of the material, ρ is the density of the material, L is the length of the rod and n is the mode of vibrations under consideration, the frequency f of ultrasonic waves is given by the following relation.

$$f = \frac{n}{2L}\sqrt{\frac{Y}{\rho}} \tag{2.1}$$

where $n = 1, 2, 3, \ldots$

Here, it is desired that the frequency of the ac current and frequency of the vibration of the rod should be same to achieve resonance condition. For this purpose, the rod should be initially magnetised by a dc current through a second coil. If this dc magnetic field is stronger than the ac magnetic field, the frequency of the vibration of the bar will be equal to the applied frequency and hence resonance occurs.

2.2.2 Magnetostriction Method

We require an alternating current of frequency greater than 20 kHz. For this purpose, the Colpitt's oscillator circuit is most suitable. This circuit consists of a tank circuit made up of two capacitors C_1 and C_2 and a common inductor coil L connected, as shown in Fig. 2.2.

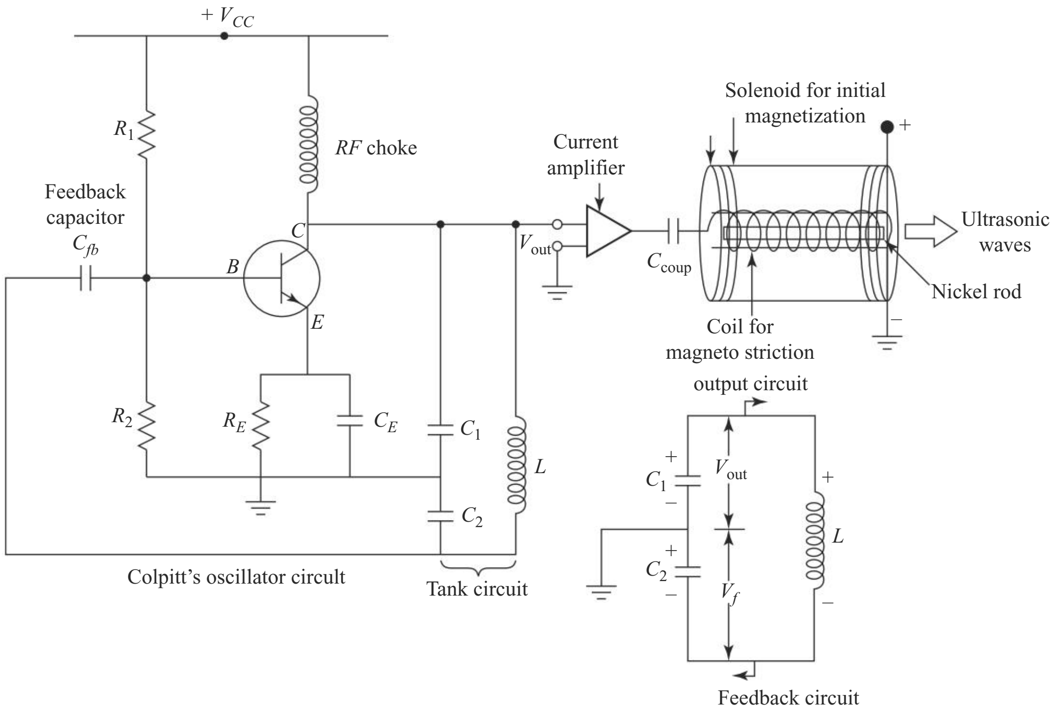


Fig. 2.2 *Magnetostriction method for production of ultrasonic waves*

The frequency of oscillations is given by

$$f = \frac{1}{2\pi\sqrt{C_T L}} \tag{2.2}$$

where,

$$C_T = \frac{C_1 C_2}{C_1 + C_2} \tag{2.3}$$

Note also that C_1-C_2-L is also a feedback circuit that produces a phase shift of π radian. The radio frequency choke decouples any ac signal on the power lines from affecting the output signal.

When the circuit is turned ON, two capacitors C_1 and C_2 of the tank circuit are charged. They get discharged through the inductor coil L, again charged with opposite polarity and this way set up oscillations of frequency given by Eq. (2.2). The output voltage of the transistor amplifier appears across the capacitor C_1 and feedback voltage is developed across the capacitor C_2 of the tank circuit. V_{out} across C_1 and V_f across C_2 are 180° out of phase. The feedback voltage to the transistor provides a positive feedback. The transistor provides 180° phase shift and a further phase shift of 180° is provided by the $C_1 - C_2$ voltage divider. So feedback is properly phased to produce continuous undamped oscillations in the tank circuit.

The transistor is biased with the help of the resistances R_1, R_2, *RF Choke* and R_E. The oscillations are fed back to the transistor base through the feedback capacitor C_{fb}. The appropriate frequencies are amplified by transistor action and the oscillations corresponding to them are sustained. Now the oscillations at the output terminals of the tank circuit are fed to operational amplifier/current amplifier to raise the level of oscillations. The output of the operational amplifier is given to the magnetostriction coil through the coupling capacitor C_{coup}. There is one more coil through which a dc current and hence a dc magnetic field is applied to initially magnetise a nickel rod and this magnetic field is higher than the ac magnetic field which sets up vibrations in the nickel rod. The resonance condition is achieved by keeping the frequency of the ac field (given by Eq. (2.2)) equal to the natural frequency of vibration the of rod (given by Eq. (2.1)). Hence, ultrasonic waves in the surrounding medium are produced. By suitably selecting the material and length of the rod, the desired frequency of the ultrasonic waves can be produced.

The *advantages* of this method are that a large output power can be produced and materials used in this method are inexpensive. The *disadvantages* are that frequencies greater than 100 kHz and single frequency oscillations cannot be generated.

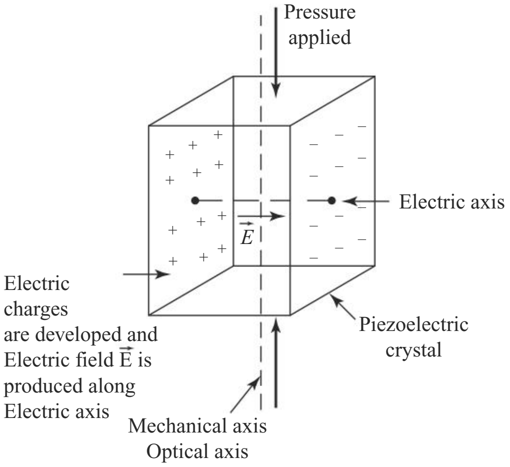


Fig. 2.3 *Piezoelectric principle for producing ultrasonic waves*

2.2.3 Piezoelectric Principle

When certain crystals (known as piezoelectric crystals) like quartz, ammonium phosphate, Rochelle salt, tourmaline, etc. (Fig. 2.3), are stretched or compressed along a mechanical axis, an electric potential difference is produced along an electrical axis.

It is also observed that when an alternating potential difference is applied along the electric axis, the crystal is set into mechanical vibrations along the corresponding mechanical axis. If the frequency of electric oscillations is

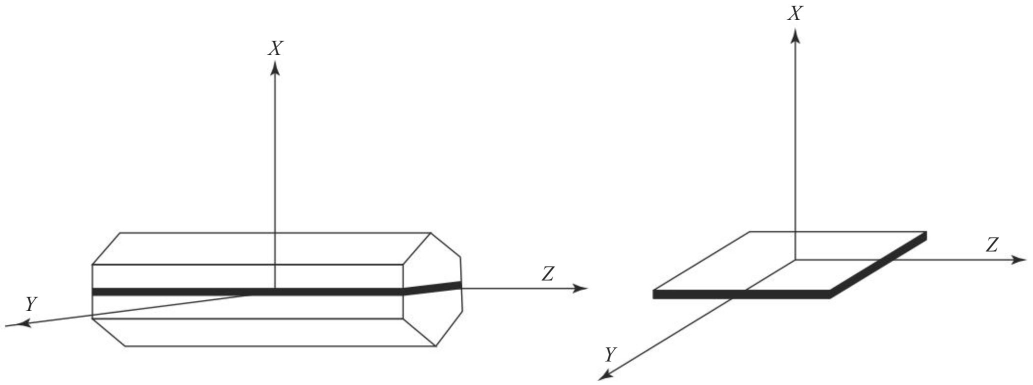


Fig. 2.4 *(a) Quartz crystal; (b) X-cut Quartz crystal*

equal to the natural frequency of vibrations of the crystal, the resonance occurs resulting into vibrations of large amplitude. This phenomenon is known as *piezoelectric effect.* This phenomenon can also be used for the production of ultrasonic waves. In the surrounding air medium, ultrasonic waves are produced depending on the frequency of the crystal oscillator. The alternating potential difference is obtained by a Hartelly oscillator.

Quartz crystal is widely used as transducers to produce ultrasonic waves. The natural quartz crystal has the shape of a hexagonal prism with a pyramid attached to each end as shown in the Fig. 2.4(a).

The crystal should be cut in a particular direction and a thin slab is taken to make the transducer. The *z*-axis is taken along the longest dimension of the natural crystal and it is known as the *optic axis.* The three lines which pass through the opposite corners of the crystal forming its three *x*- axes, are known as *electrical axes.* Similarly, the three lines which are perpendicular to the sides of the hexagon form the three *y*-axes, which are known as *mechanical axes.* Thin plates of the quartz crystal cut perpendicular to one of its *x*-axis are known as *x-cut plates.* It generates longitudinal mode of ultrasonic vibrations of frequencies up to several hundred kHz. Similarly, thin plates cut perpendicular to one of its *y*-axis are known as *y-cut plates.* It generates transverse mode of vibrations of frequencies in the range 1 MHz to 10 MHz.

The frequency of the length vibrations of an *x*-cut crystal is given by

$$f = \frac{n}{2t}\sqrt{\frac{Y}{r}} \tag{2.4}$$

where n is an integer, Y is Young's modulus along the appropriate direction, ρ is the density of the crystal and t is the thickness of the crystal.

2.2.4 Piezoelectric Method

The experimental arrangement is shown in Fig. 2.5.

The circuit diagram of a piezoelectric ultrasonic generator using a transistor is basically a Hartley oscillator which is similar to the Colpitt's oscillator with some modification. Here, two inductor coils L_1 and L_2 are placed across a common capacitor C and the centre of the inductor L_1 and L_2 is tapped and earthed.

When the circuit is switched ON, the capacitor of the tank circuit gets charged. A tank, or oscillatory circuit is formed by two inductor coils L_1 and L_2 and a capacitor connected in parallel to it. After it is fully charged, the capacitor discharges through coils L_1 and L_2 and again gets charged with opposite polarity. This way, oscillations are set up of frequency f given by Eq. (2.5).

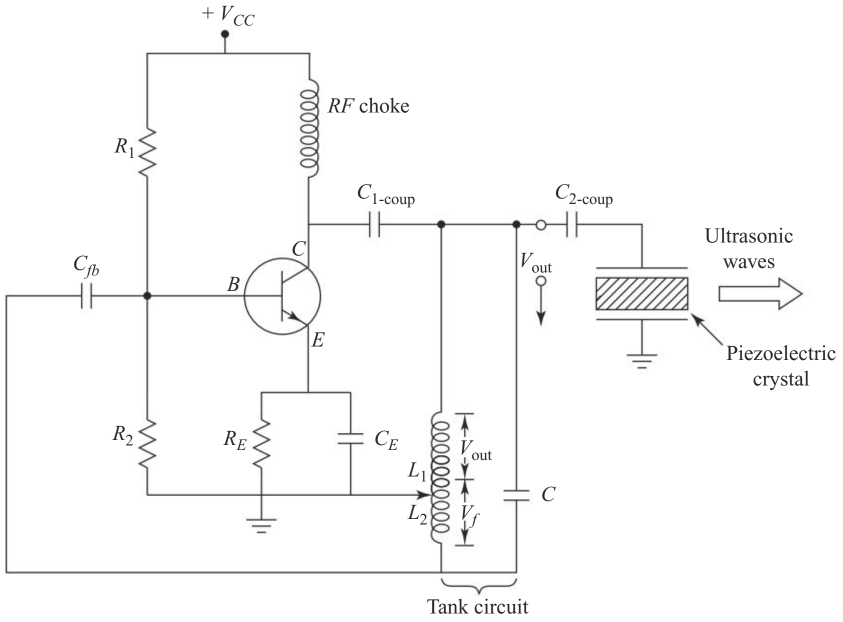


Fig. 2.5 *Piezoelectric method for production of ultrasonic waves*

$$f = \frac{1}{2\pi\sqrt{CL_T}} \tag{2.5}$$

where, total inductance is

$L_T = L_1 + L_2 + 2M$

M = mutual inductance between L_1 and L_2. Note here that

$L_1 - L_2 - C$ is also the feedback network producing a phase change of π radian.

Now output voltages of the transistor amplifier appear at the collector and hence across the inductor L_1 and feedback voltage across the inductor coil L_2. V_{out} and V_{fb} are 180° out of phase. It provides positive feedback to the transistor. The transistor also produces a 180° phase shift and 180° phase shift by an $L_1 - L_2$ voltage divider. In this way, feedback is properly phased to produce continuous undamped oscillations.

The transistor is biased using R_1, R_2, R_E and *RF* choke (radio frequency choke). The tank circuit is coupled to the transistor through the coupling capacitor $C_{1\text{-coup}}$. Capacitor C_{fb} provides positive feedback to the transistor amplifier. The oscillating voltage generated by the tank circuit of frequency f given by Eq. (2.5) is applied to the two electrodes (in which the piezoelectric crystal is fixed) through the coupling capacitor $C_{2\text{-coup}}$. The piezoelectric crystal will start vibrating with a frequency f given by Eq. (2.4). If these two frequencies are made equal, the resonance condition is established and maximum amplitude of vibrations occurs. The desired frequency can be obtained by suitably taking the material of appropriate length.

The striking advantages of this method over the magnetostriction method are (i) high frequencies up to 500 MHz can be produced, (ii) a single frequency output can be obtained, and (iii) a large range of frequencies can be covered using different transducers.

DETECTION OF ULTRASONICS

2.3

The ultrasonic waves can be detected by the following methods.

2.3.1 Piezoelectric Detector

This is a quantitative method. The quartz crystal is used as a detector of ultrasonic waves. One pair of faces of the quartz crystal is exposed to ultrasonic waves (Fig. 2.6) if they are present in the air medium.

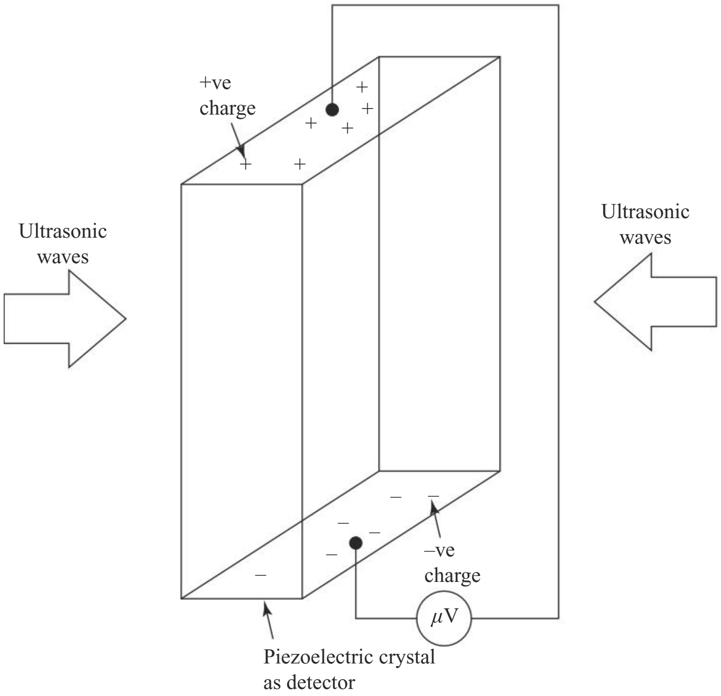


Fig. 2.6 *Piezoelectric detector*

On the other pair of faces, varying electric charges are produced. This generates an alternating voltage of very small value. It is amplified and then detected/measured by some suitable micro-voltmeter. A CRO will be useful to measure both frequency and strength of the ultrasonic signal.

2.3.2 Kundt's Tube Method

This method is useful to detect ultrasonic waves of relatively large wavelength. Lycopodium powder is sprinkled in the tube. When ultrasonic waves are passed through the tube, heaps at the nodal points and pits at anti-nodal points on the surface of the powder are formed [See Fig. 2.7(a)]. So the uniform surface is changed into heaps and pits regularly spread over the surface, which is the indication of the presence of ultrasonic waves.

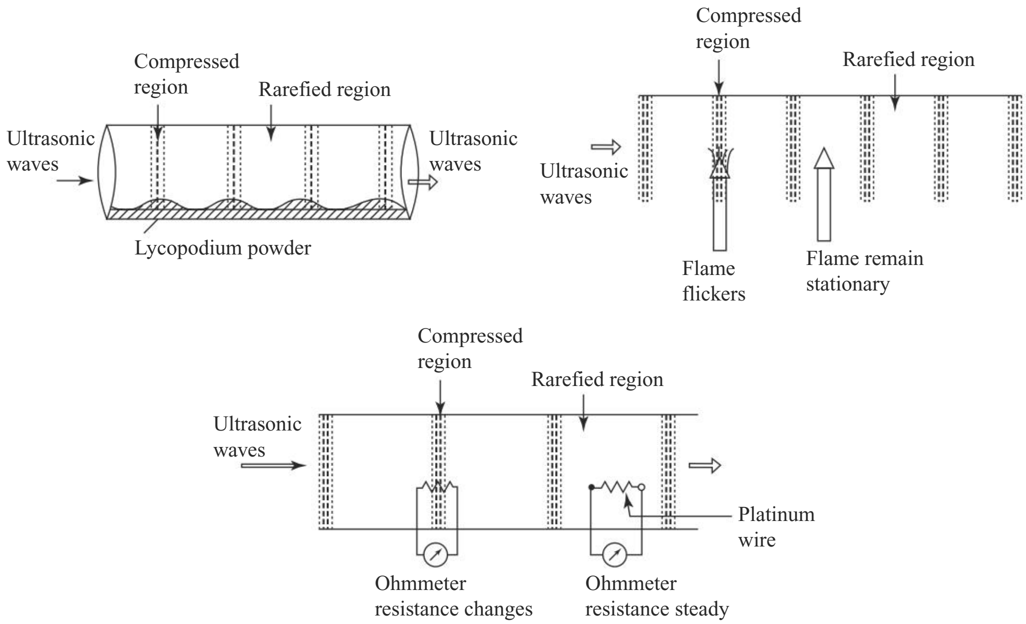


Fig. 2.7 *(a) Kundt's tube method; (b) Sensitive flame method; (c) Thermal detector method*

2.3.3 Sensitive Flame Method

A sensitive flame is moved in a medium where ultrasonic waves are present, the flame flickers at nodes and remains stationary at antinodes [Fig. 2.7(b)].

2.3.4 Thermal Detector Method

Platinum wire is very sensitive to change in the temperature. The resistance changes immediately to any change in the temperature. So a fine platinum wire connected with a bridge circuit is allowed to move in the medium of ultrasonic waves. The temperature of the medium varies due to alternate compressions and rarefactions [Fig. 2.7(c)]. There is a variation of temperature at nodes while at antinodes, the temperature remains constant. So the resistance of the platinum wire varies at nodes and remains constant at antinodes. The variation in the resistance of the platinum wire with respect to time can be detected by using a sensitive bridge arrangement, which will be in the balanced position when the platinum wire is at antinodes.

PROPERTIES OF ULTRASONICS

2.4

(i) Ultrasonic waves have frequencies greater than 20 kHz, so they are highly energetic.

(ii) The speed of propagation of ultrasonic waves depends upon their frequency, i.e., it increases with increase in frequency. It depends also on the medium of propagation. The speed will be highest in the solid medium and lowest in the gaseous medium and intermediate in the liquid medium similar to audible sound waves. The wavelengths are of the order of few millimetres or less.

(iii) Due to the small wavelengths, ultrasonic waves show negligible diffraction. So they can be transmitted over long distances (few km) without any appreciable loss of energy. Audible sound waves cannot be transmitted directly for more than a few metres.

(iv) The intense beam of ultrasonic waves has a disruptive power, so it can cause bubbles in the liquids.

(v) When ultrasonic waves are propagated in a liquid bath from one end, a stationary wave pattern is produced due to the reflection of waves from the opposite end. The density of the liquid thus varies from layer to layer (vertical planes) of the liquid along the direction of propagation. In this way, a plane diffraction grating is formed which can diffract light if it passes through such a medium. This phenomenon can be used to find the wavelength of ultrasonic waves if the wavelength of light is known.

WAVELENGTH OF ULTRASONIC WAVES

2.5

Debye and Sears in America observed the phenomenon of diffraction of light by ultrasonic waves passing through a liquid (1932). When ultrasonic waves are propagated in a liquid kept in a constant-temperature bath, the density varies from layer to layer due to periodic variation of pressure. If a monochromatic beam of light is passed through this liquid incident at right angles to the direction of propagation of waves and tangential to layers, the liquid behaves as a diffraction grating. Such a grating is known as an *acoustical grating.* This acoustical grating behaves in the same way as a ruled transmission grating of a plane glass plate. Hence, this method is used for finding the unknown frequency of the ultrasonic waves and hence wavelength and velocity of the waves in a liquid.

The experimental arrangement is shown in Fig. 2.8. The light from a monochromatic source S of light of known wavelength λ is focused on the narrow slit in the collimator (marked by – O) by the lens L_1. The light from the slit now passes through the convex lens L_2 which turns it in a parallel beam. This parallel beam of light passes through the ultrasonic cell. This ultrasonic cell consists of a rectangular glass tank containing the liquid of known density and is transparent to light.

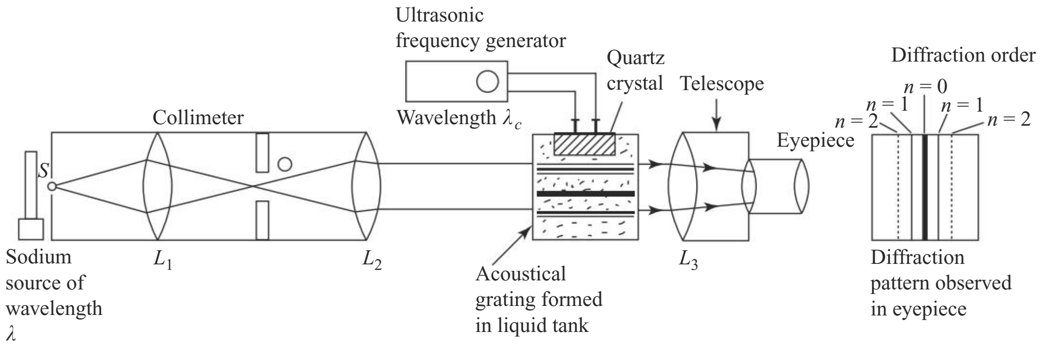


Fig. 2.8 *Schematic representation of the experimental arrangement for measurement of ultrasonic wavelength*

The quartz crystal which produces ultrasonic waves is dipped into the liquid and is fixed at one of the vertical walls of the cell attached to the oscillator. The ultrasonic waves are produced under resonance condition at this wall of the cell and propagate through the liquid. The waves are reflected from the opposite side of the cell and form a stationary wave pattern which acts as an acoustical grating. The piezoelectric crystal is so placed in the cell to form the acoustical grating in a direction perpendicular to the direction of propagation of light. The light emerging from the cell is focused by the lens L_3 and can be seen through the eyepiece of a telescope. Initially, when the crystal is at rest, a single image of the slit is observed. But when the crystal is excited by an oscillator of appropriate frequency, ultrasonic waves are produced in the liquid, the intensity of the central image of the slit decreases and a number of diffracted patterns are observed on both the sides of the centre. The angular separation θ between the direct image of the slit and the diffracted image of n^{th} order is measured. The wavelength of ultrasonic waves can be calculated by using the theory of diffraction grating. It should be noted that the grating element is same as the wavelength of ultrasonic waves. If $\lambda_{\text{ultrasonic}}$ is the wavelength of ultrasonic waves to be found and if λ is the wavelength of the monochromatic light used then

$$\lambda_{\text{ultrasonic}} \sin \theta_n = n\lambda$$

or

$$\lambda_{\text{ultrasonic}} = \frac{n\lambda}{\sin \theta_n} \tag{2.6}$$

In this way, $\lambda_{\text{ultrasonic}}$ can be calculated.

If f be the frequency of ultrasonic waves then the velocity of ultrasonic waves in the liquid can be calculated by following relation

$$v = f\lambda_{\text{ultrasonic}} \tag{2.7}$$

APPLICATIONS OF ULTRASONIC WAVES

2.6

There are numerous applications in almost all fields of science, technology and engineering, medicine, oceanography, astronomy, biosciences etc.

Detection of Flaws or Cracks in Metals Ultrasonic waves can be used to detect cracks or cavities in different metals. We know that flaw/cavity in the metal produces a change in the medium due to which reflection of ultrasonic waves takes place. The speed of ultrasonic waves through cavity in the metal will be different than in the solid region of the metal. So when ultrasonic waves pass through a metal having some hole or cavity inside it, a large amount of reflection occurs. Some reflection also takes place from the back surface of the metal. As shown in Fig. 2.9, from the ultrasonic generator the ultrasonic waves are allowed to fall on the transmitting transducer which is placed on the surface of the metal. Some portion of the waves go straight to the opposite end of the metal and are reflected, and some part of the waves get reflected from the cavity. The reflected pulses from the cavity and from the back surfaces are received by a receiving transducer and are amplified and applied to one set of plates of a Cathode Ray Oscilloscope (CRO). The transmitted signal and reflected signal from the flaw and back surface of the metal produce different peaks. The position of the second peak on the time base of the oscilloscope will give the distance of flaw from the front surface of the metal.

Cleaning and Clearing Ultrasonic waves can be used for cleaning machine parts, utensils, washing clothes, and removing dust and soot from the chimney.

How does Ultrasonic Cleaning work? Ultrasonic cleaning is the rapid and complete removal of contaminants from objects by immersing them in a tank of liquid flooded with high-frequency sound waves. These nonaudible sound waves create a scrubbing-brush action within the fluid.

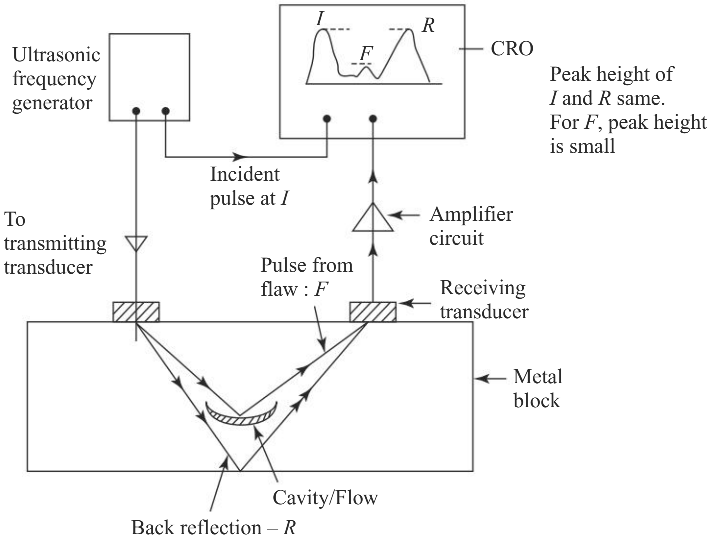


Fig. 2.9 *Flaw detection in metals by ultrasonic technique*

The process is brought about by high-frequency electrical energy that is converted by a transducer into high-frequency sound waves—ultrasonic energy. Its ability to clean even the most tenacious substances from items derives from the core of the unit: the transducer. The cleaning power of a unit stems from the transducer's performance. The efficiency of the transducer will affect both the cleaning time and efficacy achieved during the cleaning cycle. A poor-quality transducer will use more power and take longer to clean items than a good transducer.

The ultrasonic energy enters the liquid within the tank and causes the rapid formation and collapse of minute bubbles: a phenomenon known *as cavitation.* The bubbles travel at high speed around the tank causing them to implode against the surface of the item immersed in the tank in an enormous energy release, which gently lifts contamination off the surface and innermost recesses of intricately shaped parts.

It is this ability to clean box joints, hinges and threads quickly and effectively that has made ultrasonic cleaners the first choice for many industries for over 25 years.

Sonar Sound Wave Navigation and Ranging (SONAR) is used for detection of submarines, iceberg and other objects in ocean. One can determine the presence of submerged submarines, icebergs, rocks or an enemy aircraft in the sky by this technique. In this technique, a sharp ultrasonic beam is directed in various directions into the sea. After its reflection, it is picked up by the ultrasonic detector. The reflected waves from any direction are due to presence of some reflecting medium like submarines, icebergs, rocks, etc., in the sea. The idea of the distance of the body is obtained by the time interval between the generation of ultrasonic waves and their return after reflection. If the submerged body is moving, there will be a change in frequency of the echo signal due to Doppler effect. It helps to determine the velocity of the body and predicts its direction.

Depth of the Sea The ultrasonic waves are highly energetic and show a negligible diffraction effect. So they can be used to find the depth of the sea. The time interval between sending and receiving the wave from the sea is recorded. As the velocity of the waves is known, the depth of the sea can be estimated. Depth of sea' $d = vt/2$.

Medical Applications of Ultrasonic Waves Ultrasonic techniques are complementary to other physical methods used in surgery, therapy and diagnosis. Most medical applications of ultrasound are based on the properties of longitudinal waves in the frequency range of 1–15 MHz. Ultrasonic waves travel at similar velocities (about 1500 m/s) in most biological tissues, and are absorbed at a rate of about 1 dB cm^{-1} MHz^{-1}. Absorption occurs chiefly due to relaxation processes. It leads to thermal effects in biological systems. Mechanical effects, such as streaming and cavitation, are also important in certain situations, particularly at low frequencies. Highly focused ultrasound is used in neurosurgery; it is the only method for producing trackless damage deep in the brain. In vestibular surgery, ultrasonic irradiation is used routinely for the treatment of Ménière's disease; it can alleviate the symptoms without damage to hearing. Therapeutic applications include the treatment in physiotherapy of various soft-tissue ailments, and the production of aerosols for inhalation. The examination of soft tissues is possible both by pulse-echo and continuous-wave techniques. Ultrasonic diagnosis provides information about the position and extent of characteristic impedance discontinuities; this information cannot be obtained directly by any other method. One- and two-dimensional displays, and time-position waveforms, are produced by pulse-echo techniques. Continuous-wave techniques include those based on the Doppler frequency shift of ultrasound reflected by moving structures.

Applications in Processing, Testing and Communications In processing applications, ultrasonic vibration at high power densities is used to produce a physical or chemical change in material. In liquids, the effective agent is often cavitation, which causes emulsification and dispersion, useful for cleaning and atomisation.

Focused ultrasound is used for surgery. Echo-sounding techniques are used for testing materials as well as at sea. They have recently been applied in medical diagnosis. Ultrasonography by image-forming methods may be an alternative method. The high-frequency elastic properties of a material can be measured with great precision and give information about its structure.

The most important communications application is in ultrasonic delay lines, which are useful in radar and data-handling systems.

Signaling Due to smaller wavelength, the ultrasonic waves can be concentrated into a sharp beam. So they can be used for signaling in a particular direction.

Soldering and Metal Cutting Ultrasonic waves can be used for drilling, cutting, soldering, and room-temperature welding of different metals. To solder aluminum, ultrasonic waves along with an electrical soldering iron is used.

Formation of Alloys Different constituents of alloys have different densities. So they can be kept mixed uniformly by a beam of ultrasonic waves. Thus, it is easy to get an alloy of uniform composition.

Ultrasonic Mixing A colloid solution, an emulsion of two nonmiscible liquids (like oil and water), can be mixed by simultaneously subjecting to ultrasonic radiations. Most of the emulsions like polishes, paints, food products and pharmaceutical preparations are prepared by using ultrasonic mixing.

Effect on Coagulation and Crystallization Processes The suspended particles in the liquid can be brought quite close to each other by ultrasonics, so that coagulation may take place. Once their size is increased they can be processed easily. The crystallisation rate is improved by ultrasonic waves. When melt is put to crystallisation, the size of crystals can be made smaller and more uniform by the use of ultrasonics.

Ultrasonics in Metallurgy The grain size can be refined and trapped gases can be removed by irradiating melt with ultrasonic waves during the process of cooling.

Destruction of Lower Life Animals like rats, frogs, fishes, etc., can be killed or injured by high intensity ultrasonics. This way there is no need to use poisonous liquids for this purpose which can be harmful to human beings.

COMPARISON OF MAGNETOSTRICTIVE VERSUS PIEZOELECTRIC TRANSDUCERS FOR POWER ULTRASONIC APPLICATIONS

2.7

Magnetostrictive and piezoelectric are two fundamental transducer designs used for power ultrasonic applications. Magnetostrictive transducers utilise the magnetostrictive property of a material to convert the energy in a magnetic field into mechanical energy. Piezoelectric transducers utilise the piezoelectric property of a material to convert electrical energy directly into mechanical energy

Comparisons Between the Two Techniques There are relative advantages and disadvantages of the two technologies presented above. Some of these are discussed below in some detail.

2.7.1 Frequency Range

Magnetostrictive transducers are frequency limited within a given frequency range. Typical ultrasonic frequencies range from 20 kHz to 200 kHz. Due to physical size limitations of the magnetostrictive transducer (frequency is dependent on the length of the transducer with higher frequency requiring a shorter and shorter length), it is inherently limited to operate at frequencies below approximately 30 kHz.

Piezoelectric transducers do not suffer frequency barriers. The manufacturer can choose an appropriate piezoelectric design and drive it at a selected output frequency over the entire ultrasonic range by utilising harmonic multiples of the primary resonant frequency. This makes piezoelectric transducers a more versatile choice from the standpoint of frequencies available.

2.7.2 Audible Noise

Normally, an adult human can hear sounds in the range of frequencies of 20 Hz to 18 kHz. For magnetostrictive systems, the operating range is from 18 to approximately 30 kHz, i.e., their first subharmonic (which will be half of the operating frequency) is always within the human audible range. The limitation imposed on magnetostrictive systems due to high energy present at the first subharmonic of the ultrasonic frequency may often seem extremely loud to the human ear.

For piezoelectric transducers the typical energy used is equivalent to 40 kHz and has its first subharmonic at 20 kHz which is above the normal adult hearing limit. However, the second harmonic is in the audible range (at 2.kHz), with very low energy compared to first harmonic, and so the sound level is very low and can be ignored. Therefore, it will be a costly affair to bring a magnetostrictive system below the 85 db level according to OSHA standards (a government agency in the Department of Labor to maintain a safe and healthy work environment: for 85 db level and above for, hearing protection is required). By appropriate design and acoustic shielding, higher-frequency piezoelectric equipment can easily be brought under OSHA limits.

2.7.3 Transducer Reliability

Both types of transducers are mature, highly reliable, engineered devices with today's technology. It is interesting to note that the generators used to drive modern piezoelectric ultrasonic generators have a two-year warranty because their uses advanced semiconductors in their fabrication. On the other hand, magnetostrictive generator designs commonly employing older SCR technology are more prone to failure under load due to high amperage and switching frequency. Due to this, typical magnetostrictive generator warranties are limited to one year.

2.7.4 Sweep Frequency

Sweeping frequency is an important feature for uniform cleaning. Main ultrasonic suppliers provide piezoelectric transducers with sweeping frequency due to the market demand for this highly effective feature. For a magnetostrictive transducer, a large inertial mass puts restriction to shift the frequency as rapidly as is required to get good sweep frequency. So these systems have not moved to a sweep frequency system.

2.7.5 Energy Efficiency

Piezoelectric transducers are extremely efficient due to the direct conversion of electrical to mechanical energy in a single step. Direct application of the power to the piezoelectric materials causes it to change shape and create the sound wave. Almost 95% of the power delivered to the transducer is used to do cleaning. Modern ultrasonic generators used to drive piezoelectric transducers have an overall system efficiency of 70% or higher.

On the other hand, magnetostrictive systems rely on the double conversion of electrical energy to magnetic energy to mechanical to produce the sound wave. Due to the energy lost in heating of the coils and the effects of magnetic hysteresis, magnetic systems are usually less than 50% efficient.

2.7.6 Frequency versus Application

Frequency influences strongly on the diameter of the cavity. Low frequencies produce large diameter cavity and high frequencies, a small diameter cavity. This is also is true for energy per cavitations. However, the density of cavitations is high with high-frequency systems and low with low frequency systems. The total energy is equal for both high and low frequencies.

For most cleaning applications, 40 kHz is considered the precision cleaning frequency. Low-frequency systems are better in high mass applications where the soil to be removed is extremely difficult.

Concluding Remark Today's piezoelectric ultrasonic transducers are reliable, efficient devices which can be used with confidence in all power ultrasonic applications. They are capable of providing a wider range of frequency and waveform characteristics and higher electrical conversion efficiency than their magnetostrictive counterparts. So, piezoelectric transducers are a better choice for advanced technology.

Solved Problems

1 *Calculate the length of nickel rod required to produce ultrasonic waves of frequencies of 50 kHz and 100 kHz. Consider the fundamental mode of vibrations of the rod. Given, Density of nickel = 8908 kg/m*^{3}*, Young's modulus of nickel = 2.14 × 10*^{11} *N/m*^{2}.

Solution Given, $n = 1$, YM $= 2.14 \times 10^{11}$ N/m^2, $\rho = 8909$ kg/m^3

(i) For frequency $f = 50$ kHz

$$f = \frac{n}{2L}\sqrt{\frac{Y}{\rho}}$$

$$L = \frac{n}{2f}\sqrt{\frac{Y}{\rho}} = \frac{1}{2 \times 50 \times 10^3}\sqrt{\frac{2.14 \times 10^{11}}{8908}} = 0.048 \text{ m}$$

(ii) For frequency $f = 100$ kHz

$$f = \frac{n}{2L}\sqrt{\frac{Y}{\rho}}$$

$$L = \frac{n}{2f}\sqrt{\frac{Y}{\rho}} = \frac{1}{2 \times 100 \times 10^3}\sqrt{\frac{2.14 \times 10^{11}}{8908}} = 0.049 \text{ m}$$

For a frequency of 50 kHz, the required length of the nickel rod is 0.049 m and for double the frequency, i.e., 100 kHz, the required length of the nickel rod is half, i.e., 0.0245 m.

2 *A nickel rod having 5 cm length is vibrating at resonance. Calculate the fundamental frequency of vibration for which ultrasonic waves are generated. Given, Young's modulus of nickel = 2.14 ×* *10*^{11} *N/m*^{2}*, density of nickel = 8908 kg/m*^{3}.

Solution Given, $L = 5 \times 10^{-2}$ m, $n = 1$, YM $= 2.14 \times 10^{11}$ N/m^2

$$\rho = 8909 \text{ kg/m}^3$$

$$f = \frac{n}{2L}\sqrt{\frac{Y}{\rho}} = \frac{1}{2 \times 5 \times 10^{-2}}\sqrt{\frac{2.14 \times 10^{11}}{8908}} = 49.013 \text{ kHz}$$

3 *Calculate the operating frequency of an oscillator circuit for magnetostriction method if inductor L = 15 mH, capacitors* C_1 *= 0.001 µF and* C_2 *= 0.01 µF. Calculate the length of the nickel rod required to produce the ultrasonic waves at resonance. Calculate the speed of waves produced. Given, Young's modulus of nickel = 2.14 × 10*^{11} *N/m*^{2}*, Density of nickel = 8908 kg/m*^{3}.

Solution Given, $n = 1$, for fundamental mode of vibration

$$C_1 = 0.001\ \mu\text{F},\ C_2 = 0.01\ \mu\text{F},\ Y = 2.14 \times 10^{11} \text{ N/m}^2,\ \rho = 8908 \text{ kg/m}^3,\ L = 15 \text{ mH}$$

$$C_T = \frac{C_1 C_2}{C_1 + C_2} = \frac{0.001 \times 0.001}{0.001 + 0.001}\ \mu F$$

Total capacitance is

$$C_T = 9.09 \times 10^{-4}\ \mu F = 909 \times 10^{-12}\text{F}$$

Operating frequency is

$$f = \frac{1}{2\pi\sqrt{C_T L}} = \frac{1}{2 \times 3.14\sqrt{909 \times 10^{-12} \times 15 \times 10^{-3}}} = 43.123\ \text{kHz}$$

Now length of nickel rod at, frequency of vibration at the fundamental mode of vibration of the nickel rod is

$$L = \frac{n}{2f}\sqrt{\frac{Y}{\rho}} = \frac{1}{2 \times 43.123 \times 10^{3}}\sqrt{\frac{2.14 \times 10^{11}}{8908}} = 0.056\ \text{m}$$

Now,

$$\lambda = 2L \Rightarrow v = \sqrt{\frac{Y}{\rho}} \quad \text{for } n = 1$$

$$\therefore\ v = \sqrt{\frac{2.14 \times 10^{11}}{8908}} = 4901.3\ \text{m/s}$$

4 *A quartz crystal having 3 mm thickness is vibrating at resonance. Calculate the fundamental frequency of vibration for which ultrasonic waves are generated. Given, Young's modulus for quartz =* 7.9×10^{10} *N/m*^{2} *and density of quartz = 2650 kg/m*^{3}*.*

Solution Given, for fundamental frequency, $n = 1$

Thickness of quartz crystal, $t = 0.003$ m

$$Y = 7.9 \times 10^{10}\ \text{N/m}^2$$

$$\rho = 2650\ \text{kg/m}^3$$

$$f = \frac{n}{2t}\sqrt{\frac{Y}{\rho}} = \frac{1}{2 \times 0.003}\sqrt{\frac{7.9 \times 10^{10}}{2650}} = 9.099 \times 10^{5}\ \text{Hz}$$

$$f = 0.909\ \text{MHz}$$

5 *Calculate the operating frequency of the oscillator circuit for the Piezoelectric method if inductors* $L_1 = 1000\ \mu H$, $L_2 = 100\ \mu H$, *capacitor* $C = 20$ *pF, mutual inductance between the coil,* $M = 20\ \mu H$. *Calculate the thickness of the quartz crystal required to produce the ultrasonic waves of the above frequency. Calculate wavelength and speed of ultrasonic waves. Given, Young's modulus for quartz =* 7.9×10^{10} *N/m*^{2} *and density of quartz = 2650 kg/m*^{3}*.*

Solution $L_1 = 1000\ \mu\text{H},\ L_2 = 100\ \mu\text{H},\ M = 20\ \mu\text{H},\ C = 20\ \text{pF}$

$$L_T = L_1 + L_2 + 2M$$

$$\therefore\ L_T = [1000 + 100 + (2 \times 20)]\ \mu\text{H} = 1140\ \mu\text{H} = 1.14\ \text{mH}$$

Operating frequency f

$$f = \frac{1}{2\pi\sqrt{L_T C}} = \frac{1}{2 \times 3.14\sqrt{1.14 \times 10^{-3} \times 20 \times 10^{-12}}}$$

Now for fundamental mode of vibration, $n = 1$

$$f = \frac{n}{2t}\sqrt{\frac{Y}{\rho}}$$

$$t = \frac{1}{2 \times f}\sqrt{\frac{7.9 \times 10^{10}}{2650}}$$

$$t = \frac{1}{2 \times 1.054 \times 10^{6}}\sqrt{\frac{7.9 \times 10^{10}}{2650}} = 2.59 \times 10^{-3}\ \text{m}$$

$$t = 2.6\ \text{mm}$$

Wavelength of ultrasonic wave $\lambda = 2t = 2 \times 2.29 \times 10^{-3}\ \text{m} = 5.18\ \text{mm}$

Speed of ultrasonic waves is,

$$v = \lambda f$$

$$= (5.18 \times 10^{-3}) \times (1.054 \times 10^{6})$$

$$\therefore\ v = 5459\ \text{m/s}$$

Multiple-Choice Questions

1. Which one from the following is a correct characteristic of ultrasonic waves?
   - (a) Ultrasonics are sound waves of very long wavelength.
   - (b) Ultrasonics are sound waves of very high frequency.
   - (c) Ultrasonic waves move faster than sound waves.
   - (d) Ultrasonic waves move slower than sound waves.

2. Ultrasonic waves have frequency
   - (a) greater than 20 kHz
   - (b) less than 20 kHz
   - (c) between the range 20–20000 Hz
   - (d) less than 20 Hz

3. Ultrasonic waves are
   - (a) electromagnetic waves of very large wavelength
   - (b) electromagnetic waves of very high frequency
   - (c) sound waves of very large wavelength
   - (d) sound waves of very high frequency

4. Ultrasonic waves cannot be produced by
   - (a) radio frequency oscillator with diaphragm loudspeaker
   - (b) radio frequency oscillator with quartz crystal
   - (c) radio frequency oscillator with nickel rod
   - (d) all of the above methods

5. The magnetostriction phenomenon is useful for producing
   - (a) ultrasonic waves
   - (b) electromagnetic waves
   - (c) light
   - (d) heat

6. The piezoelectric phenomenon is observed in a
   - (a) quartz crystal
   - (b) NaCl crystal
   - (c) nickel rod
   - (d) iron rod

7. Which oscillator from the following list is not useful for production of ultrasonic waves?
   - (a) Colpitt's oscillator
   - (b) Hartley oscillator
   - (c) Radio frequency oscillator
   - (d) Audio frequency oscillator

ANSWERS TO MCQS

1. (b) 2. (a) 3. (d) 4. (a) 5. (a) 6. (a) 7. (d)

Practice Problems

1. A nickel rod is used to produce ultrasonic waves of frequency *f*. If we want to produce an ultrasonic frequency of 2*f*, what should be the length of the rod? **[Length *L* of the rod is halved]**

2. A quartz crystal having thickness ***t*** is vibrating at resonance. The fundamental frequency of vibration is *f*. If the thickness is made twice its original value, what will be its new frequency? If thickness is reduced to half what will be its frequency? **[New frequency (*f*/2); 2*f*]**

3. We want to generate an ultrasonic wave of frequency *f* by both the popular methods, i.e., magnetostriction and piezoelectric methods. What should be the length of nickel rod for the magnetostriction method and thickness of quartz crystal for the piezoelectric method?

Given, Young's modulus of nickel = 2.14×10^{11} N/m^2,

Density of nickel = 8908 kg/m^3.

Young's modulus for quartz = 7.9×10^{10} N/m^2

Density of quartz = 2650 kg/m

[Length of nickel rod = (1.113) thickness of quartz]

Short-answer Questions

1. What is ultrasonics?

2. Which principles are used to produce ultrasonic methods?

3. What is magnetostriction?

4. What is piezoelectric effect?

5. Which method is suitable for producing high-frequency waves of the order of 1 MHz?

6. Why cannot a loudspeaker be used for production of ultrasonic waves?

7. Explain the term 'SONAR'.

8. Explain acoustical grating.

9. What are the major differences between audible waves and ultrasonic waves?

10. What are the properties of ultrasonic waves?

11. Explain briefly the medical applications of ultrasonic waves.

Long-answer Questions

1. Define ultrasonics. Describe piezoelectric method for their production. How are they used for SONAR (depth sounding) exploration?

2. Explain the phenomenon of magnetostriction. How will you produce high-frequency sound waves with its help?

3. What are ultrasonic waves? Discuss their properties. How are they produced and detected?

4. Explain in detail how the ultrasonic pulse technique is used for nondestructive testing of materials.

5. Give an account of the methods used in the production of ultrasonic waves.

6. Write an essay on ultrasonics and their applications.

7. What are ultrasonic waves? Describe some of their applications.

8. Enumerate the different methods for the production of ultrasonic waves and describe one of them in detail.

9. How will you determine the wavelength of these waves?

10. Discuss the method of measurement of frequency of ultrasonic waves.

11. Discuss magnetostrictive versus piezoelectric transducers for power ultrasonic application.

Questions from Previous Exams

(i) Define piezoelectric effect. **(GTU- Jan' 2009)**

When pressure is applied to one pair of opposite faces of crystals like quartz, tourmaline, Rochelle salt, etc., cut with their faces perpendicular to its optic axis equal and opposite charges appear across its other faces.

(ii) Explain the terms magnetostriction and piezoelectric effect. Discuss any one method of production of ultrasonic waves. **(GTU- Jan' 2009)**

Refer Section 2.2.

(iii) Give the properties of an ultrasonic wave. **(GTU- Jan' 2010)**

Frequency is greater than 20 kHz

It can travel long distance.

Speed of propagation depends upon frequency.

(iv) Describe the principle and the method of production of ultrasonic waves by magnetostriction method. **(GTU- Jan' 2010)**

Refer Section 2.2.

(v) Calculate the frequency to which a piezoelectric oscillator circuit should be tuned so that a piezoelectric crystal of thickness 0.1cm vibrates in its fundamental mode to generate ultrasonic waves.(Young modulus and the density of the material of crystal are 80 GPa and 2654 kgm^{-3}) **(GTU- Jan' 2009)**

Given $E = 80\text{ GPa} = 80 \times 10^9\text{ Pa}$, $\rho = 2654\text{ Kgm}^{-3}$, $t = 0.1\text{ cm} = 0.1 \times 10^{-2}\text{m}$

The frequency of vibration is given by

$$f = [1/2t](E/\rho)^{1/2}$$

$$[1/(2 \times 0.1 \times 10^{-2})](80 \times 10^9/2654)^{1/2}$$

$$f = 0.915\text{MHz}$$

(vi) Calculate the thickness of a quartz plate needed to produce ultrasonic wave of frequencies (i) 2 MHz (ii) 30 kHz. (Given ρ = 2650 kg/m^3 and Young's modulus = 8 × 10^{10} N/m^2) **(GTU- Jan' 2010)**

Given $f = 2\text{ MHz} = 2\times10^6\text{ Hz}$, $\rho = 2654\text{Kgm}^{-3}$, Young's Modulus = $8 \times 10^{10}\text{ N/m}^2$

$$t_1 = ?$$

The frequency of vibration is given by

$$f = 1/2t\ (E/\rho)^{1/2}$$

Therefore,

$$t_1 = 1/2f\,(E/\rho)^{1/2}$$

$$t_1 = 1.37 \times 10^{-3}\text{m}$$

Given $f = 30$ kHz $= 30 \times 10^3$Hz, $\rho = 2654$ kgm^{-3}, Young's Modulus $= 8 \times 10^{10}$ N/m^2

$$t_2 = ?$$

The frequency of vibration is given by

$$f = 1/2t\,(E/\rho)^{1/2}$$

Therefore,

$$t_2 = 1/2f\,(E/\rho)^{1/2}$$

$$t_2 = 0.09157 \text{ m}$$

(vii) What is the absorption coefficient? Explain how to determine the absorption coefficient. (GTU- Jan' 2010)

(viii) Frequency range of audible sound wave is 20 Hz < f < 20 kHz. (GTU- June' 2010)

(ix) What is piezoelectric method? (GTU- June' 2010)

When pressure is applied to one pair of opposite faces of crystals like quartz, tourmaline, Rochelle salt, etc., cut with their faces perpendicular to its optic axis equal and opposite charges appear across its other faces.

(x) Calculate the frequency at which 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. (GTU- June' 2010)

Given, Young's Modulus $Y = 80$ GPa, and density of crystal material $= 2654$ kgm^{-3},

The frequency of vibration is given by

$$f = 1/2t\,(Y/\rho)^{1/2}$$

$$f = 2.74 \times 10^6 \text{ Hz}$$

(xi) What is acoustic grating? Explain the acoustic grating method to determine the velocity of ultrasonic waves in liquid. (GTU- June' 2010)

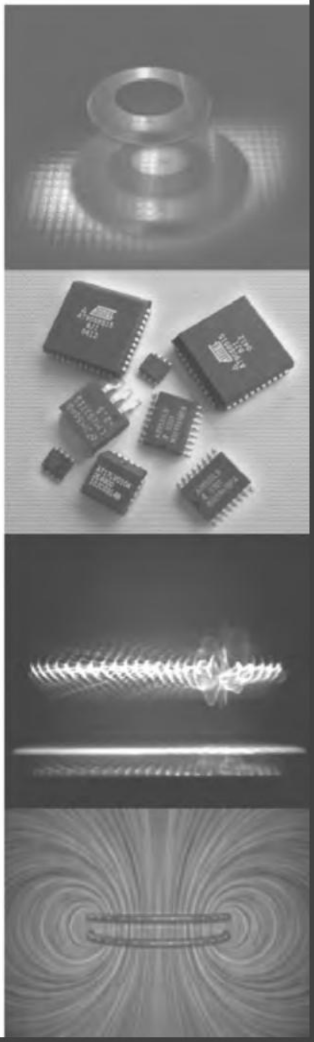
Refer Section 2.5

The McGraw-Hill Companies

Crystal Physics

3

A diamond made up of carbon differs much in terms of physical properties from graphite which is also made of carbon. What causes this difference? If we look at a crystal of quartz or calcite, what causes them to exhibit physical properties like piezoelectricity and double refraction? How can the different faces and the crystal orientations be identified? The fundamental theories about all these questions are discussed in this chapter.



INTRODUCTION

3.1

Many solids used in engineering are crystalline in structure, that is, their atoms, ions or molecules are arranged in a regular, repetitive manner, forming a three-dimensional pattern which may be obtained by the three-dimensional repetition of a certain pattern unit. This repeating three-dimensional pattern of atoms sometimes controls the external shape of the crystals as in the case of a small cake of table salt. However, if the corners of a sodium-chloride crystal are rounded off, the internal pattern remains intact as it is not altered by changing the external shape. A study of the geometric form and other physical properties of crystalline solids with the help of X-rays, electron beams, and neutron beams, etc., is called the science of *crystallography.* The knowledge of this science is essential for understanding the crystalline solid state. On this depends the understanding of certain important engineering properties. To an engineer, a solid is an aggregate of matter that retains its shape in contrast to a liquid or gas that has no shape except that of the vessel containing it. It may further be recognised that there is a time factor to be considered, i.e., a solid may retain its shape but not possibly for long under certain conditions. For instance, plastics, paints and pitch are but a few materials that are solid in some respects and liquid in another but they are outside the boundary of crystallography. Some aspects of crystalline structure are studied with the help of X-rays, neutron beams, and electron beams. However before, we apply such methods for the determination of crystal structure, it is important to study briefly the various systems of crystals, particularly those which are necessary for our study.

CLASSIFICATION OF SOLIDS

3.2

It is observed that most of the solids are found in either of the two categories, viz., crystalline or amorphous.

Crystalline These solids are characterised by a perfect or nearly perfect periodicity of atomic structure. This geometric regularity of structure provides a simple conceptual picture of a crystal and helps a lot in gaining knowledge of the physical properties of the solid.

Crystalline solids or, in general, crystals are the solids composed of atoms, ions or molecules, which are arranged in a regular, periodic way throughout in a three-dimensional pattern. Each atom, ion or molecule is fixed at a definite point in space at a definite distance from and in a definite angular orientation to all others surrounding it. This implies that the internal spatial symmetry of atomic, ionic or molecular orientation is an essential feature of the crystalline state. Therefore, crystals develop well-defined geometric shapes or faces and angles between them.

Further, it is to be noted that the crystals are either in single crystalline or polycrystalline form. If the regular, periodic arrangement of atoms is extended throughout the body of the solid without any interruption, the crystal is known as a *single crystal*, and if an aggregate of many such small crystals or crystallites or grains separated by well-defined boundaries known as *grain boundaries* are formed, the crystal is known as *polycrystalline.* Diamond is an example of a single crystal, while copper wire is an example of a polycrystalline solid.

Crystals are anisotropic substances, meaning thereby they possess different physical properties along different directions and planes. In cubic crystals, the electrical conductivity, for example, should be different along edges of the cube and that along the body diagonal and face diagonal.

Amorphous In such solids, though the atoms or molecules are strongly bonded yet there is hardly any geometrical regularity or periodicity in the way in which the atoms are arranged in space.

Therefore, in amorphous solids, the constituent atoms or molecules are randomly arranged so that no regular, periodic arrangement of atoms or molecules exists.

Due to the randomised arrangement of molecules, they are considered to be *isotropic*, i.e., the physical properties measured along all directions are identical. Plastic, for example, has the same thermal or electrical conductivity along any direction.

Distinction Between Crystalline and Amorphous Solids

1. Crystalline solids have a regular arrangement of particles (atoms, ions or molecules) whereas amorphous solids have a completely random particle arrangement.
2. Crystalline solids have different physical properties (e.g., thermal conductivity, electrical conductivity, refractive index, etc.) in different directions, i.e., they are anisotropic; while amorphous substances are isotropic, i.e., they have the same physical properties in all directions.
3. The cooling curve for amorphous substance is smooth, while the curve of crystalline substance has breaks, the middle of which corresponds to the process of crystallisation. In the process of crystallisation, some energy is liberated which compensates the loss of heat and hence the temperature remains constant.

4. In case of a crystalline substance, the melting point is very sharp; while amorphous substances do not have sharp melting points. For example, as the temperature of glass is gradually raised, it softens and starts flowing without any sharp change from solid state to liquid state.

The distinction between crystalline and amorphous solids is, however, not absolute.

CRYSTAL STRUCTURE

3.3

Crystals appear when a substance changes from one state to another state, i.e., in phase transformation. The basic conditions for the formation of crystalline solid are as follows:

(i) A change from the liquid phase to solid crystallisation by the process of solidification.

(ii) If a change from one solid phase to another solid phase takes place, there is an alteration of the shape of the crystal structure and re-crystallisation occurs.

(ii) A change from the gaseous phase to the solid crystallisation by the process of sublimation.

As a crystal is a three-dimensional body made up of regular and periodic three-dimensional patterns of atoms or molecules in space called the crystal structure, it may be described in terms of an idealised geometrical concept called a space lattice.

Lattice Points and Space Lattice Points can be imagined in space about which atoms or molecules are located. Such points are called *lattice points.* The totality of such points is called *space lattice* or *crystal lattice.* The three-dimensional space lattice (*3-D space lattice*) may be defined as a finite array of lattice points in three dimensions in which each and every lattice point has identical surroundings in the array.

3.3.1 Basis and Crystal Structure (Fig. 3.1)

Every lattice point can be associated with one or unit assembly of atoms or molecules identical in composition called the *basis* or *pattern.* The regular periodic three-dimensional arrangement of a basis is called *crystal structure.* Space lattice is imaginary. Crystal structure is real.

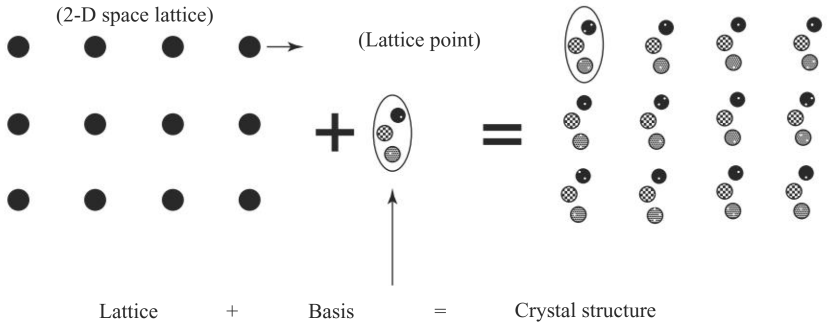


Fig. 3.1 *Showing crystal structure formed by adding lattice and basis*

3.3.2 Bravais and Non-Bravais Lattice

A *Bravais* lattice is one in which all lattice points are identical in composition. If the lattice points are not identical then the lattice is called *non-Bravais* lattice.

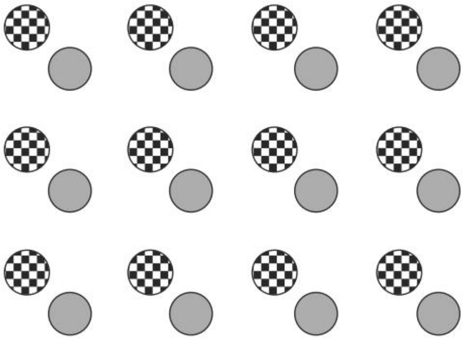
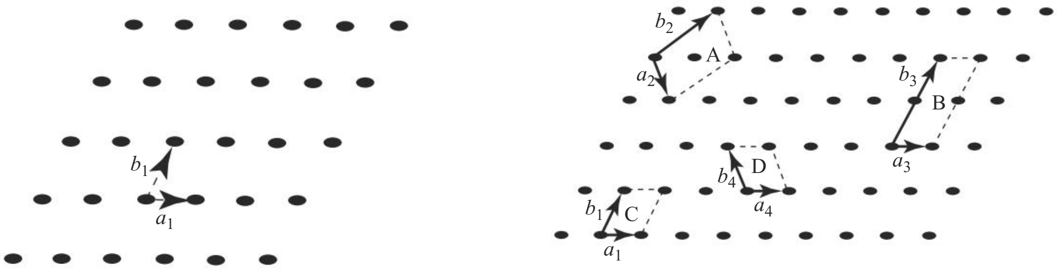


Fig. 3.2 *Bravais and non-Bravais lattice*

The set of lattice points ▦'s together constitutes a Bravais lattice. Similarly, the set of lattice points ●'s together constitutes a Bravais lattice. But the set of all lattice points ▦'s and ●'s together constitutes a non–Bravais lattice. Hence, a non-Bravais lattice could be considered as the superposed pattern of two or more interpenetrating Bravais lattices.

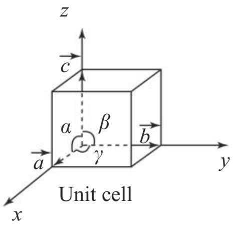
3.3.3 Unit Cell and Lattice Parameters

In every crystal, some fundamental group of particles is repeated. Such a fundamental group of particles is called a ***unit cell***. A unit cell is so chosen that it will represent the symmetry of the crystal. Hence, the unit cell with maximum symmetry is chosen. Unit cells are the basic building blocks of the crystal. If these unit cells are translated in three dimensions, they will generate the crystal. A crystal lattice is described by a unit cell. Each unit cell is described by three lattice parameters a, b and c which are the length of the vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$ and the three angles α, β, γ between these vectors as shown in Fig. 3.3(c). They are nothing but the intercepts of the faces and the interfacial angles. All together they constitute the *lattice parameters.* Two-dimensional lattice representation, in general, is shown in Fig. 3.3(a) and its various possible unit-cell representations (A, B, C and D) are shown in Fig. 3.3(b).



(a) Two-dimensional representation of a single crystal lattice

(b) Two-dimensional representation of a single crystal lattice showing various possible unit cells



(c) Unit cell representation by axial lengths and angles

Fig. 3.3 *Unit-cell representation in two and three dimensions*

3.3.4 Primitive Cell (Fig. 3.4)

Sometimes reference is made to a primitive cell. A primitive cell may be defined as a geometrical shape which, when repeated indefinitely in three dimensions, will fill all space and it consists of lattice points only at corners. It consists of only one atom per unit cell. Therefore, unit cells may be primitive (simple) or non-primitive.

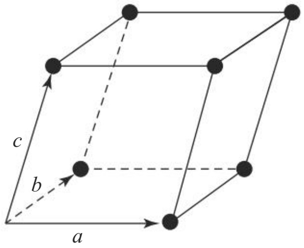


Fig. 3.4 *A generalised primitive unit cell*

THE CRYSTAL SYSTEMS AND BRAVAIS LATTICES

3.4

Based on lattice parameters, crystals are classified into seven basic systems. If atoms are placed only at corners, seven crystal systems yield seven lattices. But few more lattices could be constructed by placing atoms at the face centre, body centre, etc. Bravais showed that there are 14 such lattices existing in nature. Hence, the name Bravais space lattices. Each crystal system differs from the other in lattice parameters.

(1) ***Cubic Crystal System (Isometric)*** $a = b = c$ and $\alpha = \beta = \gamma = 90°$
(Simple or primitive, face centered (fcc) and body centered (bcc))
Example NaCl

(2) ***Tetragonal Crystal System*** $a = b \neq c$ and $\alpha = \beta = \gamma = 90°$
(Simple and body centered)
Example SO_2

(3) ***Orthorhombic Crystal*** $a \neq b \neq c$ System and $\alpha = \beta = \gamma = 90°$
(Simple, face centered (fcc), body centered (bcc) and base centered)
Example KNO_3

(4) ***Monoclinic Crystal System*** $a \neq b \neq c$ and $\alpha = \beta = 90° = \gamma$
(Simple and Base centered)
Example Na_2SO_3

(5) ***Triclinic Crystal System*** $a \neq b \neq c$ and $\alpha \neq \beta \neq \gamma \neq 90°$
(Simple)
Example $CaSO_4$

(6) ***Trigonal Crystal System (Rhombohedral)*** $a = b = c$ and $\alpha = \beta = \gamma \neq 90°$
(Simple)
Example $CaCO_3$

(7) ***Hexagonal Crystal System*** $a = b \neq c$ and $\alpha = \beta = 90°$, $\gamma = 120°$
(Simple)
Example ZnO

The fourteen Bravais space lattices are as shown in Fig. 3.5.

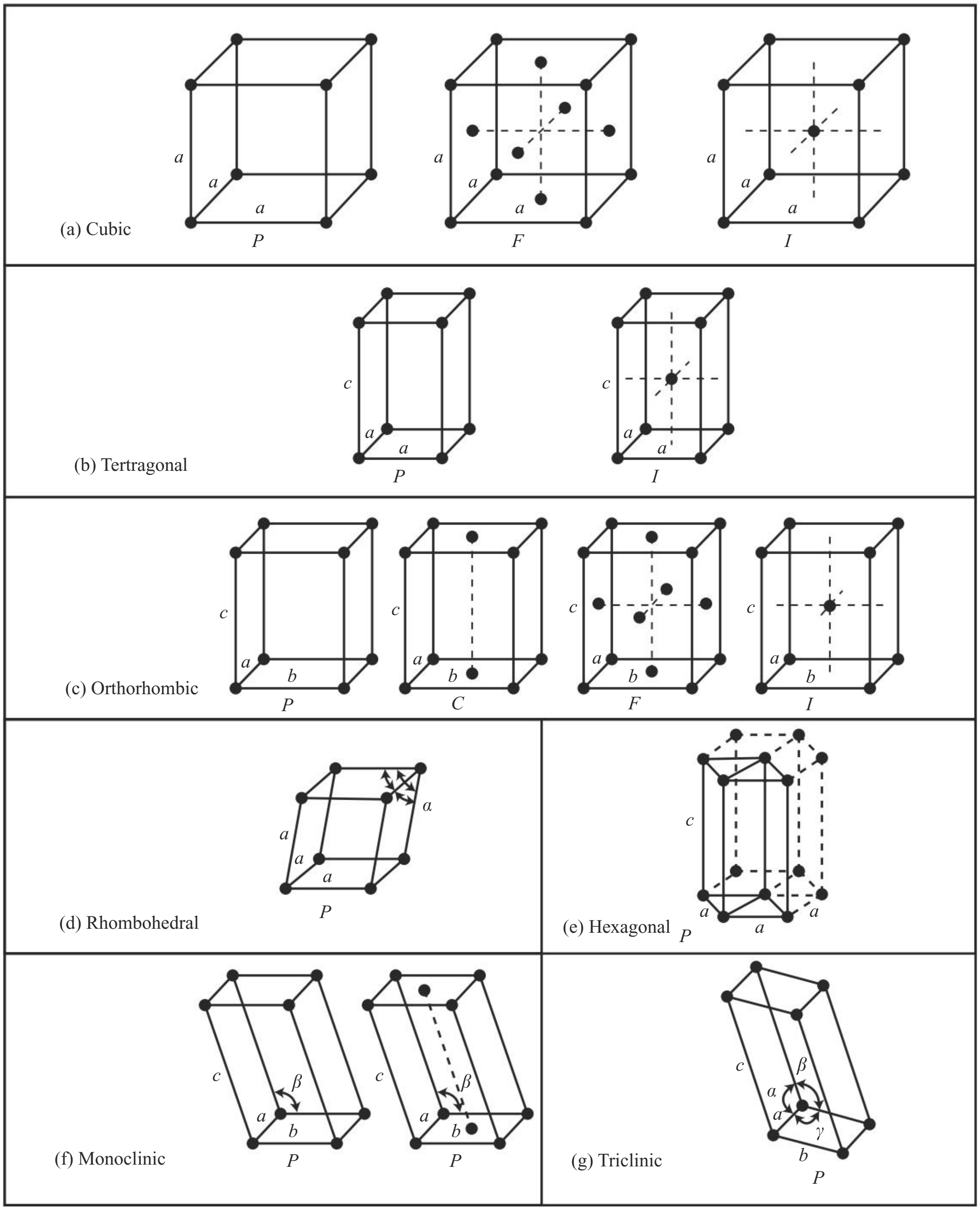


Fig. 3.5 *The fourteen Bravais space lattices*

SPACE LATTICES OF CUBIC SYSTEMS

3.5

The simplicity of the cubic systems is associated with the fact that for them the lattice parameters *a*, *b* and *c* are equal in magnitude and perpendicular to each other. Consider Fig. 3.6 which represents three different types of space lattices of cubic system and leading to different concentration of atoms in special planes. In Fig. 3.6(a), the lattice points are at the corners only. In Fig. 3.6(b), the lattice points are at the corners as well as the centre of the unit cell, and in Fig. 3.6(c), the points are at the corners as well as at the centre of each face.

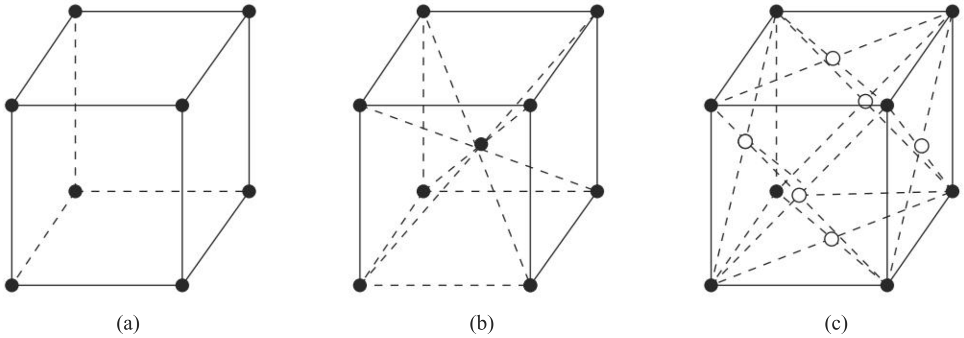


Fig. 3.6 *Space lattices of cubic system; (a) Simple cubic lattice; (b) Body centered cubic lattice; (c) Face centered cubic lattice*

They have respectively been called by Bravais as *simple cubic (sc)* lattice, *body centred cubic (bcc)* lattice and *face centred cubic (fcc)* lattice. Thus, the crystals belonging to the cubic systems have three types of lattices depending upon the situations of the lattice points in the unit cells.

The above-mentioned unit cells (*sc*), (*fcc*), and (*bcc*) are sometimes represented by the symbols *P*, *F* and *I* respectively. It is the symmetry consideration which decides which of these can be had in each system, e.g., in the cubic system, we have *P*, *F* and *I* lattices, while triclinic, hexagonal rhombohedral systems have the *P* lattice only.

MILLER INDICES FOR DIRECTIONS AND PLANES

3.6

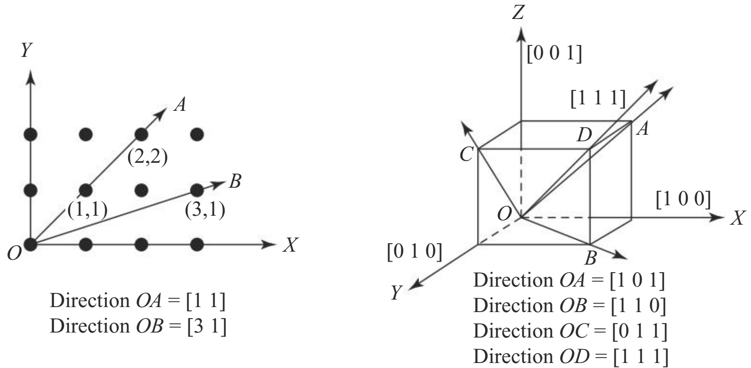
3.6.1 Miller Indices for Directions in Crystals

In crystals there exist various directions like crystallographic axes, face diagonal, body diagonal, etc., which are very important and various crystallographic planes which contain concentration of atoms. Since many planes as well as directions are identical, it is very much essential to label these directions and planes for crystal analysis.

Arrows in two dimensions show directions. ***The directions are described by giving the coordinates of the first whole numbered point [(x, y) in two dimensions, and (x, y, z) in three dimensions] through which each of the direction passes***. Directions are enclosed within square brackets. For this, we use three indices *u, v, w* enclosed in braces without using comma as [*u v w*].

We should note here that the Miller indices for directions are integers and have no common factor larger than unity. The indices of direction are simply the coordinates to the point such that a vector from the origin to the point is the desired direction. The direction specified by, say, the symbol [*u v w*] may be obtained as follows:

Move from the origin over a distance *ua* along the *X*-axis; *vb* along the *Y*-axis; and *wc* along the *Z*-axis. The vector joining the point so obtained with the origin gives the direction indicated by this symbol [*u v w*]. Thus, in a cubic crystal the direction of the *X*-axis is given by [100], the *Y*-axis by [010] and *Z*-axis by [001] as shown in Fig. 3.7. Negative signs may also be used to describe particular directions in a lattice. For example, directions *AO, BO, CO* and *DO* are represented by $[\bar{1}\ 0\ \bar{1}]$, $[\bar{1}\ \bar{1}\ 0]$, $[0\ \bar{1}\ \bar{1}]$, $[\bar{1}\ \bar{1}\ \bar{1}]$ respectively. This can be obtained by shifting the origin of our coordinate system to points *A, B,* or *C* or *D* as the case may be and find the coordinate of the second point *O* in each case. Put the coordinates so obtained in braces. Here, the origin can be shifted to any point so that the unit cell direction can be properly indexed. Figure 3.7 shows the direction indices of some lines in a cubic crystal.



(a) Direction in 2 dimensions

(b) Directions in 3 dimensions

Fig. 3.7 *Miller indices for directions in 2 dim,ensions and 3 dimensions*

A full set of equivalent directions in a crystal is represented by the symbol of the kind [*u v w*]. For instance, the direction of a form, say, all the body diagonals of a cube are indicated by ⟨1 1 1⟩, that is, this is shortened for the eight equivalent diagonal directions.

3.6.2 Miller Indices for Planes in Crystals

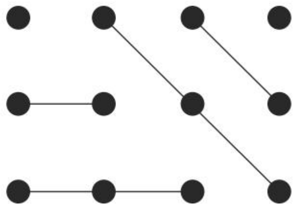


Fig. 3.8 *Different lattice planes*

The crystal may be regarded as made up of an aggregate of a set of parallel equidistant planes, passing through the lattice points, which are known as ***lattice planes***. These lattice planes can be chosen in different ways in a crystal. The problem in designating these planes was solved by ***Miller*** who evolved a method to designate a set of parallel planes in a crystal by three numbers **(*h k l*)** called ***Miller indices of planes.***

Steps Involved in Determining Miller Indices of a Given Set of Parallel Planes Miller indices may be defined as the reciprocals of the intercepts made by the plane on the crystallographic axes when reduced to smallest numbers.

Consider a plane *ABC* as shown in Fig. 3.9 in which one of the planes belongs to the set of parallel planes with Miller indices (*h k l*). Let *x, y* and *z* be the intercepts made by the plane along the three crystallographic axes *X, Y* and *Z* respectively. The following four steps are involved to obtain Miller indices (*h k l*).

1. Determine the coordinates of the intercepts made by the plane along the three crystallographic axes.
2. Express the intercepts as multiples of the unit cell dimensions, or lattice parameters along the axes.
3. Determine the reciprocals of these numbers.
4. Reduce them into the smallest set of integral numbers and enclose them in simple brackets or parentheses (No commas to be placed between indices).

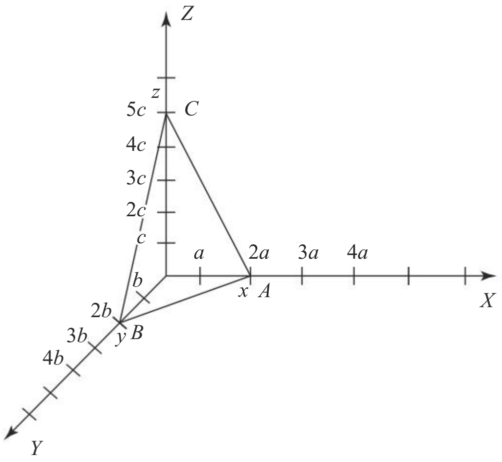


Fig. 3.9 *Plane having Miller indices (h k l)*

According to Fig. 3.9, the steps involved are the following:

1. The intercepts $x = 2a$, $y = 2b$ and $z = 5c$. In general, it can be $x = pa$, $y = qb$, $z = rc$.
2. The multiples of lattice parameters are $\frac{x}{a} = \frac{2a}{a} = 2$, $\frac{y}{b} = 2$, and $\frac{z}{c} = 5$
3. Taking the reciprocals $\frac{a}{x} = \frac{1}{2}$, $\frac{b}{y} = \frac{1}{2}$, and $\frac{c}{z} = \frac{1}{5}$
4. Reducing the reciprocals to smallest set of integral numbers by taking LCM.

$$10 \times \frac{1}{2},\ 10 \times \frac{1}{2},\ 10 \times \frac{1}{5}$$

$$= \quad 5, \quad 5 \quad 2$$

Miller indices of the *ABC* = (*h k l*) = (5 5 2)

Note

(a) All parallel equidistant planes have the same Miller indices.

(b) If the Miller indices have the same ratio, then the planes are parallel.

(c) If the plane is parallel to any of the axes then the corresponding Intercept is taken to be ∞.

(d) Negative indices are represented by putting a bar over the digit.

(e) All equivalent planes are represented by the bracket {h k l}. For example, the set of planes (100), (010), (001), $(\bar{1}00)$, $(0\bar{1}0)$, $(00\bar{1})$ which repesents all six principle planes of a cube are represented by a single index {100}.

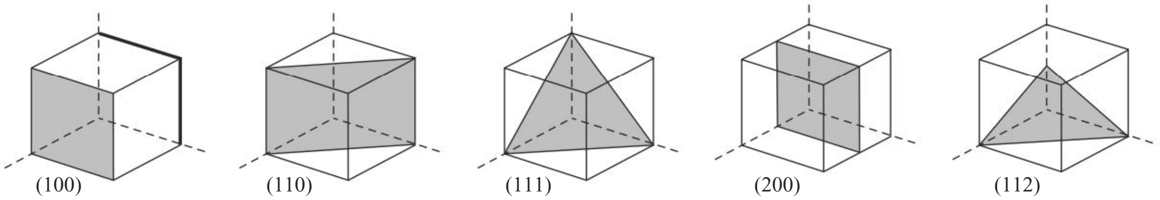

(100) (110) (111) (200) (112)

Fig. 3.10 *Miller indices of some basic planes for cubic structure*

Important Features of Miller Indices of Crystal Plane

The important features of Miller indices for the cubic crystal are the following:

1. If a plane is parallel to any one of the coordinate axes, then its intercept is at infinity. The index is taken to be zero for the intercept of the plane at infinity.
2. All equally spaced parallel planes have the same Miller indices (*h k l*) or vice versa.
3. The indices (*h k l*) do not define a particular plane, but represent a set of parallel planes.
4. It is only the ratio of the indices which is important in this notation.

RELATION BETWEEN INTERPLANAR DISTANCE AND CUBE EDGE

3.7

Consider a lattice plane *ABC,* which is one of the planes belonging to the set of planes with Miller indices (*h k l*). Let *x, y* and *z* be the intercepts made by the plane along the three crystallographic axes *X*, *Y* and *Z* respectively.

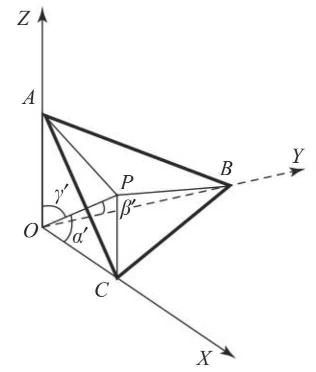


Fig. 3.11 *Representation of the interplanar distance* d_{hkl} *and the cube edge a*

Let *OP* be the perpendicular drawn form the origin to the plane. Let α', β' and γ' be the angles made by *OP* with the crystallographic axes *X*, *Y* and *Z* respectively. Let another consecutive plane parallel to *ABC* pass through the origin *O*. Let *a*, *b* and *c* be the lattice parameters. *OP* is called the interplanar spacing and is denoted by d_{hkl}.

From Fig. 3.11, the direction cosines can be found.

From right angle ΔOPC, $\cos\alpha' = \frac{OP}{OC} = \frac{d_{hkl}}{x}$

From right angle ΔOPB, $\cos\beta' = \frac{OP}{OB} = \frac{d_{hkl}}{y}$

From right angle ΔOPA, $\cos\gamma' = \frac{OP}{OA} = \frac{d_{hkl}}{z}$

We know that, $h = \frac{a}{x}, k = \frac{b}{y}, l = \frac{c}{z}\ \mathrm{v} \Rightarrow x = \frac{a}{h}, y = \frac{b}{k}, z = \frac{c}{l}$

Therefore, $\cos\alpha', = \frac{h}{a} d_{hkl}, \cos\beta' \frac{k}{b} d_{hkl} \cos = \frac{l}{c} d_{hkl}.$

Now for the rectangular Cartesian coordinate system, we have

$$\cos^2\alpha' + \cos^2\beta' + \cos^2\gamma' = 1$$

$$\left(\frac{h}{a}\right)^2 d_{hkl}^2 + \left(\frac{k}{b}\right)^2 d_{hkl}^2 + \left(\frac{l}{c}\right)^2 d_{hkl}^2 = 1$$

$$\Rightarrow \qquad d_{hkl}^2 = \frac{1}{\left(\frac{h}{a}\right)^2 + \left(\frac{k}{b}\right)^2 + \left(\frac{l}{c}\right)^2}$$

$$d_{hkl} = \sqrt{\frac{1}{\left(\frac{h}{a}\right)^2 + \left(\frac{k}{b}\right)^2 + \left(\frac{l}{c}\right)^2}} \tag{3.1}$$

which is the expression for the interplanar spacing. For a cubic lattice $a = b = c$,

$$d_{hkl} = \frac{a}{\sqrt{h^2 + k^2 + l^2}} \tag{3.2}$$

SOME IMPORTANT CRYSTAL STRUCTURES

3.8

Some of the important crystal structures are diamond structure, zinc blende structure, barium titanate structure, sodium chloride structure, etc. Some of them are discussed here.

3.8.1 Diamond Structure

This structure shows one unit cell of the diamond structure. Diamond is composed of pure carbon. In essence, the diamond structure can be viewed as a face-centred cubic array with half of the tetrahedral holes filled. It is important to note, however, that diamond is an example of a network covalent compound. In the diamond structure, the atoms are connected by covalent bonds, with each carbon atom bonded to four others in a tetrahedral geometry. In essence, a sample of diamond is one large molecule.

Fig. 3.12 *Diamond structure*

3.8.2 Zinc Blende Structure

Zinc sulphide (ZnS) exists as two forms, zinc blende and wurtzite. Zinc blende is cubic zinc structure in which the sulphide ions form a face-centered cubic array and the zinc ions fill one-half of the tetrahedral holes. This illustration shows a side view of the zinc blende structure. The big spheres represent the sulphide ions and the small spheres represent the zinc ions. The sulphide ions are quite large (atomic radius = 184 pm) relative to the size of the zinc ions (74 pm).

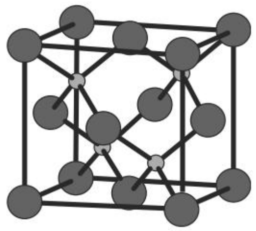


Fig. 3.13 *Zinc blende structure*

3.8.3 Barium Titanate Structure

It is well known that barium titanate belongs to the group of compounds having structures of the perovskite type1. The ideal perovskite structure has a simple cubic lattice, with one formula-weight per cell. It was early recognised that for many of these compounds, including perovskite ($CaTiO_3$) itself as well as barium titanate, the structure was not truly cubic, but was actually a slightly deformed modification of it. Perovskite itself is generally believed to be monoclinic.

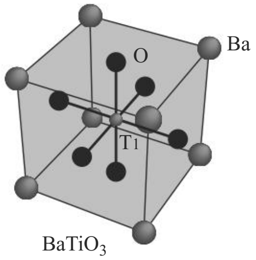


Fig. 3.14 *Barium titanate structure*

3.8.4 Sodium Chloride Structure

Sodium chloride forms crystals with face-centered cubic symmetry. In these, the larger chloride ions are arranged in a cubic close-packing, while the smaller sodium ions fill all the cubic gaps between them.

Fig. 3.15 *Sodium chloride structure*

Each ion is surrounded by six ions of the other kind; the surrounding ions are located at the vertices of a regular octahedron. This same basic structure is found in many other minerals and is commonly known as the halite or rock-salt crystal structure. It can be represented as a face-centered cubic (fcc) lattice with a two-atom basis. The first atom is located at each lattice point, and the second atom is located halfway between lattice points along the fcc unit cell edge.

It is held together by an ionic bond which is produced by electrostatic forces arising from the difference in charge between the ions.

3.9 ATOMIC PACKING FACTOR OF THE CUBIC SYSTEM

Some important definitions are as follows:

Coordination Number (CN) It is the number of equidistant nearest neighbours that an atom has in a crystal structure.

Nearest Neighbour Distance It is the distance between two nearest neighbours in a crystal structure.

Atomic Packing Factor (APF) or Packing Fraction It is the fraction of space occupied by atoms in a unit cell. It is defined as the ratio of volume occupied by atoms in a unit cell to the volume of the unit cell. If the number of atoms per unit cell are 'n', V_a is the volume of atoms in the unit cell and V is the volume of the unit cell then,

Atomic packing fraction, $$APF = \frac{nV_a}{V} \tag{3.3}$$

Lattice Constant In a cubic lattice the distance between atoms remains constant along crystallographic axes and is called lattice constant.

3.9.1 Simple Cubic Structure

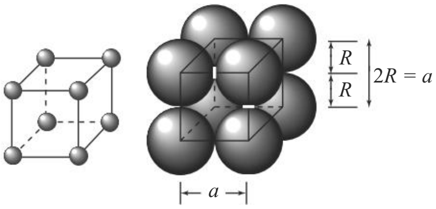


Fig. 3.16 *Simple cubic structure*

In a simple cubic structure, each atom consists of 6 equidistant nearest neighbours. Hence its coordination number is 6.

Eight unit cells share each atom at the corner. Hence only 1/8th of the volume of the atom lies in each cell. Since the atoms are present only at corners, the number of atoms per unit cell is given by

$$n = \frac{1}{8} \times 8 = 1 \text{ atom}$$

In this structure the atoms touch each other along the sides of the cube. Therefore, $a = 2R$, where R is the radius of each atom.

We know that the atomic packing fraction is given by

Atomic packing fraction, $$APF = \frac{nV_a}{V}$$

$$APF = \frac{1 \times \frac{4\pi R^3}{3}}{a^3}$$

$$= \frac{4\pi R^3}{3a^3}$$

$$= \frac{4\pi R^3}{3(2R)^3}$$

$$= \frac{4\pi}{24}$$

$$= 0.5235$$

3.9.2 Face Centred Cubic (FCC) Structure

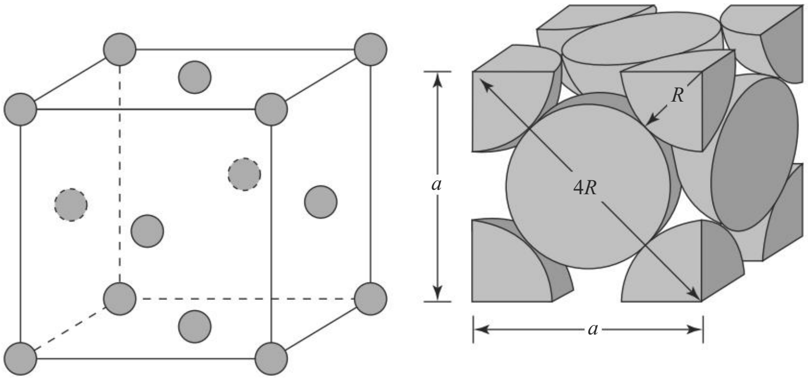


Fig. 3.17 *Face centered cubic structure*

In fcc structure in addition to atoms at corners, atoms are present at face centres. Each atom consists of 12 equidistant nearest neighbours. Hence, the coordination number is 12.

The number of atoms per unit cell is

$$n = \frac{1}{2} \times 6 + \frac{1}{8} \times 8 = 4$$

In this structure atoms touch each other along the face diagonal. Therefore,

$$(4R)^2 = a^2 + a^2;$$

where R is the atomic radius and a is the lattice constant

$$16\,R^2 = 2a^2$$

$$a = 2\sqrt{2}R$$

Atomic packing fraction, $APF = \frac{nV_a}{V}$

$$APF = \frac{1 \times \frac{4\pi R^3}{3}}{a^3}$$

$$= \frac{4\pi R^3}{3(2\sqrt{2}R)^3}$$

$$= \frac{\sqrt{2}\pi}{6}$$

$$= 0.7405$$

Hence atoms occupy 74.05% of the volume of the unit cell.

3.9.3 Body Centred Cubic (BCC) Structure

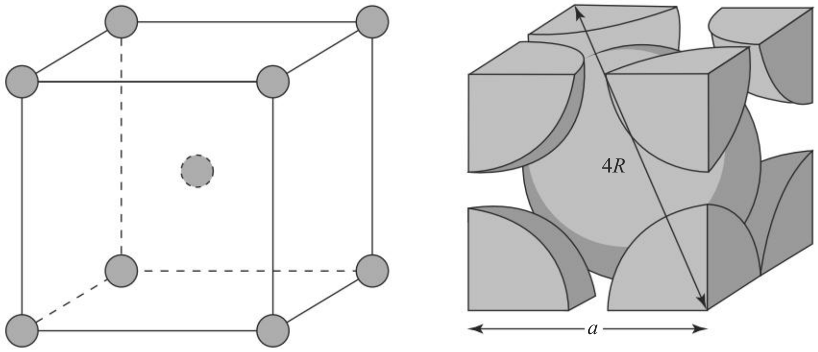


Fig. 3.18 *Body centered cubic structure*

Each atom has 8 equidistant nearest neighbours. Hence the ***coordination number*** is 8. Since there are eight atoms at corners and 1 atom at the body centre, the number of atoms per unit cell is given by

$$n = \frac{1}{8} \times 8 = 2 \text{ atoms}$$

Also, in this structure the atoms touch each other along the body diagonal. Therefore, $(4R)^2 = (\sqrt{2R})^2 + a^2$; where R is the atomic radius and a is the lattice constant

$$16R^2 = 3a^2$$

$$a = \frac{4R}{\sqrt{3}}$$

Atomic packing fraction, $APF = \dfrac{nV_a}{V}$

$$APF = \frac{1 \times \frac{4\pi R^3}{3}}{a^3}$$

$$= \frac{4\pi R^3}{3\left(\frac{4R}{\sqrt{3}}\right)^3}$$

$$= \frac{\sqrt{3}\pi}{8}$$

$$= 0.6802$$

Hence atoms occupy 68.02% of the volume of the unit cell.

SYMMETRY ELEMENTS AND CLASSES OF CRYSTALS

3.10

The various types of crystals may be classified into certain groups depending upon their external and internal symmetry. The most important elements of external symmetry are centre, axes and planes. Now let us define these elements of external symmetry.

Centre of Symmetry A crystal possesses a centre of symmetry when it has a point such that all straight lines drawn through it will pass through a pair of similar points lying on opposite sides of the centre of symmetry and at the same distance from it. The centre of a cube is the centre of symmetry, whereas a tetrahedron has no centre of symmetry.

Axis of Symmetry A crystal is said to possess n-fold rotation axes when the rotation over $(360/n)$ degrees brings the crystal into self-coincidence; for instance, if on rotating a crystal through 180°, 120°, 90°, or 60° about an axis brings the crystal into coincidence with its initial position, then the axis of rotation is said to be 2–, 3–, 4–, 6–, fold axis, On this criterion the principal axes of a cube are 4-fold axes while the body diagonal axes of a cube are 3-fold. A 2-fold axis in a cube would be a line through the coordinates 0, ½, 0 and 1, ½, 1. The c-axis of a hexagonal crystal unit is a 6-fold axis.

Plane of Symmetry When a plane can be drawn in the crystal, which contains the centre of the crystal, such that one half of the crystal is the reflection of the other half, the crystal is said to possess a plane of symmetry. There are, for example, six diagonal planes of symmetry in a cube.

When the seven crystal systems are combined with symmetry operations referred to above, we obtain 32 crystal classes so that all crystals must belong to one of them. In addition to elements of external symmetry of crystals, there exists an internal symmetry in the matter of placement of the atoms on or around lattice points. The symmetry is determined by certain point operations. When a point in a lattice is subject to certain operations which reproduce this point to build up the structure and symmetry of the unit cell, they accomplish essentially the same purpose as the elements of external symmetry. The point operations are rotation, reflection, inversion, rotation inversion, gliding and screw axes.

(i) The operation of ***rotation*** is the same as already dealt with in the heading ‘axis of symmetry’. It may be 2–, 3–, 4–, or 6-fold and indicates that a point within a crystal can be brought into self-coincidence by the operation described.

(ii) A ***reflection*** operation produces the same effect as a plane of symmetry described earlier.

(iii) A crystal possesses an ***inversion*** centre, when for every point located at a distance r relative to the centre there exists a similar point at r and on the same line through this centre.

(iv) A ***rotation–inversion*** operation is one when the crystal can be brought into self-coincidence by combining the processes of rotation and inversion.

(v) A ***gliding*** plane exists in a crystal when it combines a reflection plane with a translation parallel to this plane so that the structure is brought into self-coincidence by the movement and reflection across the specified plane.

(vi) A crystal possesses a ***screw axis*** when a point combines rotation with translation parallel to the axis like the working of a screw. The direction of rotation may be clockwise or counter-clockwise and existence of 3-fold, 4-fold, 4-fold and 6-fold rotations has been discovered.

When all the operations listed in the external as well as internal symmetry categories are applied to the seven crystal systems, we find that 230 groups of crystals result in place of 32 classes obtained by the application only of the external symmetry. The 230 groups are called space groups and all the crystals can be listed into one of these groups.

RELATION BETWEEN LATTICE CONSTANT AND DENSITY

3.11 Consider a cubic crystal having the lattice constant of the cell $= a$

Let n be the number of atoms per unit cell.

The molecular weight of the crystalline substance or compound $= M$

The density of the unit cell $= \rho$

Avagadro's number $= N_A$

[Avagadro's number N_A gives the number of molecules per kilogram mole of the substance]

$$\text{Mass of each molecule} = \frac{\text{M}}{N_A}$$

Hence, mass of n molecules (i.e., mass in each unit cell) $= \dfrac{M}{N_A} n$

Density of a unit cell is

$$\rho = \frac{\text{Mass of the unit cell}}{\text{Volume of the unit cell}}$$

$$= \frac{Mn}{N_A a^3}$$

where volume of the unit cell is a^3

$$a^3 = \frac{Mn}{N_A \rho}$$

$$a^3 = \frac{\text{Molecular weight} \times \text{Number of atoms per unit cell}}{\text{Avogadro's number} \times \text{Density}}$$

The above relation connects the lattice constant a with that of the density of the substance. So we can calculate theoretically the lattice parameter a.

Concluding Remark Engineering branches like Materials science, Electronics, Metallurgy, etc., require basic knowledge of the physics of crystal structure to understand new materials which is helpful to develop new kind of devices.

Solved Problems

1 *Sodium chloride crystals have an fcc crystal structure. The density of NaCl is 2.18 × 10^3 kg/m^3. Calculate the distance between two adjacent atoms. The atomic weight of sodium is 23 and that of chlorine is 35.5.*

Solution Given, the density of NaCl, $r = 2.18 \times 10^3$ kg/m^3

Molecular weight of sodium chloride

$$= \text{Atomic weight of Na} + \text{Atomic weight of Cl}$$

$$= 23 + 35.5 = 58.5$$

Number of molecules per unit cell for fcc structure $= n = 4$

Avogadro's number $= N_A = 6.023 \times 10^{23}$ molecules/mole

Using the relation between density and lattice constant;

$$a^3 = \frac{Mn}{N_A \rho}$$

$$a^3 = \frac{Mn}{N_A \rho}$$

$$= \frac{58.5 \times 4}{6.023 \times 10^{23} \times 2.18}$$

$$a = 5.63 \times 10^{-8} \text{ cm}$$

The distance between adjacent atoms is half of this value $= a/2$

$$= \frac{5.63 \times 10^{-8}}{2} \text{ cm}$$

$$= 2.81 \times 10^{-8} \text{ cm}$$

2 *Copper has a face centred cubic structure and its atomic radius is 1.278 × 10^{-10} m. Calculate the density of Cu. [Given, Atomic weight of Cu = 63.5]*

Solution Radius $r = 1.278 \times 10^{-10}$ m $= 1.278 \times 10^{-8}$ cm

Density can be obtained from the relation $a^3 = \dfrac{Mn}{N_A \rho}$

For fcc structure, atomic radius r and the lattice constant a are related by

$$r = a\frac{\sqrt{2}}{4}$$

$$\Rightarrow \qquad a = 2\sqrt{2}r$$

$$\rho = \frac{Mn}{N_A a^3}$$

$$= \frac{M}{N_A} \frac{n}{(2\sqrt{2}r)^3}$$

$$= \frac{63.54}{6.023 \times 10^{-23}} \frac{4}{\left(2\sqrt{2} \times 1.28 \times 10^{-8}\right)^3}$$

$$= 8.98 \text{ g/cm}^3$$

3 *Draw the family of planes {110} in cubic system.*

Solution The family of planes {110} in cubic system are

(a) (110), (b) (110), (c) (011),
(d) $(\bar{1}\,\bar{1}\,0)$, (e) $(\bar{1}\,0\,\bar{1})$, (f) $(0\,\bar{1}\,\bar{1})$

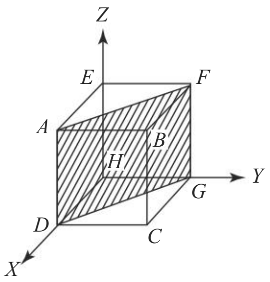
To draw plane (110)

Step 1 Draw a unit cell of cubic structure and suitably choose the three axes.

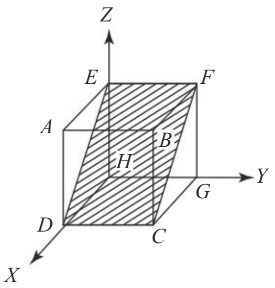
Step 2 For Miller indices of plane ($h\,k\,l$), i.e., for (1 1 0), take the reciprocal of the indices. This will give the intercept along the three axes. $\frac{1}{h}, \frac{1}{k}, \frac{1}{l}$ i.e., as given in the problem $\frac{1}{1}, \frac{1}{1}, \frac{1}{0}$ gives the intercept 1 along the x-axis, 1 along y-axis and • along z-axis. Therefore, the given plane is parallel to z-axis.

Step 3 Draw the plane parallel to z-axis and keeping intercept 1, 1 along x-and y-axis.

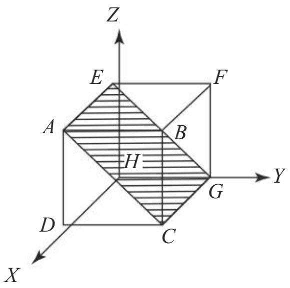
Similarly, other planes can be drawn following the same procedure.



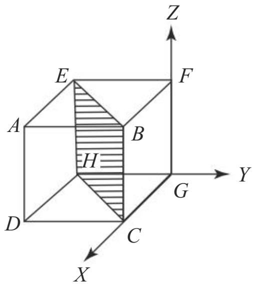
(a) Plane (110)



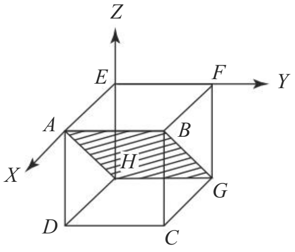
(b) Plane (101)



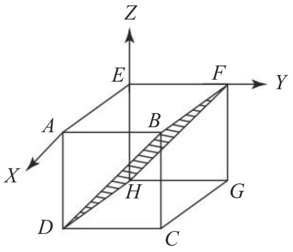
(c) Plane (011)



(d) Plane $(\bar{1}\bar{1}0)$



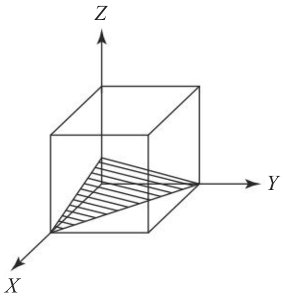
(e) Plane $(\bar{1}0\bar{1})$



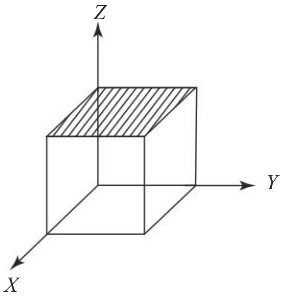
(f) Plane $(0\bar{1}\bar{1})$

4 *Sketch the planes (1 1 2), (0 0 1), (2 0 0) and in a cubic cell.*

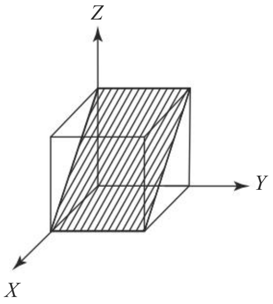
Solution



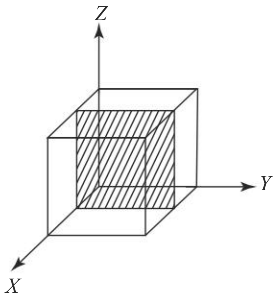
(a) Plane (1 1 2)



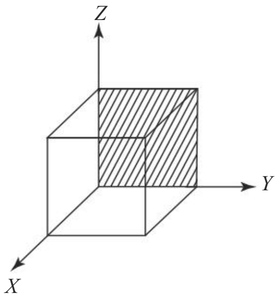
(b) Plane (0 0 1)



(c) Plane (1 0 1)

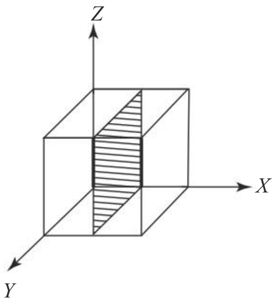


(d) Plane (2 0 0)

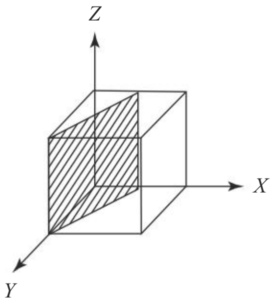


(e) Plane ($\bar{1}$ 0 0)

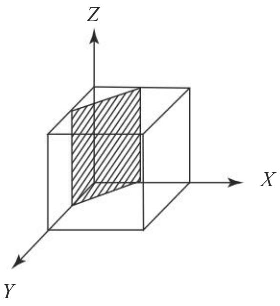
5 *Label the Miller indices for the following figures.*



(a)



(b)

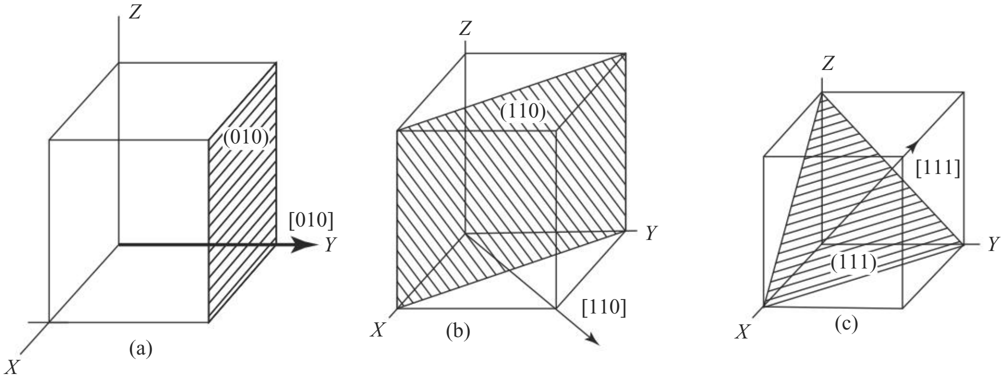


(c)

Solution (a) (2 0 0) (b) (2 1 0) (c) (2 2 0)

6 *Sketch the following pair of planes and directions:*
(a) (010), [010], (b) (110), [110], (c) (111), [111]

Solution



(a) (b) (c)

7 *Aluminium is an fcc crystal with lattice constant a = 0.405 nm. Calculate the number of unit cells present in an aluminium foil of 0.005 cm thickness and two sides of 25 cm length.*

Solution Given, lattice constant $a = 0.405$ nm.

Sides of the Al foil = 25 cm = 25×10^{-2} m.

Aluminium foil of thickness = 0.005×10^{-2} m

$$\text{Volume of the foil} = \text{Area} \times \text{thickness} = (25 \times 10^{-2}\ \text{m})^2 \times 0.005 \times 10^{-2}\ \text{m}$$

$$= 3.125 \times 10^{-6}\ \text{m}^3.$$

$$\text{Volume of the unit cell} = a^3 = (0.405 \times 10^{-9}\ \text{m})^3 = 6.6430 \times 10^{-29}\ \text{m}^3.$$

$$\text{Number of unit cells} = \frac{\text{Volume of the A1 foil}}{\text{Volume of the unit cell}} = \frac{3.125 \times 10^{-6}}{6.6430 \times 10^{-29}} = 4.7041 \times 10^{22}$$

8 *In a given lattice with lattice constants, a = b = 2.5 Å, c = 1.8 Å, α = β = γ = 90° calculate the lattice spacing between (111), (110) and (100) planes. Draw your conclusions.*

Solution For a given tetragonal system, the lattice spacing is given by substituting $a = b$ in the following equation:

$$d_{hkl} = \frac{1}{\sqrt{\left(\frac{h}{a}\right)^2 + \left(\frac{k}{b}\right)^2 + \left(\frac{l}{c}\right)^2}}$$

$$= \frac{1}{\sqrt{\left(\frac{h^2 + k^2}{a^2}\right) + \frac{l^2}{c^2}}}$$

$$d_{111} = \frac{1}{\sqrt{\left(\frac{1^2 + 1^2}{2.5^2}\right) + \frac{1^2}{1.8^2}}} \qquad [h = k = l = 1;\ a = 2.5,\ c = 1.8]$$

$$= \frac{1}{\sqrt{0.32 + 0.308}} = 1.26\ \text{Å}$$

For plane (110), put $h = 1$, $k = 1$, $l = 0$ in the equation below:

$$d_{hkl} = \frac{1}{\sqrt{\left(\frac{h^2 + k^2}{a^2}\right) + \frac{l^2}{c^2}}}$$

$$d_{110} = \frac{1}{\sqrt{\left(\frac{1^2 + 1^2}{2.5^2}\right) + 0}}$$

$$= \frac{1}{\sqrt{0.32}} = 1.768\ \text{Å}$$

For (100) plane, put $h = 1, k = 0, l = 0$, in the equation below:

$$d_{hkl} = \frac{1}{\sqrt{\left(\frac{h^2 + k^2}{a^2}\right) + \frac{l^2}{c^2}}}$$

$$d_{100} = \frac{1}{\sqrt{\left(\frac{1^2 + 0^2}{2.5^2}\right) + 0}}$$

$$= \frac{1}{\sqrt{\left(\frac{1}{2.5^2}\right)}} = 2.5\ \text{Å}$$

As $d_{111} = 1.26$ Å, $d_{110} = 1.768$ Å, $d_{100} = 2.5$ Å $\quad d_{111} < d_{110} < d_{100}$

Or they are in the ratio of

$$d_{100} : d_{110} : d_{111} = 1 : 0.707 : 0.577$$

$$d_{100} : d_{110} : d_{111} = 1 : \frac{1}{\sqrt{2}} : \frac{1}{\sqrt{3}}$$

9 *Lead is an fcc crystal with an atomic radius of 1.746 Å. Calculate the spacing between the planes (200), (220), (222) and (111).*

Solution For fcc crystal structure, the atomic radius r and lattice constant a is related by

$$a = \frac{4r}{\sqrt{2}} = \frac{4 \times 1.746}{\sqrt{2}} = 4.93\ \text{Å}$$

The interplanar spacing is given by

$$d_{hkl} = \frac{a}{\sqrt{h^2 + k^2 + l^2}}$$

$$d_{200} = \frac{4.93}{\sqrt{2^2 + 0^2 + 0^2}} = 2.465\ \text{Å}$$

$$d_{220} = \frac{4.93}{\sqrt{2^2 + 2^2 + 0^2}} = 1.74\ \text{Å}$$

$$d_{222} = \frac{4.93}{\sqrt{2^2 + 2^2 + 2^2}} = 1.423\ \text{Å}$$

$$d_{111} = \frac{4.93}{\sqrt{1^2 + 1^2 + 1^2}} = 2.84\ \text{Å}$$

10 *Classify the following unit cells into proper crystal systems.*

(i) $a = 10.8$ Å, $b = 9.47$ Å, $c = 5.2$ Å, $\alpha = 41°$, $\beta = 83°$, $\gamma = 93°$

(ii) $a = b = 10.73$ Å, $c = 14.3$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$

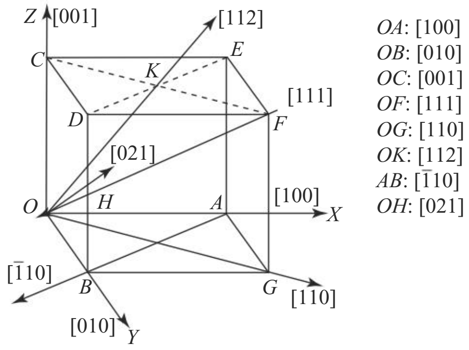
Solution (i) As per the given data, $a \neq b \neq c$, $\alpha \neq \beta \neq \gamma \neq 90°$, which is a characteristic parameter for triclinic system.

(ii) As per the data, it is $a = b \neq c$, $\alpha = \beta = 90°$, $\gamma = 120°$, suggesting that the given system is a hexagonal system.

11 *Draw the following directions in a cubic cell.*

[100], [010], [001], [111], [110], [112], , [021]

Solution



12 *A certain orthorhombic crystal has a ratio of .*

Find the Miller indices of faces whose intercepts are

0.241 : 1 : 0.188

0.858 : 1 : 0.754

0.429 : ∞ : 0.126

Solution The intercepts in terms of the unit axial dimensions are as given below:

$$\frac{1}{2} : 1 : \frac{1}{2}$$

$$2 : 1 : 2$$

$$1 : \infty : \frac{1}{3}$$

Taking the reciprocals of these numbers and reducing them to the smallest whole numbers, the Miller indices come out as follows:

$$(212)$$
$$(121)$$
$$(103)$$

13 *Assuming that lattice points of lattice parameter a in a bcc structure are occupied by spherical atoms of radius r, calculate the free volume per unit cell*

Solution (a) Volume of the unit cell = a^3

For bcc structure; $$a = \frac{4r}{\sqrt{3}} \Rightarrow r = \frac{a\sqrt{3}}{4}$$

Now volume of all the atoms in a unit cell is as there are 2 atoms per unit cell in the bcc unit cell.

$$v = 2 \times \left(\frac{4\pi}{3}\right) r^3$$
$$= \left[\frac{8\pi}{3}\right]\left[\frac{a\sqrt{3}}{4}\right]^3$$
$$= \frac{\pi a^3 \sqrt{3}}{8}$$

Thus, free volume is = $$V - v = \left[a^3 - \frac{\pi a^3 \sqrt{3}}{8}\right] = a^3\left[1 - \frac{\pi\sqrt{3}}{8}\right]$$

Multiple-Choice Questions

1. A space lattice represents a
   - (a) crystal
   - (b) three-dimensional lattice arrays of points only
   - (c) two-dimensional lattice arrays of points
   - (d) amorphous solid

2. The number of atoms per unit cell for a primitive unit cell
   - (a) is always one
   - (b) depends on crystal structure
   - (c) can be any number
   - (d) any one from one, two and four

3. Comparing the Atomic Packing Fraction (APF) for Simple Cubic (SC), Body Centred Cubic (BCC) and Face Centred Cubic structure, we can infer that is correct.
   - (a) $(APF)_{BCC} > (APF)_{FCC} > (APF)_{SC}$
   - (b) $(APF)_{BCC} > (APF)_{SC} > (APF)_{FCC}$
   - (c) $(APF)_{FCC} > (APF)_{BCC} > (APF)_{SC}$
   - (d) $(APF)_{FCC} > (APF)_{SC} > (APF)_{BCC}$

4. Crystals are basically
   - (a) homogeneous and isotropic solids
   - (b) homogeneous and anisotropic solids
   - (c) isotropic solids only
   - (d) highly ordered isotropic solid

5. The relation of length of axes of unit cell and relation of angle between axes in a monoclinic crystal system is

(a) $a = b = c$; $\alpha = \beta = \gamma = 120°$

(b) $a = b \neq c$; $\alpha = \beta = \gamma \neq 120°$

(c) $a \neq b \neq c$; $\alpha = \beta = 90° \neq \gamma$

(d) $a \neq b = c$; $\alpha = \beta = 120°, \gamma \neq 90°$

6. The relation of angle between axes of a triclinic crystal system is

(a) $\alpha = \beta = \gamma = 120°$ (b) $\alpha = \beta = \gamma \neq 120°$

(c) $\alpha \neq \beta \neq \gamma \neq 90°$ (d) $\alpha \neq \beta \neq \gamma$

7. Which of the following crystals belong to the rhombohedral (trigonal) crystal system?

(a) Calcite (b) Diamond

(c) Graphite (d) sodium chloride

8. The coordination number for a faces centred cubic lattice is

(a) 12 (b) 8

(c) 6 (d) 4

9. The atomic packing factor of a faces centred Cubic crystal is

(a) 0.52 (b) 0.68

(c) 0.74 (d) 1

10. Which of the following groups of planes are similar?

(a) (1 1 0), [1 1 0], (1 0 1), (0 1 1)

(b) ($\bar{1}$ 0 1), (1 $\bar{1}$ 0), (1 $\bar{1}$ 0 0)

(c) (1 $\bar{1}$ 1), (1 1 $\bar{1}$), ($\bar{1}$ 1 1)

(d) [1 0 1], [0 1 1], [1 1 0]

11. Which relation between radius of atom and lattice constant represents the fcc structure?

(a) $r = a/2$ (b) $r = a\frac{\sqrt{3}}{4}$

(c) $r = a\frac{\sqrt{2}}{4}$ (d) $r = a$

12. The relation between direction and plane in a cubic crystal is such that

(a) $(1\ 1\ 1) \perp [1\ 1\ 1]$ (b) $(1\ 1\ 1) \parallel [1\ 1\ 1]$

(c) $(1\ 1\ 1) = [1\ 1\ 1]$ (d) $(1\ 1\ 1) \neq [1\ 1\ 1]$

13. Which of the following is not a Bravais lattice?

(a) Simple cubic

(b) Body centred cubic

(c) Face centred cubic

(d) Base centred cubic

14. The ratio of the interplanar distance for a simple cubic system is

(a) $d_{100} : d_{110} : d_{111} = 1 : \frac{1}{\sqrt{2}} : \frac{2}{\sqrt{3}}$

(b) $d_{100} : d_{110} : d_{111} = 1 : \frac{1}{\sqrt{2}} : \frac{1}{\sqrt{3}}$

(c) $d_{100} : d_{110} : d_{111} = 1 : \frac{1}{\sqrt{2}} : \frac{1}{2}$

(d) $d_{100} : d_{110} : d_{111} = 1 : \frac{1}{\sqrt{2}} : \frac{\sqrt{2}}{3}$

15. Lattice constant for a cubic system and density of a crystal are related by

(a) $a = \sqrt{\frac{Mn}{N_A\rho}}$ (b) $a^3 = \sqrt[3]{\frac{Mn}{N_A\rho}}$

(c) $a^3 = \frac{Mn}{N_A\rho}$ (d) $a = \frac{Mn}{N_A\rho}$

ANSWERS TO MCQS

1. (b) 2. (a) 3. (c) 4. (b) 5. (c) 6. (c) 7. (a) 8. (a) 9. (c)

10. (c) 11. (c) 12. (a) 13. (d) 14. (b) 15. (c)

Practice Problems

1. Calculate the interplanar spacing for a (311) plane in a simple cubic lattice whose lattice constant is 2.109×10^{-10} m. **[6.358×10^{-11} m]**

2. Show that $d_{101} : d_{100} : d_{001} :: \frac{a}{\sqrt{2}} : a : a$

3. If the interplanar distance for a plane having Miller indices (110) is 2.86 Å, calculate the lattice constant. **[4.044×10^{-10} m]**

4. The atomic packing efficiency of a crystal is 68%, and the closest distance of approach between its neighbouring atom is 2.62 Å. Calculate density of the crystal if the mass of a single atom is 50.04 amu. **[600 kg/m^3]**

5. Show that the ratio of d_{111} / d_{200} in lead is 1.156.

6. Ni has fcc structure with a lattice constant of 3.52 Å. Calculate the interplanar spacings for planes (101), (123), (320). **[2.49 Å, 0.941 Å, 0.976 Å]**

7. In a crystal of NaCl, calculate the interplanar spacing of planes (111), (110), (100) whose lattice parameter is 5.64 Å.

8. For a body centred cubic lattice, find the number of nearest neighbours, the nearest neighbours distance and the spacing of (110) planes.

9.What are Bravais lattices and Miller indices in crystallographic notations ? Show analytically that the (111) planes are perpendicular to [111] direction in a simple cubic crystal.

10. Draw the (110) and (111) planes and [110] and [111] directions in the simple cubic crystal. What do you infer from these diagrams? Then explain the use of Miller indices notation in crystal studies.

11. Diamond crystallises in ZnS structure. Calculate its density, given that the cube edge for diamond is 3.57 Å and the atomic weight of carbon is 12.01. (Take ***n*** = 8 since, ZnS structure consists of two interpenetrating FCC structure) **[$\rho = 3506$ kg/m^3]**

12. For an fcc cubic crystal, the interplanar spacing of (l 10) plane is 2 Å. Calculate the atomic radius.
[$a = 2\sqrt{2}$ Å, $r = 1$ Å]

13. For planes (101) (112) and (001), sketch the planes in the fcc structure.

14. In a simple cubic crystal, find the ratio of (i) spacing of the (110) and (111) planes, and (ii) nearest neighbour distance to the next nearest neighbour distance. **[(i) 1: 0.816 (ii) 1 : 1.414]**

Short-answer Questions

1. What is a crystal?

2. What are crystalline and amorphous materials?

3. State the difference between crystalline and amorphous materials.

4. Explain in brief the lattice.

5. State the basic conditions for the formation of crystalline solids.

6. What do you mean by space lattice?

7. What is motif or basis? Define crystal structure.

8. What is Bravais and non-Bravais lattice?

9. What is a crystal structure?

10. Define a unit cell and a primitive unit cell.

11. What are lattice parameters? How are they defined? What is their importance?

12. What is coordination number?

13. Define atomic packing factor or density of packing.

14. Give the relation between the density of the crystal and lattice constant.

15. What do you mean by Miller indices for planes and directions?

16. What are crystallographic axes?

17. How are the solid materials classified from the crystal structure point of view?

18. Sketch the plane for the Miller indices of (1 0 0), (1 1 0), (1 1 1).

19. Show the directions for the Miller indices [1 0 0], [1 1 0], [1 1 1].

20. Show by a diagram how the above directions and planes are related.

21. Name the seven crystal systems and state the condition of lattice parameters.

22. What are the possible space lattices of a cubic system?

23. What is the coordination number for simple cubic, body centred cubic and face centred cubic structures?

24. What is the atomic packing factor for simple cubic, body centred cubic and face centred cubic structures?

25. What is the relation between atomic radius and lattice parameters for simple cubic, body centred cubic and face centred cubic structures?

Long-answer Questions

1. (a) Define the terms: single crystal, unit cell, space lattice and Miller indices.

(b) Derive the relation for the number of atoms per unit cell for cubic type of crystal and hence estimate the total number of atoms per unit cell of simple cubic, body centred cubic and face centred cubic.

2. Draw a diagram to indicate the structure of a sodium chloride crystal and with its help explain the following terms: (i) unit cell, and (ii) lattice parameters of the unit cell.

3. Draw different crystallographic planes with indices {1 1 1} of a set of planes in the unit cell of a simple cubic structure.

4. (a) Draw neat diagrams to indicate the important plane systems in a simple cubic crystal and find a relation between the spacing of the planes in the indicated systems.

(b) What are Miller indices? Indicate what should be the Miller indices for a typical plane belonging to each system.

5. Explain the following terms as applied to crystals (i) space lattice (ii) primitives, and (iii) coordination number. Find the coordination numbers for bcc and the fcc structures.

6. Write short notes on Miller indices.

7. What is a primitive cell of a 2-dimensional lattice?

8. Draw the unit cell for a simple cubic structure and show the principle planes in it. How many atoms effectively contribute to this unit cell ?

9. Draw the (110) and(111) planes and the [110] and [111] directions in a simple cubic crystal. What do you infer from these diagrams? Hence, explain the use of Miller indices notation in crystal studies.

10. What are Bravais lattices and Miller indices in crystallographic notations? Show by diagram that the (111) planes are perpendicular to the [111] direction in a simple cubic crystal.

11. With reference to crystals, state what you understand by the terms (i) unit cell, (ii) coordination number, and (iii) packing factor.

12. Derive the relation for the number of atoms per unit cell for a cubic type of crystal and hence estimate the total number of atoms per unit cell of the simple cubic, body centered cubic and face centered cubic.

13. What do you understand by packing density or packing factor? Show that the packing factors for a simple lattice, body centered lattice and face centred lattice are $p/6$, $\frac{\sqrt{3}\,\pi}{8}$, 0.7405 respectively.

14. Define atomic radius. Calculate the atomic radii in case of simple cubic, fcc and bcc lattices.

15. Show that for bcc and fcc crystal structures, the lattice constants are given by $a_{bcc} = 4r/\sqrt{3}$ and $a_{fcc} = 4r/\sqrt{2}$ where r is the atomic radius.

16. What do you understand by Miller indices of a crystal plane? Show that the spacing between consecutive planes defined by Miller indices (hkl) is given by $d_{hkl} = \dfrac{1}{\sqrt{\left(\dfrac{h^2}{a^2} + \dfrac{k^2}{b^2} + \dfrac{l^2}{c^2}\right)}}$, where a, b and c are primitives along the three mutually perpendicular axes.

Questions from Previous Exams

(i) Define the unit cell **(GTU- Jan′ 2009)**

It is a smallest geometric figure, the repetition of which gives the actual crystal structure.

(ii) Define lattice and basis. **(GTU- Jan′ 2009)**

A three dimensional collection of points in space is called crystalline lattice.

A unit assembly of atoms identical in composition, arrangement and orientation is called basis.

(iii) What are Miller indices? Explain with proper example how to determine Miller Indices. **(GTU- Jan′ 2009)**

(Refer Section: 3.6)

(iv) Calculate the interplana spacing for a (3, 1, 1,) plane in a simple cubic lattice whose lattice constant is 2.109×10^{-10} m. **(GTU- Jan′ 2009)**

Given Miller Indices (h, k, l,) = (3, 1, 1,)

Lattice constant $a = 2.109 \times 10^{-10}$ m.

Interplana distance $d = a/(h^2 + K^2 + l^2)^{1/2}$

$d = 2.109 \times 10^{-10}\ \text{m}/(11)^{1/2}$

$d = 2.109 \times 10^{-10}\ \text{m}/3.31$

$d = 6.358 \times 10^{-11}$ m

(v) Which short notes on crystal system **(GTU- Jan′ 2009)**

(Refer Section: 3.4)

(vi) What are Miller indices? Draw the plane from given Miller Indices. (1 1 0), (−1 0 0) (iii) (1 1 2) **(GTU- Jan′ 2010)**

(Refer Section: 3.6)

(vii) Show that $d = a/\ (h^2 + K^2 + l^2)^{1/2.}$ **(GTU- June′ 2010)**

(Refer Section: 3.7)

(viii) Find the Miller indices of a plane which intercepts at $a/2$, $b/2$ along X-axis and Y-axis respectively and parallel to Z-axis in a simple cubic unit cell. Draw a (0 1 1) plane in a cubic system. **(GTU- June′ 2010)**

(Refer Section: 3.6)

(ix) Define the term atomic radius and packing fraction. Calculate the above for SC, FCC and BCC structures. It is defined as half of the distance between two nearest neighbors in a crystal of a pure element. **(GTU- June′ 2010)**

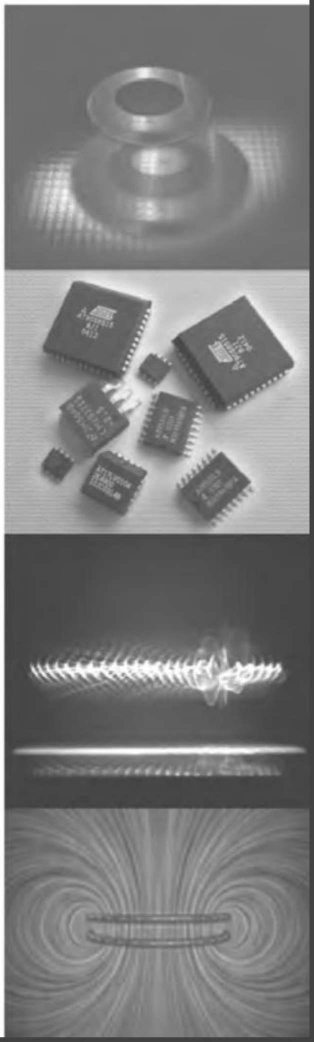
(Refer Section: 3.9)

The McGraw-Hill Companies

Band Theory of Solids

4

If we touch the wiring of our house, nothing will happen to us but unfortunately if the open end of a wire is touched, we could get an electric shock. If, instead, we touch the interior part of an electronic instrument like a mobile phone, we may not feel anything but that the device is working. What is the reason behind this varied electrical nature of different materials? How one can explain it theoretically so that this knowledge could be put to various tasks? You will find in this chapter the basic theory that explains how materials are characterised and categorised according to their electrical properties.



ORIGIN OF ENERGY–BAND FORMATION IN SOLIDS

4.1 According to the free-electron theory of metallic solids, we have assumed that the electrons move in a constant potential region (conveniently taken as zero potential) inside a one-or three-dimensional potential well. This theory explains successfully the phenomena of electrical conductivity in metals, and thermionic emission of electrons from certain metals but fails to explain why there is a difference in electrical conduction for some materials which can act as good electrical conductors, why some are good electrical insulators and why some others are semiconductors.

For the case of a single atom, there are single energy levels like K, L, M levels as we know. But for a solid, a significant change is observed due to the atoms which are closely spaced and, therefore, their energy levels form energy bands as the energy levels of the atoms are very close to each other. If N is the number of atoms in the crystal, it is found that each of the energy levels of an atom splits into N levels of energy. According to Pauli's exclusion principle, only two electrons of opposite spin can occupy the same state, hence the maximum number of electrons which these N states can occupy is $2N$. The explanation is as follows.

If the atoms are brought in close proximity as in crystalline solids, the valence electrons of adjacent atoms interact, so that they constitute a single system of electrons common to the entire crystal, and their outermost electronic orbits overlap.

According to Pauli's exclusion principle, no more than two interacting electrons may have the same energy level, so it is obvious that single lines will be split up into pairs of lines and new levels must be established which are discrete but only infinitesimally different in energy content. The pattern of permissible energy levels become inevitably more complex than for the constituent atoms on their own. Thus, the allowed electronic energy states are perturbed and split into several closely spaced states. The group of such perturbed energy levels which are very close together forms an energy band. Each band contains as many discrete levels as there are atoms in the crystal. The individual energies within the band are so close together that for many purposes, the energy band may be considered to be continuous.

Let us calculate the spacing between different energy levels in an energy band. A crystal of one-milligram mass contains about 10^{19} atoms. If we assume the valency band to be s-band, it contains 2×10^{19} levels. Consider the width of the energy band to be 2 eV. Then, it is obvious that 2×10^{19} levels per milligram spread over an energy bandwidth of 2 eV. Hence, spacing between different levels $= 2/2 \times 10^{19} = 10^{-19}$ eV. Even though energy levels are discrete, the picture of a band as a continuum of energy levels is a very good approximation. This splitting of the single energy level of an isolated atom into a band of energy in the case of a solid is responsible for most of the electrical, magnetic and optical properties of that solid.

4.1.1 Concept of Valence and Conduction Bands

The valence band consists of a group of states containing the outermost or the valence electrons of an atom. It is these electrons which have the highest energy and are most affected in an assembly of atoms constituting the crystal lattice. The band is, obviously, the highest occupied band and like those above it, is shared by the crystal as a whole (Fig. 4.1). Now at low temperatures, all the electrons will exist in the state of lowest possible energy in the absence of an external supply of energy such as radiation. This band may be completely filled when it contains all the electrons that may occupy the band or partially filled when some of the energy levels are vacant. This situation represents a balanced electronic configuration in which there is no net movement of electrons in any particular direction although the orbital of each electron is effectively spread over a sizeable region of the crystal lattice.

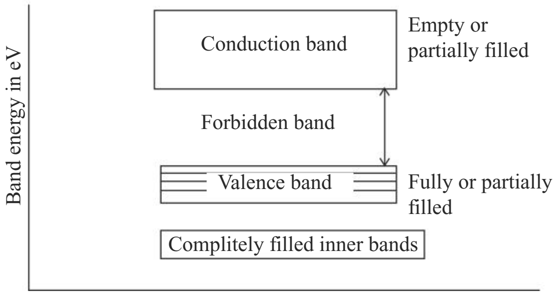


Fig. 4.1 *Energy bands in solids*

The next higher permitted energy band is known as the **conduction band**. It may either be empty or partially filled with electrons. It is the lowest unfilled energy band. In a conduction band, the electrons can move freely and hence are known as **conduction electrons**. The energy gap between the valence band and conduction bands is known as the **forbidden energy gap**. Now the covalent forces of the crystal lattice have their source in the valence band. If a valence electron absorbs enough energy, it jumps across the forbidden energy gap and enters the conduction band. An electron in the conduction band can easily move in the conduction band than it can jump back to the valence band from where it had come earlier. However, if a conduction electron happens to radiate too much energy, it will suddenly reappear in the valence band once again.

If an electron is ejected from the valence band, a covalent bond is broken and a positively charged hole is left behind. This hole can travel to an adjacent atom by acquiring an electron from that atom which involves breaking an existing covalent bond and then making a new covalent bond by filling up the hole. It is to be noted carefully that holes are filled by electrons which move from adjacent atoms without passing through the forbidden energy gap as shown in Fig. 4.2.

Here, the flow of conduction electrons and flow of holes in the valence band is independent of each other. This indicates the hole current in the valence band and electron current in the conduction band. Although holes flow with ease, they experience more opposition than electron flow in the conduction band, and it is observed that conduction electrons move almost twice as fast as the holes.

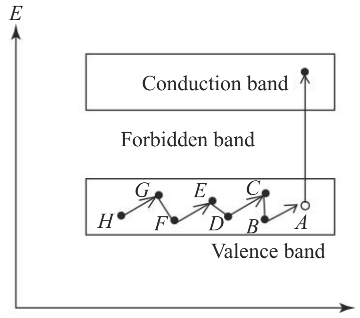


Fig. 4.2 *Valence bands, conduction band and forbidden band*

CLASSIFICATION OF MATERIALS INTO CONDUCTORS, SEMICONDUCTORS AND INSULATORS

4.2

The electrical conduction properties of different elements and compounds can be explained in terms of the electrons having energies in the valence band and conduction bands. The electrons lying in the lower energy bands, which are normally filled, play no part in the conduction process.

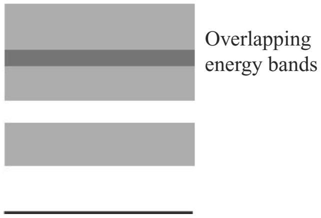


Fig. 4.3 *The overlapping energy bands in solids*

4.2.1 Conductors (Metals)

It is observed that the majority of pure elements in crystalline form have metallic properties. They have a large number of free electrons available for electrical conductions and, therefore, they are good conductors of electricity and heat. In conductors, the energy gap between the valence bands and conduction bands does not exist and the two energy bands actually overlap because the lowest levels in the conduction band require less energy than the highest levels in the valence band. Some electrons 'overflow' into levels in the valence band (Fig. 4.3).

In this case, the electrons have a continuous distribution of permitted energies. One can say that the valence band is only partially filled as in the case of sodium. Figure 4.4 illustrates this case in which (a) the energy levels of a sodium atom, and (b) the energy bands of sodium metal are depicted. A sodium atom has one 3s electron in its valence shell. Therefore, the lowest energy band or the valence band in a sodium crystal is half-filled because each level in the band, like each level in the atom, is able to contain two electrons having opposite spins.

When an electric field is set up across a piece of sodium metal, electrons easily acquire enough energy to raise them to the next higher level while remaining in their valence energy band. This acquired energy is in the form of kinetic energy and permits motion of electrons towards a positive pole, and the motion of electrons contribute to an electric current. Since the raising of some electrons to higher energy levels also opens up lower levels which were formerly occupied, electrons moving against the electric field can be decelerated to lower energy levels so the resultant movement of the electrons is in one direction. This explains how sodium and alkaline earth metals are good conductors of electricity as compared to other crystalline

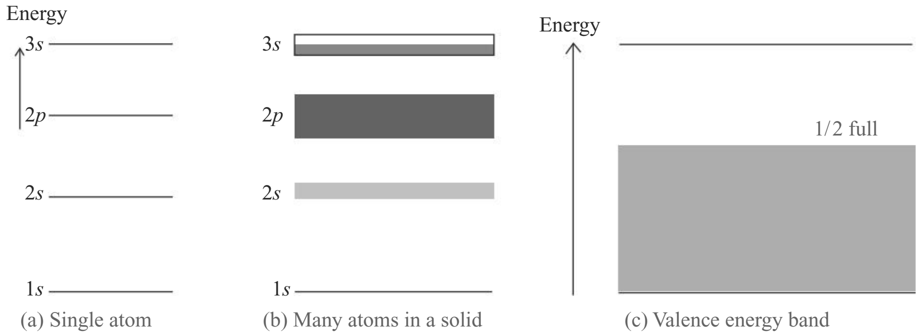


Fig. 4.4 *Energy bands for sodium (a) Single atom, core electron levels having two electrons per level; (b) Atoms in a solid form energy bands due to many atoms being closely spaced; (c) Half-occupied valence band for sodium metal.*

solids with energy bands that are only partially filled. Metals, therefore, have high electric conductivity at all temperatures, their resistivity being very low of the order of 10^{-8} ohm-metre. The proportion of valence electrons that are free to be accelerated increases slightly with temperature and this tends to increase the conductivity. However, this effect is more than counter–balanced by the increasing interaction between the electrons and the crystal lattice as the temperature rises, which diminishes the mobility of the electrons. The electrical conductivity of a metal, therefore, falls with rising temperature, i.e., its resistivity rises.

The freedom of movement of the valence electrons is also responsible for the high thermal conductivity of metals. Due to the number of effective free electrons increasing (slightly) with temperature, a temperature gradient through the material produces a gradient in the concentration of the free electrons in it. The rapid diffusion of electrons then tends to equalise the electron concentration and conveys the surplus energy between the regions of high and low temperatures. 'Electrical conduction' consists of the diffusion of electric charge through the material under conditions of nonuniform electrical potential; while 'thermal conduction' consists of diffusion of thermal energy under conditions of nonuniform temperature. In metals, both processes take place essentially by the same mechanism, and some connection between the two coefficients of conductivity is therefore to be expected. It is experimentally observed that the ratio of thermal conductivity (K) and electrical conductivity (σ) at a given temperature is approximately constant for a large range of metals and independent of the conductor. This is the well-known **Wiedemann–Franz's Law**. It is observed that this ratio is directly proportional to temperature in Kelvin, i.e., its temperature coefficient should be 3.67×10^{-8} per kelvin. This is in agreement with the theory provide strong confirmation of the electron theory of metals. The law of Wiedemann and Franz does not hold for nonmetals.

In addition to the conduction of heat through a metal by the free electrons in it, there is also transference of thermal energy by the vibrations of the ions of the crystal lattice. The rate of conduction of heat on this account is comparable with that occurring in poor (nonmetallic) conductors of heat, and it is normally swamped by the much more rapid process of electronic conduction. It, therefore, has little effect on the validity of the Wiedemann–Franz Law.

4.2.2 Semiconductors

Semiconductors are a class of materials for which the electrical conductivities are intermediate between conductors and insulators, e.g., silicon (Si), germanium (Ge), carbon (C), selenium (Se), etc. They exhibit an electrical resistivity of the order of 10^2 ohm-m, with a dielectric constant around 20 and have a varying positive or negative temperature coefficient of resistance.

They exhibit crystal structure of diamond and, as in diamond, an energy gap separates the top of the filled valence energy band where the electrons are all in the ground state, from a vacant higher band (conduction band) as in insulators. The width of the forbidden gap between the valence and conduction band is small, for example, in silicon it is only 1.1 eV (Fig. 4.5) as compared to 6.0 eV in diamond.

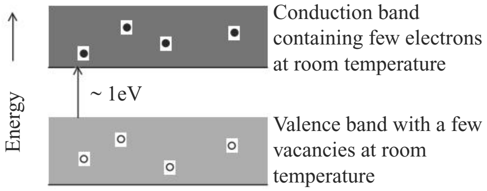


Fig. 4.5 *Energy band diagram for a semiconductor with smaller energy gap between conduction band and valence band*

This implies that at low temperatures, there will be virtually no electrons in the conduction band and the conductivity will be very low because enough energy is required to raise an electron in silicon to move it across the forbidden gap to the conduction band. But at ordinary room temperature, a small amount of electrons have sufficient kinetic energy of thermal origin to jump across the forbidden gap and enter the conduction band above it. These electrons are sufficient to permit a limited amount of current to flow when an external electric field is applied. Therefore, silicon has an electrical resistivity lying between that of conductors and insulators and so is termed a semiconductor. Other substances belonging to this category are germanium and galena (PbS) which was the crystal of early radio crystal sets. In germanium, the width of the forbidden gap is 0.72 eV, somewhat less than that of silicon (1.1 eV) at a given temperature. Therefore, germanium has more electrons in the conduction band and its conductivity is therefore higher than that of silicon. Given below is the comparative resistivities of different classes of materials:

Resistivity of copper (good conductor) = 1.7×10^{-8} ohm-m

Resistivity of germanium (semiconductor) = 0.5 ohm-m

Resistivity of silicon (semiconductor) = 2×10^2 ohm-m

Resistivity of glass (insulator) = 9×10^{11} ohm-m

4.2.3 Insulators

The materials such as glass, plastics, rubber, ceramics, and diamond have virtually all the valence electrons bound very firmly to the nuclei of their atoms and, therefore, in the absence of free electrons, the conductivity is extremely low. According to the energy band theory of solids, the conduction band is completely empty, the valence band is fully occupied by electrons at 0 K and a large forbidden gap exists between them as shown in Fig. 4.6. The forbidden gap is greater than 5 eV so that even in the presence of an external electric field, electrons do not acquire sufficient energy to cross the forbidden gap. Such materials are termed insulators on account of their high resistivity, which is of the order of 10^7 ohm-m.

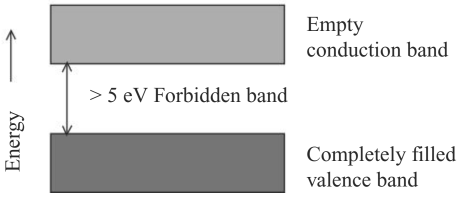


Fig. 4.6 *The energy band diagram of an insulator with large forbidden band between conduction band and valence band*

Increase in temperature enables some electrons to go to the conduction band which, in fact, accounts for the negative resistance temperature coefficient of insulators.

The energy gaps in the case of insulators and semiconductors vary in size from material to material.

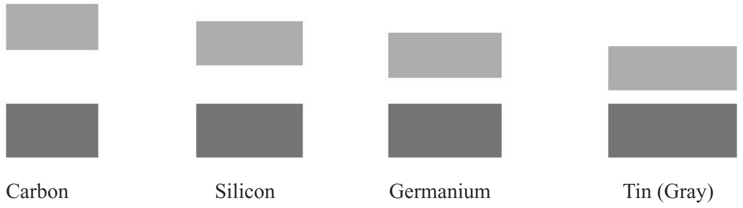


Fig. 4.7 *Comparative size of energy gaps (not to scale) for diamond (carbon, silicon, germanium and tin (gray)*

Figures 4.7 depicts the comparative size of the energy gaps for diamond (carbon), silicon, germanium and tin (grey). Each has the same diamond cubic lattice structure. Therefore, they are similar except for the atomic structure beneath their valence shells. Each atom is surrounded by four others occupying the corner points of a tetrahedron to which it is bound by electron-pair bonds, The structure may be described by an FCC point lattice in which each lattice point corresponds to two atoms, one located at (0, 0, 0) and the other at $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$. Tin (grey), with a smaller energy gap, will conduct electricity much more readily than elements with bigger gaps because there is a greater probability that an electron from the valence band can be raised to the next higher energy band or conduction band. In the case of diamond, however, the two lower energy bands are completely filled with electrons and there is a forbidden gap of 6 eV between the top of the higher of these bands and the conduction band above it. This means that at least 6 eV of additional energy must be given to an electron in the diamond crystal, if it is to have sufficient kinetic energy in order to be transferred to the conduction band. This is about 100 times as great as the average energy of vibration of the atoms of the crystal lattice.

This tremendous amount of energy cannot be easily provided by an electric field less than 6×10^5 volts/cm well over 10^{10} times greater than the electric field intensity needed to cause a current to flow in sodium. Generally, all materials with a forbidden gap larger than diamond (6 eV) are insulators. In this case, the chance of electrons being set free for conduction is almost zero. However, if a sufficiently strong electric field is applied, the electrons may gain the required energy from it and the insulation may break down. The insulation strength, therefore, depends on the width of the forbidden band between the conduction and valence bands. The chance of electrons appearing in the conduction band increases when the temperature is raised. The conductivity, therefore, increases with temperature and may even become relatively high. For example, if a glass rod is heated to a temperature just below its melting point, a large enough current may be passed through it to maintain its temperature and even to melt it. The chief application of the insulators is in electrical appliances where they are used for insulating electric connections.

To summarise the above, we can conclude the following:

(a) Materials with unfilled or overlapping energy bands are electrical conductors with electrical conductivity of the order of 10^8 or 10^7 mhos per metre.

(b) Materials with filled energy bands and large forbidden zones (energy gaps) are electrical insulators unless a considerable energy is supplied to raise electrons beyond the energy gap into the next energy band.

(c) Materials with filled energy bands and small forbidden zones are semiconductors. Their conductivities are intermediate between conductors and insulators.

TYPES OF DIODES

4.3

Adding a known amount of impurities (dopants) in a controlled fashion in any intrinsic (pure) semiconductor material can produce an extrinsic semiconductor (doped) materials on the basis of the dopant they can be classified as *n* type (pentavalent impurities, electrons as majority carriers and holes as minority carriers) or *p* type (trivalent impurities, holes as majority carriers and electrons as minority carriers) semiconductors.

These extrinsic materials find hardly any use individually, but find tremendous use if diodes are made out of them. Diodes are basic electronic devices and a major breakthrough in the field of solid state electronics. These devices have changed modern life with much cheaper and economic size (miniaturization) electronic circuits. One can say that modern life cannot be imagined without these devices. It is a must-know technology for every scientist and engineer. It begins with a basic junction diode.

4.3.1 Basic *p-n* Junction Diode or Simple Diode

When a p-type semiconductor is suitably joined to an n-type semiconductor, the contact surface is called a ***pn*** **junction,** *and the produced junction is called a* ***pn*** **junction diode** or **semiconductor diode** or **junction diode.** It can be formed either by growing the junction, by alloying or by epitaxing. The basis of most semiconductor devices is one *pn* junction diode. Therefore, it would be prudent to understand the formation, characteristics and current components of this device.

- ***Formation of the Junction***

A simple semiconductor diode is fabricated by joining a *p*-type and an *n*-type semiconductor using a special technique (either by growing the junction or by alloying or by epitaxing). A *p-n* junction is a metallurgical boundary in a semiconductor crystal in which holes are the majority carriers on one (*p*) side and electrons are the majority carriers on the other (*n*) side. The boundary between one region of a semiconductor with *p*-type impurities and another region containing *n*-type impurities is known as a *p-n* junction which is depicted in Fig. 4.8. Semiconductor are joined and majority carriers diffuse across the junction at the instant of formation of the junction.

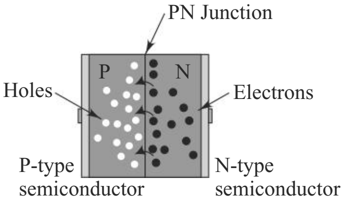


Fig. 4.8 *p-n junction: a p-n junction is formed when a p-type and n-type*

It is important to note that apart from holes as majority carriers, very few thermally generated electrons are also available as minority carriers in *p*-type semiconductors. Similarly, electrons are majority carriers and thermally generated holes are minority carriers in *n*-type semiconductors. The impurity atoms on each side are ionized and firmly fixed in the lattice through covalent bonds. Before contact, each side is electrically neutral; one hole is associated with each ionized acceptor in the *p*-type material and one electron with each ionized donor on the *n*-side.

On joining these *p* and *n* type semiconductors, because of the difference in concentration (concentration gradient) of electrons and holes on the two sides of the junction, the electrons in *n* region tend to diffuse (move towards) to the *p* region and the holes from *p* region tend to diffuse into the *n* region. This process is called '*diffusion*'. As a hole recombines with an electron, both the hole and electron disappear. Consequently, the majority holes moving out of the *p*-region leave behind on the *p*-side some of the acceptor atoms (designated as N_A) bound to the lattice, thus exposing negative ions in a previously neutral region (see Fig.4.9). Similarly, the electrons diffusing from *n* to *p*-side, undergo recombination and leave behind the uncompensated positively ionized donor (designated as N_D). The diffused charge carriers combine at the junction to neutralize each other. Hence development of immobile positive ions in the *n* region and negative ions in the *p*-region near the junction takes place, and a charge-free space called the '*depletion layer*' or '*space charge region*', of width of the order of few microns, is formed near the junction. The diffusion process continues till the materials on both sides attain equilibrium. The immobile ions form parallel rows of opposite charges across the depletion layer. Because of this charge separation, an electric potential V_B develops across the junction under equilibrium condition. This is shown in Fig. 4.9. This potential is called the '*junction potential*' or the '*barrier potential*'. The potential V_B is also called the *contact potential*, and it is an internal potential. It prevents further diffusion of majority charge carriers across the junction. It should be noted that outside the barrier, on each side of the junction, the material is still neutral.

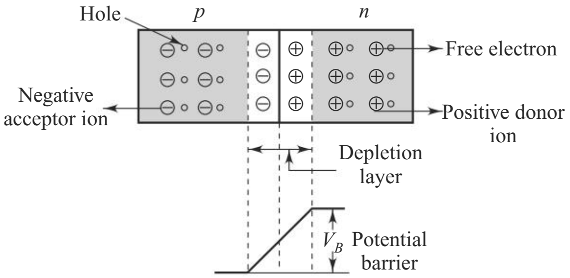


Fig. 4.9 *p-n junction under equilibrium condition and potential distribution curve showing depletion layer and potential barrier* V_B.

From the potential distribution curve, it is clear that a potential barrier V_B is set up which gives rise to electric field ($\therefore\ V_B = Ed$). This field prevents the respective majority carriers from crossing the barrier region. Thus, it is also clear that outside the depletion layer, on each side of the junction, the material is still electrically neutral. Only inside the barrier there is positive charge on *n*-side and negative charge on *p*-type.

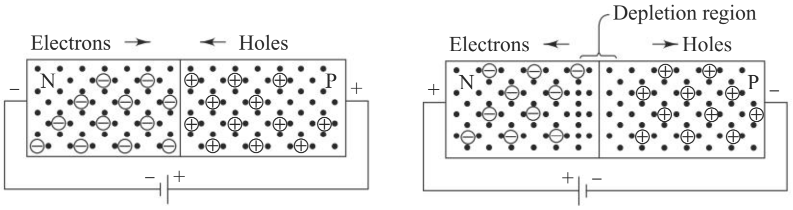
• *Applying Voltage (Biasing) across p-n Junction*

(a) Unbiased or Zero biased Junction

When no external voltage is applied across the junction, it is in steady state condition so no net charge transfer is taking place and therefore, no current flows through it.

(b) Forward Biasing

When an external dc voltage is applied to the *p-n* junction such that the positive terminal of battery is connected to *p*-region and negative terminal is connected to *n*-region then the junction is said to be forward biased. This is shown in Fig. 4.10(a) and Fig. 4.11.



(a) Forward biased *p-n* junction (b) Reverse biased *p-n* junction

Fig.4.10 *(a) Forward biased p-n junction; (b) Reverse biased p-n junction*

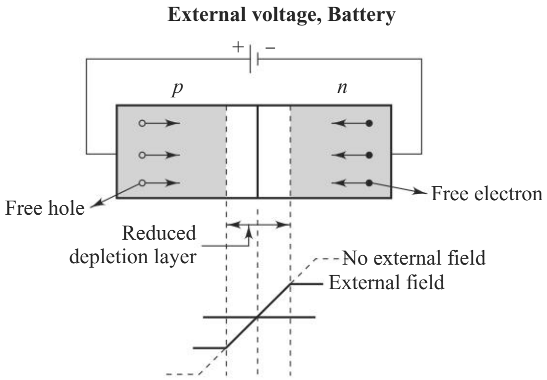


Fig.4.11 *Forward biased p-n junction Diode*

The initial effect of forward bias is to reduce/cancel the potential barrier and the width of the depletion layer. Thereafter, the holes from the *p*-region are repelled by the positive terminal of the battery towards the junction and the electrons from *n*-region are repelled by the negative terminal of the battery towards the junction. The applied voltage should be high enough so that electrons and holes acquire enough energy to overcome the potential barrier at the junction and cross the junction. The barrier voltage is usually small [0.3 volts for Ge diode and 0.7 volts for Si diode]. Hence, a small forward voltage is sufficient to completely eliminate the barrier.

Once the barrier is overcome, the junction resistance becomes almost zero. Hence, a large current flows through the junction at small voltages. The direction of current gives the forward direction of *p-n* junction. The current range is usually in mA. This is called *forward current.* Figure 4.11 shows the reduced depletion width along with the potential hill curve for external field and no external field.

(c) Reverse Biasing

When the external dc potential difference is applied to the *p-n* junction such that the positive terminal of the battery is connected to *n* region and the negative terminal of battery is connected to the *p* region and then the junction is said to be 'reverse biased'. As shown in Figs. 4.10 (b) and 4.12.

The reverse bias increases the potential barrier and also the width of the depletion layer. Under the reverse biased mode, the electrons in *n*-region and holes in *p*-region (majority carriers) move away from the junction. Hence there is no current due to the majority carriers. Here, junction resistance is very high. However in practice, a very small current (of the order of μA) flows in the circuit under reverse bias. This is the reverse current, which is due to the movement of minority free electrons in *p*-region and minority free holes in *n*-region across the junction. Even with a large reverse bias, the current flow is very small. It is of the order of *nA*.

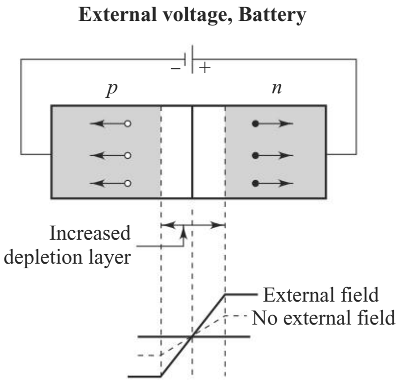


Fig. 4.12 *Reversed biased p-n junction diode*

Thus the *p-n* junction conducts when it is forward biased and practically does not conduct when it is reverse biased. This unilateral conduction can be used for rectification (the process of converting ac voltage into dc voltage).

• *Energy Bands of p-n Junction Diode under Biasing*

(a) *Unbiased or Zero biased Junction*

We learnt that the Fermi level in the *p*-type material is located close to the top of the valence band while in the *n*-type material, the Fermi level is close to the bottom of the conduction band.

While forming the *p-n* junction, due to a concentration gradient of carriers on the two sides of a junction, the carriers cross (diffuse) the junction. Such crossing continues till the Fermi levels on the two sides of the junction equalize.

The band edges in the two regions shift themselves so that the Fermi levels are properly aligned. The band picture of the unbiased *p-n* junction at equilibrium is shown in Fig. 4.13.The conduction band in the *p* region is above the conduction band in the *n* region by an energy eV_B, where V_B is the potential barrier across the junction.

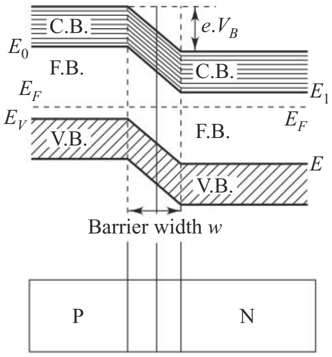


Fig. 4.13 *Energy bands in unbiased p-n junction diode*

The majority electrons in the conduction band of *n* region will face the potential barrier across the junction and hence cannot cross the junction. Thus, there cannot be any current flow.

(b) *Forward Biasing*

An external battery (dc) of voltage *V* is connected with its positive terminal to *p* region and negative to *n* region. The band picture of *p-n* junction under forward bias is shown in Fig. 4.14. Due to this forward bias, the equilibrium conditions are disturbed and the positions of Fermi level and hence the energy bands are changed. The negative terminal of the external voltage source causes an increase (eV) in electron energy and an upward shift of all energy levels on the *n*-side. Similarly, the positive terminal connected to *p* side causes an increase in hole energy and hence a lowering of all levels on *p*-side. The energy bands adjust their positions so as to suit the rise of Fermi level. Hence, the potential barrier is reduced to $e(V_B - V)$ and the barrier width is reduced. Hence, the energy required by the majority carriers to move into opposite regions is reduced to $e(V_B - V)$ from eV_B. The reduction in the height of the potential barriers does not affect the minority carrier movement. Therefore, a net current flows through the junction, which is determined by the diffusion of majority carriers. The electrons in *n* region now face a low potential barrier and can cross it more easily.

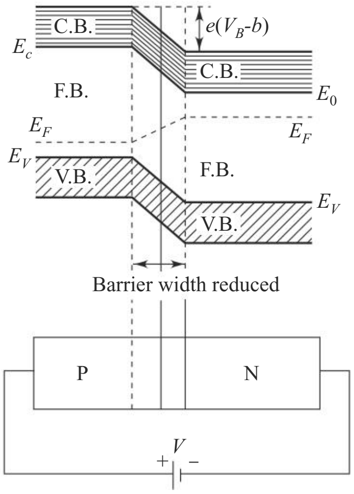


Fig. 4.14 *Energy bands of forward biased pn junction diode*

To conduct the diode, the potential of external battery should be more than the barrier potential.

(c) *Reverse Biasing*

The external battery of voltage *V* is connected with its positive terminal to *n* region and negative terminal to the *p* region. The energy-band diagram for a diode under reverse bias is shown in Fig.4.15.

So, the energy of electrons in the *n* region reduces by an amount eV. Hence the Fermi level on *n* side is shifted down by eV and the potential barrier increases to $e(V_B + V)$, thereby increasing the barrier width.

The majority electrons in the *n* region now face a greater potential barrier in crossing the junction (but the minority carriers can easily cross it).

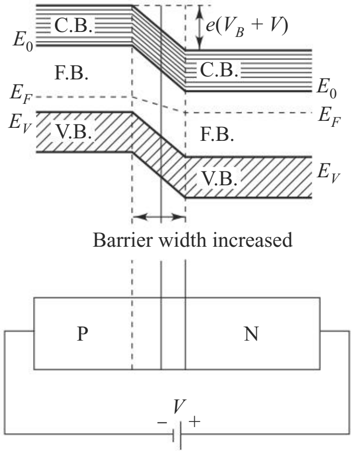


Fig. 4.15 *Energy bands of reverse biased pn junction diode*

• *Voltage–Ampere (V–I) Characteristics of Diode*

A graph that shows the variation in current in a device with the variation of applied voltage across it is called the volt − ampere characteristic of that device. Figure 4.16 shows the volt − ampere characteristic of silicon and germanium diodes. It is seen that the characteristic is not linear and hence the diode is a non-linear device. The diode acts like a closed switch in forward bias condition as it allows to conduct the diode(large current) ; and acts an open switch in reverse bias condition as it does not allow to conduct the diode (insignificantly low current or known as leakage current).

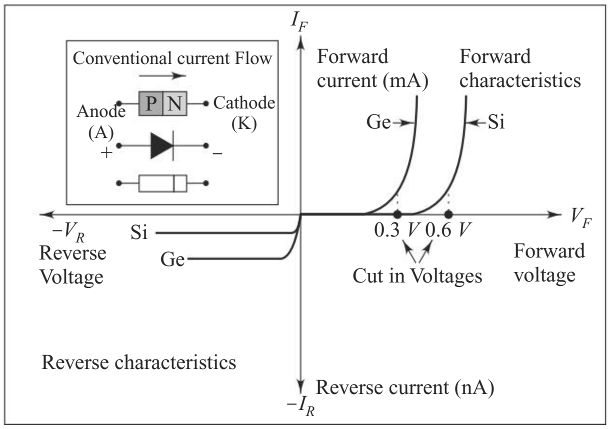


Fig. 4.16 *V-I Characteristics of p-n junction diode along with the symbol and conventional current flow direction*

4.3.2 Reverse Saturation Current

The reverse current I_0, is constant which is caused by the minority carriers. They are finite in number at a given temperature. As long as the temperature remains constant, the rate of generation of minority carriers remains constant. Therefore, the current due to their flow is the same whether the applied reverse bias voltage is small or large. Hence, the drift current due to minority carriers is known as reverse saturation current.

Differences Between Si and Ge Diode Characteristics

1. The cut-in voltage for silicon (Si) diode is 0.7 volts whereas for germanium (Ge) diode, it is 0.3 volts.

2. As in silicon diode, the reverse saturation current is very small and it can withstand high reverse voltage, it is widely used in practice..
3. The reverse saturation current in a silicon diode is very small whereas for germanium diode it is much higher.
4. The reverse breakdown voltage of a silicon diode is greater than that of germanium diode. Hence, a silicon diode can withstand higher reverse voltage compared to germanium.

Temperature Dependence of V-I characteristics of Diode

The rectifier equation or diode equation,

$$I = I_0\left(e^{\frac{eV}{kT}} - 1\right)$$

is applicable for both types of biasing if $V = V_F$ for forward bias and $V = -V_R$ in case of reverse bias. Therefore, diode characteristics are given by $I = I_0\,(e^{V/\eta V_T} - 1)$ where I_0 is reverse saturation current and depends on temperature, and V_T is volt equivalent of temperature.

$$V_T = kT$$

where k is Boltzmann's constant = 8.62×10^{-5} eV/K

T = Temperature in kelvin (K)

At room temperature (27°C or $273 \pm 27 = 300$ K)

$$V_T = 8.62 \times 10^{-5} \times 300$$

$$V_T = 26 \text{ mV}$$

It is also dependent on temperature. Due to these two parameters, diode characteristics are temperature dependent. On increasing the temperature, the diode gets more kinetic energy, and hence more *e-h* (electron-hole) pair increases which increases electrical conductivity and current. The reverse saturation current (I_s) of a *p-n* junction diode increases approximately 7% per °C for both silicon and germanium and I_0 becomes double for every 10°C rise in temperature. Figure 4.17 shows effect of temperature on *V-I* characteristics of *p-n* junction diode.

It is clear that the rise in temperature results in decrease in forward voltage across the diode and increase in forward current during forward bias, whereas the reverse saturation current increases during reverse bias.

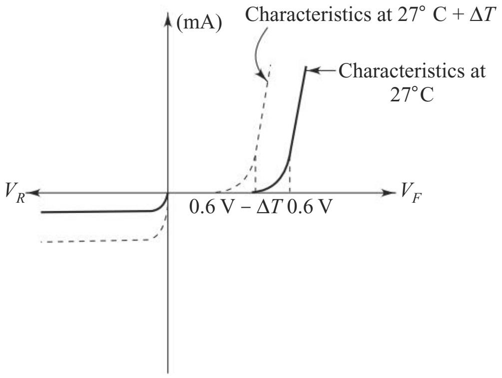


Fig. 4.17 *Effect of temperature on V-I characteristics of p-n junction diode*

4.3.3 Zener Diode

When the reverse bias across a diode is increased, a critical voltage called *breakdown voltage* is reached where the reverse current increases sharply to a high value. The breakdown region is known as the knee of the reverse characteristics as shown in Fig. 4.18. This phenomenon was explained by *C.* Zener, and so this mechanism is known as ***Zener breakdown***, the breakdown voltage is known as ***Zener voltage*** V_z and sudden increase in current is known as ***Zener current***.

The breakdown or Zener voltage depends upon the amount of doping. If the diode is heavily doped, the depletion layer will be thin and consequently, the breakdown of the junction will take place at a lower reverse voltage. Thus, Zener diode is a semiconductor diode specially designed to operate in the breakdown region of the reverse bias. Zener diodes are always mostly used in reverse bias condition. By varying the impurity concentration and other parameters, it is possible to design the breakdown voltage to suit specific applications. In Zener diodes, the breakdown phenomenon is reversible and harmless.

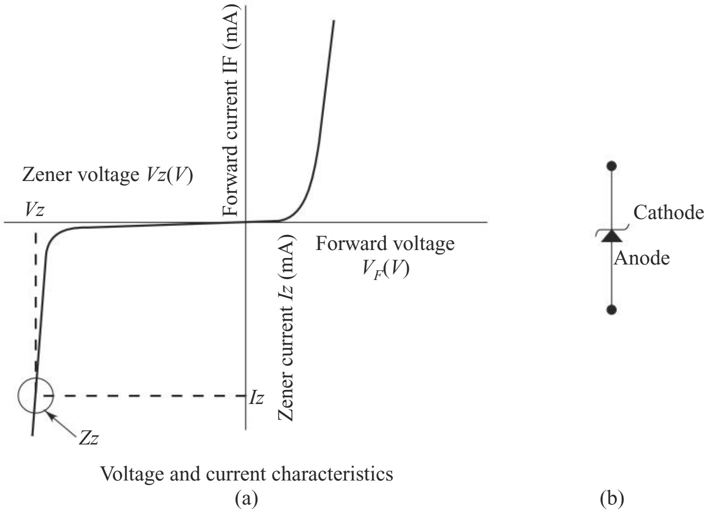
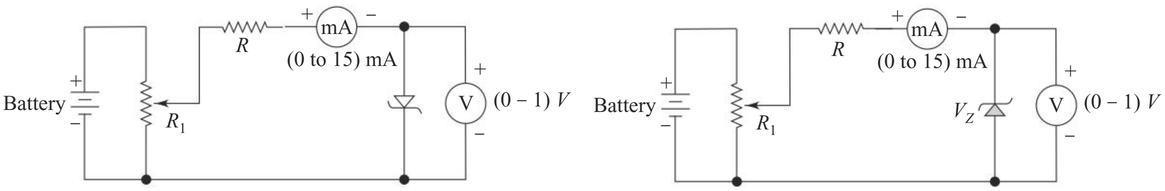


Fig. 4.18 *Zener diode (a) V-I characteristics (b) Symbol*

Figure 4.18 shows the circuit symbol and *V-I* characteristics of a Zener diode. The symbol is similar to that of *p-n* junction diode, except the bar is replaced by a Z shape as cathode. Even the forward characteristics are also alike a *p-n* junction diode. The Zener diode is rarely used in forward bias conditions. There is a significant and much useful difference in its reverse characteristics. Reverse voltage remains flat (constant) on increasing the reverse voltage, however the reverse current increases sharply at a particular voltage. This voltage at which such a sudden increase in reverse current takes place is called Zener breakdown voltage or Zener voltage, V_Z. The breakdown voltage of an ordinary diode is high, but if a reverse current above that value is allowed to pass through it, the diode is permanently damaged. Zener diodes are designed so that their Zener voltage is much lower, for example, just 2.4 volts. When a reverse voltage above the Zener voltage is applied to a Zener diode, there is a controlled breakdown which does not damage the diode. In the Zener region, the voltage across the Zener diode remains constant but the current changes depending on the supply voltage. The voltage drop across the Zener diode is equal to the Zener voltage of that diode no matter how high the reverse bias voltage is above the Zener voltage. To keep V_Z lower, we must increase the doping, thus location of Zener region can be adjusted by adjusting its doping level. Zener diodes are available in the range of 3 V to 220 V.

Experimental set-up to plot the V-I Characteristics of Zener Diode (Fig.4.19)



(a) Forward characteristics (b) Reverse characteristics

Fig. 4.19 *Circuit diagram to plot V-I characteristics of Zener diode under (a) forward bias, and (b) reverse bias conditions.*

Connect the circuit as shown in Fig. 4.19. Vary voltage across the diode and note down the readings of the voltmeter and ammeter. From these observations, we can obtain *V-I* characteristics of a Zener diode.

However, in reverse bias, increase the voltage across the Zener in steps of 1 V by adjusting the potentiometer R_1 and note down the corresponding Zener current, till you reach the specified Zener voltage V_Z. As the Zener voltage is reached, the voltage across the Zener diode will remain constant and current will increase. So adjust the potentiometer R_1 to increase the Zener current by 1 mA step and note the corresponding V.

4.3.4 Breakdown Mechanism in a Semiconductor Diode

From Fig.4.16, it is seen that the reverse saturation current in a semiconductor diode is constant irrespective of the applied reverse bias voltage. However, beyond a particular value of reverse bias, the reverse current increases suddenly because of the phenomenon of breakdown. Ordinary diodes are never operated in this region, because of the irreversible breakdown phenomenon and permanent damages to the diode. However, in zener diodes the breakdown is adjusted with controlled doping. Thus, this breakdown is a controlled one and hence it is reversible. Therefore, this diode can be used in reverse bias mode in many important applications. The breakdown occurs in reverse bias diode mainly due to two different mechanisms.

(i) Avalanche Breakdown

This type of breakdown occurs in lightly doped junctions when high reverse voltage is applied across the junction. When high reverse bias is applied to the diode, the electric field in the depletion region becomes sufficiently high. Due to this high electric field, minority electrons entering the depletion region from the *p*-side will acquire high kinetic energy. The high energy electrons will collide with host atoms and will remove valence electrons from some of the covalent bonds. The new carriers in turn, will produce additional charge carriers and the process multiplies and hence an avalanche of carriers will be produced in a very short time to give large reverse current. The avalanche breakdown is self-sustaining as long as the reverse voltage is present across the depletion region. The breakdown voltage is found to increase with temperature.

(ii) Zener Breakdown

The Zener breakdown mechanism occurs in thin, abrupt and heavily doped *p-n* junctions and requires relatively low reverse voltage for its operation. When the *p* and *n* regions of the diode are heavily doped, the depletion region becomes very thin, of the order of 10 nm. Application of reverse-bias voltage causes the conduction and valence bands to bend to the extent that the *n*-side conduction band appears opposite to the *p*-side valence band as shown in Fig. 4.20. As a result, a number of vacant states in the *n*-side conduction band are brought directly opposite to a large number of filled states in the *p*-side valence band (Fig. 4.20b). As the barrier separating the two bands is very narrow, the electrons tunnel through it under the effect of a small reverse bias. The electrons from the valence band on the *p*-side reach the conduction band on *n*-side, thus causing a sudden large reverse current from *n* to *p* side. This is known as ***Zener effect.*** Zener breakdown takes place usually at lower reverse voltages of the order of 4 volts or less.

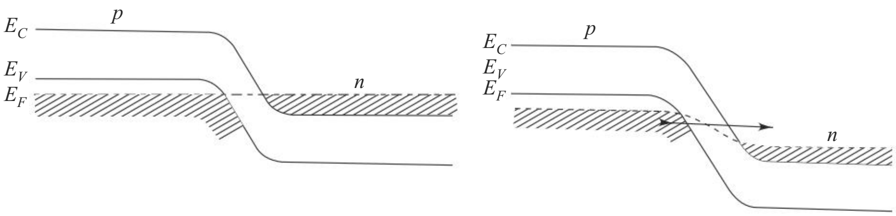


Fig. 4.20 *Energy band diagram of (a) heavily doped p-n-junction (Zener diode) (b) reverse biased junction and Zener effect*

Applications

Zener diodes are widely used in electronic circuits. Some of the important uses are the following:

1. It is used as a voltage reference in emitter − follower type voltage regulator.
2. It is used as a regulated power supply.
3. It is used in the protection circuits for MOSFET.
4. It is also used in the clipping circuits.

Among these applications it is popularly used as a voltage regulator.

4.3.5 Zener Diode as a Voltage Regulator

With the help of rectifiers and appropriate filters, we can get almost pure, i.e., ripple free dc voltage. Such dc voltage power supply finds applications in chargers, battery eliminators, electronic equipments, etc. However, the problem with this type of dc supply is that its output voltage will not remain constant in the event of fluctuations in ac input voltage or changes in the load current. This type of power supply is called an unregulated power supply. The expensive electronic instruments cannot use such a dc power supply. They need a dc power supply whose output voltage remains constant in spite of variations input voltage or load current.

The voltage regulator is an electronic circuit which supplies the constant output voltage under all the operating circumstances. Generally, the output voltage changes if there is a change in input voltage or change in load current or if the ambient temperature is changed. To provide the constant output voltage, the voltage regulator is kept in the last block of the dc regulated power supply. Thus, the power supply which provides a constant output voltage irrespective of any conditions is called a regulated power supply.

The block diagram of a regulated power supply is as shown in Fig. 4.21

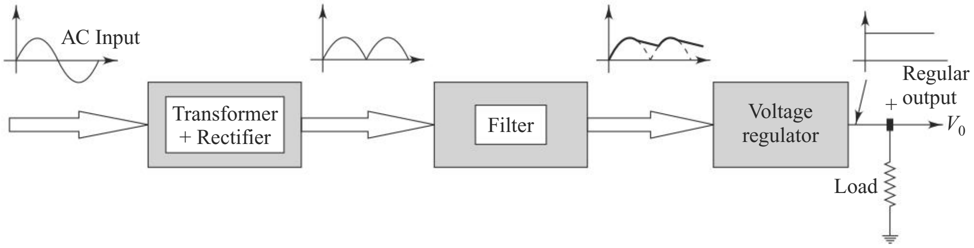


Fig. 4.21 *Block diagram of dc regulated power supply*

Working of the Circuit

The output of the unregulated power supply is connected at the input of the voltage regulator circuit. The voltage regulator is a specially designed circuit to keep the output voltage constant. The voltage regulator circuit can be either a series voltage regulator or a shunt voltage regulator. These regulator circuits are discussed in detail in the following paragraphs. Even though the output voltage V_0 of the regulated power supply remains almost constant, practically it does not remain exactly constant. It changes slightly due to changes in certain parameters.

Voltage Regulator Circuit

Voltage regulation is the ability of a circuit to maintain a constant output voltage even when either input voltage or load current varies. A Zener diode can serve as a voltage regulator. It can be used to provide a constant voltage from a source whose voltage may vary over a considerable range. Figure 4.22 shows the use of a zener diode as a voltage regulator. An unregulated power supply provides the input voltage V, R is the current limiting resistor and R_L is the load resistor. The Zener diode of a zener voltage V_Z is reverse connected across input supply voltage V_i . The load resistance R_L across which constant output is desired is connected across the diode. The role of series resistance R is to absorb the output voltage fluctuations so as to maintain constant voltage across the load. The Zener will maintain a constant voltage V_Z across the load as long as the input voltage does not fall below V_Z. A zener diode has maximum current rating equal to I_{max}. The Zener current should not exceed this limit. Minimum current may go to zero. The total current I passing through R equals the sum of diode current and load current, i.e. , $I = I_Z + I_L$

This circuit makes use of the fact that under reverse-bias breakdown voltage, the voltage across the Zener diode remains constant even if larger current is drawn. Since the load resistance R_L is across the Zener diode, the voltage across the load resistance does not vary even though current through the load changes. Hence, the voltage across the load is regulated against the variations in the load current.

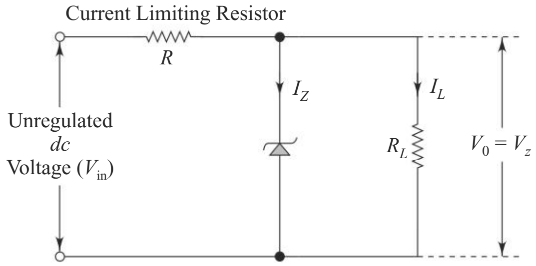


Fig. 4.22 *Zener diode is used as a voltage regulator*

The working of this circuit can be explained in two different cases as below;

(a) Regulating Action with a Varying Input Voltage (constant I_L)

Here, the load resistance R_L and load current I_L are constant and V_{in} is varying (Fig. 4.23).

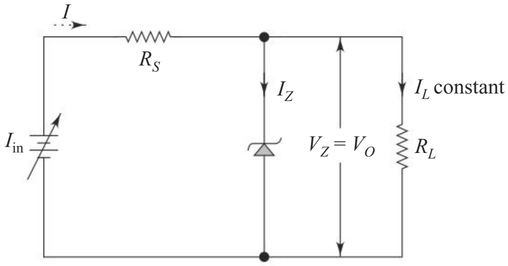


Fig. 4.23 *Regulating action with a varying input voltage*

As R_L is constant, I_L is also constant because $I_L = V_Z/R_L$. But supply current keeps changing due to change in V_{in} as,

$$I = \frac{V_{in} - V_Z}{R} \quad \text{and} \quad I = I_Z + I_L$$

If V_{in} is increased, then current I will increase. But as V_Z is constant and R_L is also constant, the load current I will remain constant. Naturally, the increase in current I will increase the Zener current I_Z. This can continue harmlessly as long as I is less than $I_{max.}$ If V_{in} is decreased, I will decrease causing I_Z to decrease as V_Z and I_L are constant. The load voltage will continue to be equal to V_Z as long as I_Z is higher than $I_{min.}$ Thus, the output voltage V_0 will remain constant as long as the Zener current is maintained between I_{min} and I_{max}. As soon as it goes outside these limits, the output voltage will not remain constant.

(b) Regulating Action with a Varying Load (V_{in} constant)

Here, the input voltage is constant and load resistor R_L is variable. If R_L increases then I_L will decrease (Fig. 4.24).

But the total current I is constant as

$$I = \frac{V_{in} - V_Z}{R}$$

Therefore with decrease in I_L the Zener current I_Z will increase. This can continue without damaging the Zener diode, as long as I_Z is less than I_{max} and the output voltage will remain constant. If R_L is reduced, I_L increases. But as I is constant, the Zener current will decrease. The output voltage will remain constant as long as I_Z is higher than I_{min}.

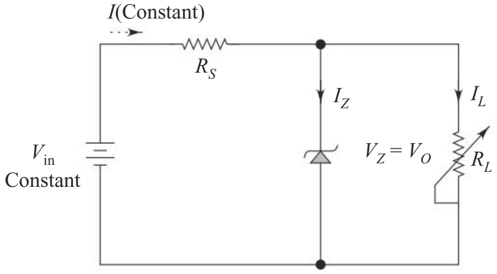


Fig. 4.24 *Regulating action with a varying load*

Thus, the output voltage remains constant as long as the Zener current is between $I_{\min}$ and the $I_{\max}$.

4.3.6 Varactor Diode or Varicaps

The depletion layer in *p-n* junction diode has a row of opposite charges on either side under equilibrium condition. This works like a capacitance. We can forward bias or reverse bias this device and accordingly, we can have diffusion capacitance and transition capacitance. Both these capacitances are important for many applications. Diffusion capacitance is significant while forward bias is applied whereas, transition capacitance is significant while reverse bias is applied.

Consider a reverse biased *p-n* junction as shown in Fig. 4.25. When reverse bias is applied, the majority carriers move away from the junction and depletion layer widens. The movement of majority carriers brings change in charge at the junction due to applied reverse voltage. This change in charge due to applied reverse voltage is called transition capacitance, barrier capacitance, junction capacitance or space charge capacitance. It is denoted by $\boldsymbol{C_t}$.

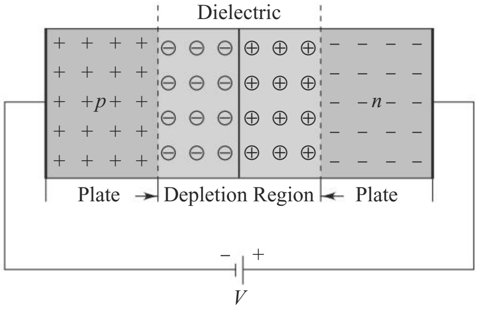


Fig. 4.25 *Reverse biased p-n junction acts as varactor diode*

$$\text{The transition capacitance} = \frac{\text{Rate of change of charge at the junction}}{\text{Rata of change of applied reverse voltage}}$$

$$C_t = \frac{dQ}{dV}$$

p and *n* regions act as plates and depletion region act as dielectric area hence,

$$C_t = \frac{\varepsilon A}{d}$$

where, A = area of cross-section of junction

d = width of depletion region

ε = permittivity of semiconductor = $\varepsilon_0 \varepsilon_r$

But the relationship between width of depletion region and barrier potential is given by

$$V_B = \frac{QNd^2}{2\varepsilon}$$

Thus, transition capacitance $\boldsymbol{C_t}$ is inversely proportional to width of depletion region. Therefore, on increasing reverse voltage, transition capacitance decreases. Here, the depletion region acts as a capacitor so the diode will perform as a variable capacitor that changes its capacitance with the applied bias voltage. Thus, a *varactor diode is a junction diode in which the junction capacitance of the diode is designed to vary with the reverse bias voltage*.

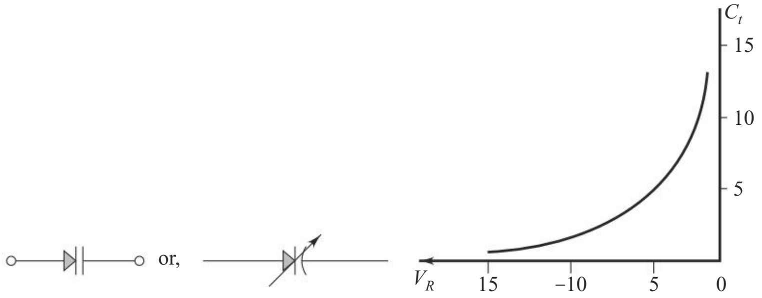


Fig. 4.26 *(a) circuit symbol of varactor diode (b) V-C_t characteristic of a typical diode*

Figure 4.26 (a) shows the circuit symbol of a varactor diode and Fig.4.26 (b) depicts V-C_t characteristic of a typical diode. The capacitance of a typical varactor can vary from 2 to 50 picofarads for a bias variation of just 2 volts. If doping level is increased, the width of the depletion layer reduces. Therefore, C_t increases with increase in doping level.

Varactor diodes are also known as ***varicaps***. A varactor diode shows normal volt–ampere characteristics as the junction diode in both forward and reverse bias conditions. The semiconductor material used for the fabrication of varactor diodes should have high carrier mobility, low dielectric constant, large band gap, low ionization potential, and high thermal conductivity. The material having most of the required properties is GaAs and therefore, GaAs is used for the fabrication of varactor diodes. They are widely used in different devices like TV receivers, F M receivers, tuned devices, communication equipments, time-varying signal applications, etc.

4.3.7 Light Emitting Diode (LED)

Display devices and light sources are ever-developing branches. CRT, LCD, Plasma, LED, etc., are some of the popular display devices. Incandescent lamps fluorescent tubelights, CFL, arc lamps, neon lamps, xenon lamps, halogen lamps, LED, LASER, etc., are light sources used in different applications. They all are classified as luminaries as they work on the principles of luminescence. LED finds its place in both the areas. It is a specially designed semiconductor diode. It is widely used due to advantages such as smaller size, less power requirement, good spectral response, etc. III-V compound type semiconductors are used instead of elemental semiconductors such as Si or Ge due to narrow band gap requirement to emit light. Different compounds of this composition are used to emit a desired wavelength. It converts electrical energy into light energy. Emission of light in LED is due to injection electroluminescence phenomenon.

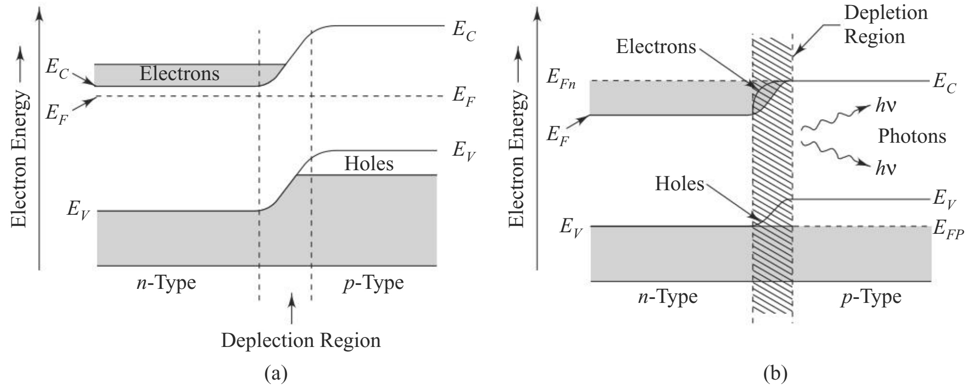
Principle

When the LED is forward biased, the electrons in the n-region will cross the junction and recombine with the holes in the p-type material producing light in the visible or infra-red region. The wavelength of light is given by,

$$\lambda = \frac{hc}{E_g} = \frac{1.24}{E_g}\,\mu\text{m/eV}$$

This effect is known as injection electroluminescence. A significant light output is obtained only when there is a large number of electron-hole recombination per second. To ensure this, the p and n regions are heavily doped in this device.

Theory The energy-band diagram of a heavily doped p-n junction is shown in **Fig**.4.27.



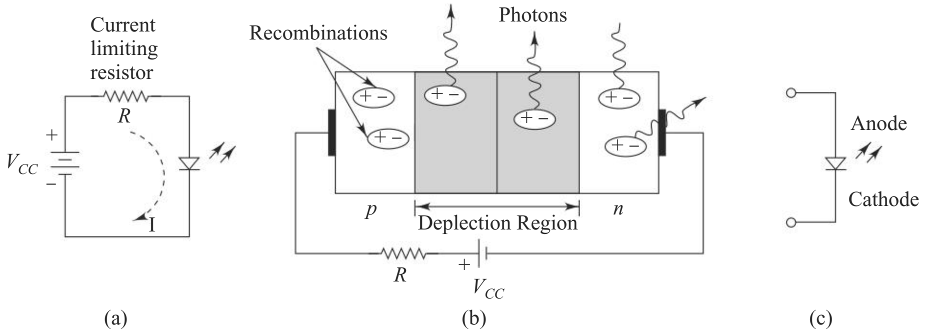
(a) (b)

Fig. 4.27 *Energy-band diagram of LED (a) Unbiased (b) forward biased conditions*

There are a majority of electrons in the conduction band of n-region and a majority of holes in the valence band of p-region (Fig. 4.27a). When such a device is forward biased, the electrons move towards the depletion region and occupy energy levels in the conduction band. Similarly, holes moves towards the depletion region and occupy energy levels in the valence band. The electrons in the conduction band are directly above the holes at the edge of the valence band (Fig. 4.27b). The situation is highly conducive for direct recombination of electrons and holes. When an electron from the conduction band jumps into the hole of the valence band, recombination occurs and the excess energy is emitted in the form of a light photon.

The colour of the emitted light depends on the type of material used.

	Material used	Colour of the emitted light
1.	Gallium Arsenide, GaAs	Infrared
2.	Gallium Arsenide-phosphide, GaAsP	Red or Yellow
3.	Gallium Phosphide, GaP	Red or Green



(a) (b) (c)

Fig. 4.28 *(a) LED biasing (b) Working principle of LED (c) Circuit symbol of LED*

LED biasing, working principle and circuit symbol of the device are shown in Fig. 4.28 (a), (b) and (c) respectively. To emit light, LED has to be in forward bias. The forward voltage across an LED is considerably greater than an ordinary diode. Typically, the maximum forward voltage for LED is between 1.2 V and 3.2 V depending on the device. The LED emits light in response to a sufficient forward current.

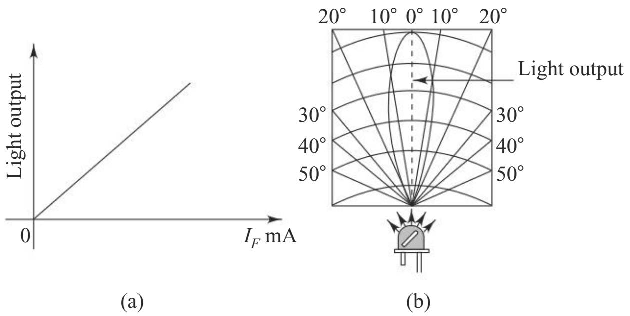


Fig. 4.29 *(a) Out put characteristics and (b) Radiation pattern of LED*

The graph of Fig. 4.29 (a) is the output characteristics of an LED. Light output of this device is directly proportional to the forward LED current. However, it is not possible to increase I_F indefinitely. Typically, the maximum value of I_F is 80 mA. Directionality of the emitted light is a very important parameter in many applications. To know the directionality of the emitted light, radiation pattern of a graph of light output is plotted against angle as shown in Fig. 4.29 (b). Radiation pattern reveals that most of the radiated energy is confined within 20° of the direction of maximum light. Thus, the LED is a highly directional source because it radiates only in a particular direction.

Construction

The structure of a surface emitting LED is shown in Fig**.** 4.30(a). These LEDs emit light in a direction perpendicular to the *p*-*n* junction plane. An *n*-type layer is grown on a substrate and a *p*-type layer is grown on the *n*-layer by the process of diffusion. The *p*-layer is made very thin to prevent loss of photons due to absorption in the layer. Metal connections are made at the edges of the *p*-layer to allow more central surface for the light to escape.

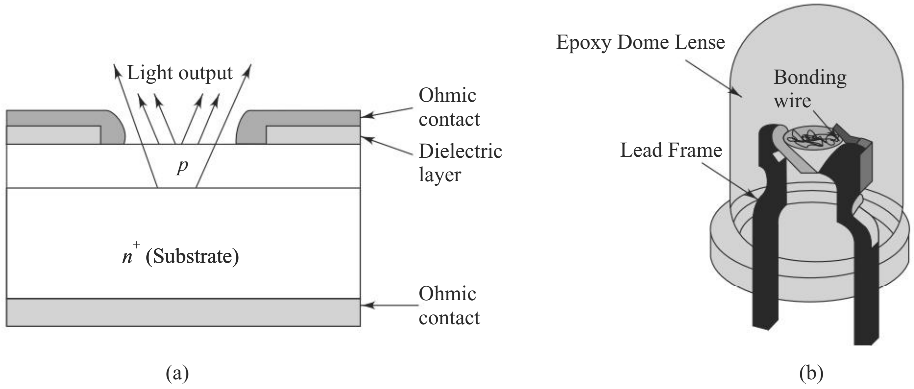


Fig. 4.30 *(a) structure and (b) construction of LED*

A metal film is deposited at the bottom of the substrate for reflecting as much light as possible towards the surface of the device and also to provide electrode connection. The light generated at the junction may not emerge from the surface of the device as it is likely to suffer total internal reflection at the semiconductor – air boundary. Therefore, the device is encapsulated in a clear epoxy resin of suitable refractive index (Fig. 4.30b). Such a structure is called as a cup type construction.

Merits of LED

1. LEDs are of small size and light in weight. Therefore, it is possible to pack a large number of LEDs in a small space while manufacturing a display.
2. They have longer life.
3. They are available in different spectral colours.
4. They are suitable at high operating speeds as they take less than 1 μs to turn on or off.
5. LEDs can be easily interfaced with the other electronic circuits.

Demerits of LED

1. Luminous efficiency of LED is low.
2. Output power is affected by changes in temperature.
3. Overcurrent can damage it easily.
4. They need larger power for their operation.

Applications

LEDs are commonly used in many applications. Discrete LEDs are used *as indicators* and *as light sources* in fibre-optic communications. A number of LEDs may be grouped *to form a display*. The LEDs may be arranged in the form of a seven-segment display where by energizing a proper combination of segments, the decimal numbers 0 to 9 may be displayed. Or they may be arranged in the form of a 5 × 7 matrix which may be used to generate a decimal number or alphabetical character. They are also used as opto-couplers. They are popularly used in the infrared remote controls.

SOLAR CELL AND ITS APPLICATIONS

4.4

One can produce electricity on the hillside or at remote places where a power grid is unavailable and we can run the electrical gadgets by getting energy as light from the sun. In this chapter, the basic principles of solar energy applications are discussed.

4.4.1 Introduction to Photovoltaic Cell/Solar Cell and its Principle

Photovoltaic effect is the phenomenon in which the incidence of light or other electromagnetic radiation upon a junction made up of two dissimilar materials, such as a metal and a semiconductor, induces the generation of an electromotive force.

The photovoltaic effect produces voltage (or a corresponding electric current) in a material upon exposure to electromagnetic radiation. Even though the photovoltaic effect is directly related to the photoelectric effect, the two processes are different and should be distinguished. In the photoelectric effect, electrons are ejected from a material's surface upon exposure to radiation of sufficient energy. On the other hand, the photovoltaic effect is different in that the generated electrons are transferred from one material to another resulting in the build up of a voltage between two electrodes.

In most photovoltaic applications, the radiation is sunlight; and for this reason, the devices making use of photovoltaic effect to convert solar energy into electrical energy are known as *solar cells.* In the case of a P-N junction solar cell, illumination of the material results in the creation of an electric current as excited electrons and the remaining holes are swept in different directions by the built-in electric field of the depletion region.

Therefore, photovoltaic is the direct conversion of light into electricity at the atomic level. The photoelectric effect was first noted by a French physicist, Edmund Bequerel, in 1839, who found that certain materials would produce small amounts of electric current when exposed to light. In 1905, Albert Einstein described the nature of light and the photoelectric effect on which photovoltaic technology is based, for which he later won a Nobel prize in Physics. The first photovoltaic module was built by Bell Laboratories in 1954. It was billed as a solar battery and was mostly just a curiosity as it was too expensive to gain widespread use. In the 1960s, the space industry began to make the first serious use of the technology to provide power aboard spacecraft. Through the space programmes, the technology advanced, its reliability was established, and the cost began to decline. During the energy crisis in the 1970s, photovoltaic technology gained recognition as a source of power for non-space applications.

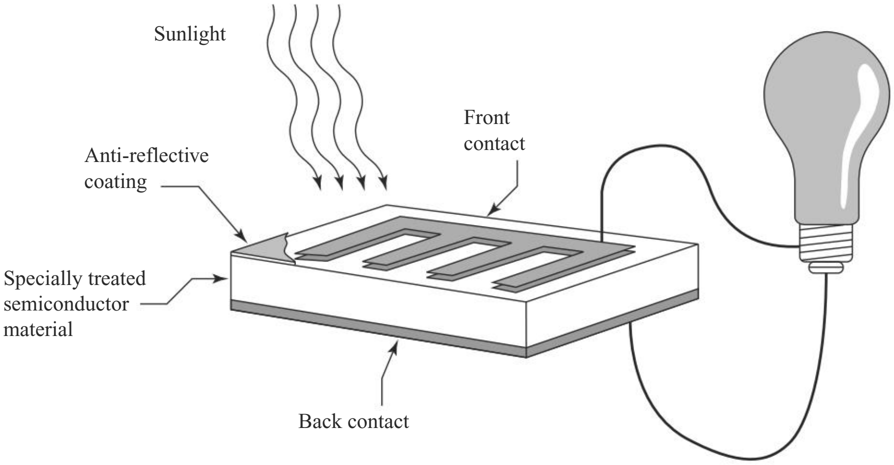


Fig. 4.31 *Schematic diagram for photovoltaic cell principle*

Figure 4.31 illustrates the operation of a basic photovoltaic cell, also called a solar cell. Photovoltaic cells, or solar cells, are made of the same kinds of semiconductor materials, such as silicon, used in the microelectronics industry. For solar cells, a thin semiconductor wafer is specially treated to form an electric field, positive on one side and negative on the other. When light energy strikes the solar cell, electrons are knocked loose from the atoms in the semiconductor material. If electrical conductors are attached to the positive and negative sides, forming an electrical circuit, the electrons can be captured in the form of an electric current—that is, electricity. This electricity can then be used to power a load, such as a light or a tool.

A number of solar cells electrically connected to each other and mounted in a support structure or frame is called a *photovoltaic module.* Modules are designed to supply electricity at a certain voltage, such as a common 12-volt system. The current produced is directly dependent on how much light strikes the module.

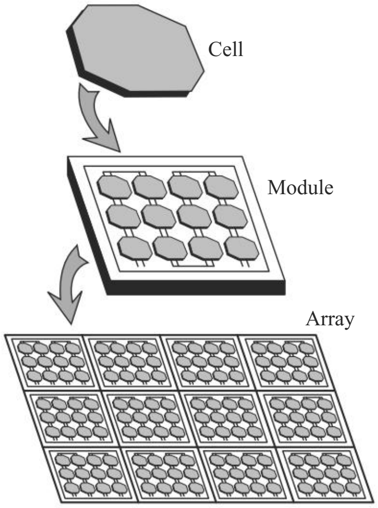


Fig. 4.32 *Photovoltaic module*

Multiple modules can be wired together to form an array. In general, the larger the area of a module or array, the more the electricity that will be produced. Photovoltaic modules and arrays produce direct-current (dc) electricity. They can be connected in both series and parallel electrical arrangements to produce any required voltage and current combination.

Today's most common photovoltaic (PV) devices use a single junction, or interface, to create an electric field within a semiconductor such as a PV cell. In a single-junction PV cell, only photons whose energy is equal to or greater than the band gap of the cell material can free an electron for an electric circuit. In other words, the photovoltaic response of single-junction cells is limited to the portion of the sun's spectrum whose energy is above the band gap of the absorbing material, and lower-energy photons are not used.

One way to get around this limitation is to use two (or more) different cells, with more than one band gap and more than one junction, to generate a voltage. These are referred to as 'multi-junction' cells (also called 'cascade' or 'tandem' cells). Multi-junction devices can achieve higher total conversion efficiency because they can convert more of the energy spectrum of light to electricity.

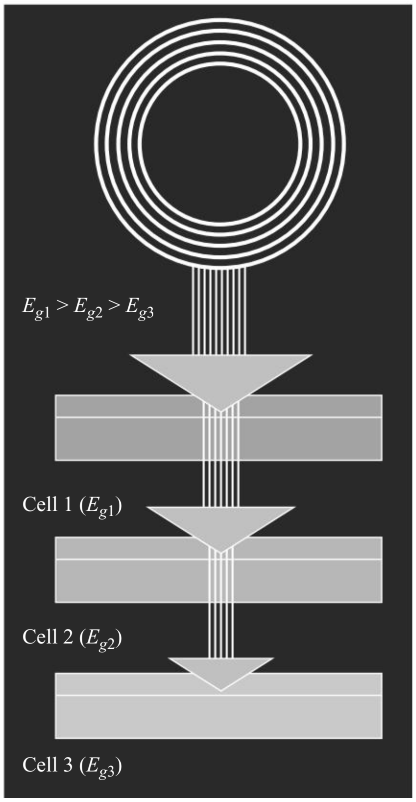


Fig. 4.33 *Typical multi-junction cell showing the energy band gap E_g for various cells*

As shown below, a multi-junction device is a stack of individual single-junction cells in descending order of band gap (E_g). The top cell captures the high-energy photons and passes the rest of the photons on to be absorbed by lower-band-gap cells.

Much of today's research in multi-junction cells focuses on gallium arsenide as one (or all) of the component cells. Such cells have reached efficiencies of around 35% under concentrated sunlight. Other materials studied for multi-junction devices have been amorphous silicon and copper indium diselenide.

As an example, the multi-junction device below uses a top cell of gallium indium phosphide, 'a tunnel junction,' to aid the flow of electrons between the cells, and a bottom cell of gallium arsenide.

The most commonly used solar technologies for homes and businesses are solar water heating, passive solar design for space heating and cooling, and solar photovoltaics for electricity.

Businesses and industry also use these technologies to diversify their energy sources, improve efficiency, and save money. Solar photovoltaic and concentrating solar power technologies are also being used by developers and utilities to produce electricity on a massive scale to power cities and small towns.

Although solar cells had many advantages, they have also negative effects and disadvantages. One of the major drawbacks in this kind of technology is the absolute volume of land that is needed to put up the solar power plant.

Most of the time, a large area is needed, and a typical value for a plant calls for 9.5 square miles of land area for its house operation.

Depending upon the cost of the local electric rates and installation, the pay off can be 14 to 20 years. Although, solar modules have 20 years warranty, the investment will be lost if one decided to move.

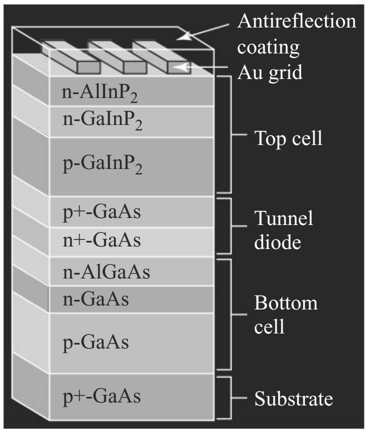


Fig. 4.34 *Typical multi-junction device*

Moreover, solar electricity is more costly compared to the current rates. As soon as the PV system is installed for use, it will generate electricity with similar cost as the current technology until about 12 years when the inverter demands replacement.

Recent utility rates have been made in greater amount every year and with the larger pressure on the reduction of carbon, the rate will sharply increase. In the long run, this increase can easily compensate the installation cost but the year it takes for the pay off is too long for most of the people. Solar electricity will not be usable at night and is less obtainable in cloudy weather conditions from the established silicon-based technologies.

Hence, there must be a complimentary power system or storage like germanium in the thin-film solar cells of amorphous silicon−germanium composition which supplies residual power, bringing forth the capacity at night-time, and utilising the background infrared radiation.

Luckily, most of the consumption is at daytime, so there is no need of storing solar energy as long as it caters to the total consumption.

During the conversion of solar cells from dc to ac with the aid of a grid tie inverter, an energy loss of about 4 to 12% occurs.

4.4.2 Photovoltaic Module Characteristics

Voltage generated by a photovoltaic module or panel depends on the material from which its cells are made and the manner of their processing as well as any interconnections of the cells made. As current drawn from the PV module increases, initially its voltage falls gradually and after it reaches its *knee point,* it falls rapidly (see the *V-I* characteristics in the accompanying figure). The knee point of the *V-I* characteristics of the photoelectric module defines the limiting conditions of its usage. This characteristic of the device is a limitation as well as an inherent safety feature—it is supplying its maximum power when you reach the "knee point" and beyond the knee point it is protecting itself. When using a PV panel or any such device, we must ensure that the peak demand

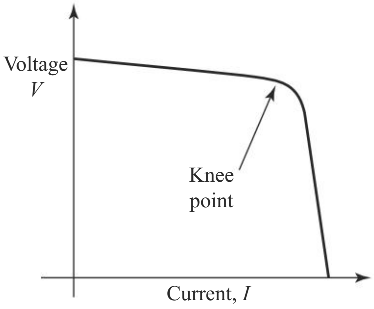


Fig. 4.35 *V-I characteristics of a photoelectric panel*

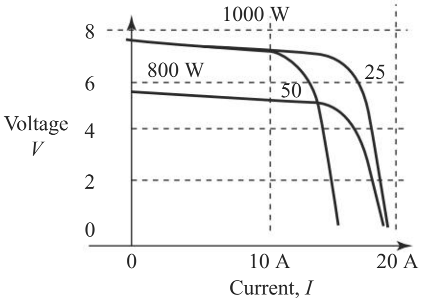


Fig. 4.36 *Effect of temperature on characteristics of solar panel*

does not exceed the maximum power the device is capable of delivering.

It is obvious that the capacity of a photovoltaic module to deliver power is dependent on the intensity of the light to which it is exposed and the temperature at which it is operating. As shown in the accompanying diagram, a photovoltaic module rated to deliver 1000 watts of power will lose its capacity and would be capable of providing only, say, 800 W if the light intensity to which it is exposed falls. Its peak power capacity (at the knee point) also falls. The maximum current it can supply and the current at which it supplies the peak power remain unaltered.

The characteristics of a photovoltaic module are dependent also on the temperature at which it is operating. As shown in the accompanying diagram, the peak power and current capacity of a solar-power generator fall as its operating temperature rises from 25°C to 50°C.

These characteristics are important to keep in mind in using these devices. Solar panels or arrays are mounted on rooftops of buildings or in open areas in such a way that they are able to get maximum light intensity by facing the sun. The solar panels or arrays are mounted in racks which can be tilted to face the sun.

4.4.3 Solar Cell

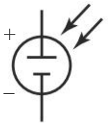
As discussed above, the solar cell (also known as photovoltaic cell) generates electric potential when it is illuminated by optical irradiation. It is used to generate electricity from sunlight; therefore it is known as solar cell. Solar cells are useful for both space and terrestrial applications, and are the source of long-duration power supply for satellites. They are an important alternative terrestrial energy source because they can convert sunlight directly to electricity with good conversion efficiency, can provide nearly permanent power at low operating cost, and are virtually nonpolluting. They are also used in watches, irrigation systems, calculators, etc.

Silicon-based solar cells are normally in use. Some of the materials used in solar-cell fabrication are single crystal Si, poly-silicon, CdTe and GaAs. Popular geometries of solar cells are flat-planed systems, thin-film devices, or concentration systems.

Sunlight is the main source of radiation energy for solar cells. We know that the intensity of solar radiation in free space at the average distance of the earth from the sun is 1367 W/m^2, which is known as the *solar constant.* But the sunlight received on the earth is less due to various absorption mechanisms in the atmosphere of the earth.

The electromagnetic radiation of wavelength range of 200–3000 nm from the sun is the main energy output used by solar cells. This energy decreases as it reaches the earth's surface depending on angle of incidence and atmospheric conditions.

Symbol of Solar Cell



Construction of Solar Cell The solar cell was invented in 1954 by Pearson. Silicon P-N junction diodes of special construction make a solar cell. The structure of a solar cell is shown in Fig. 4.37.

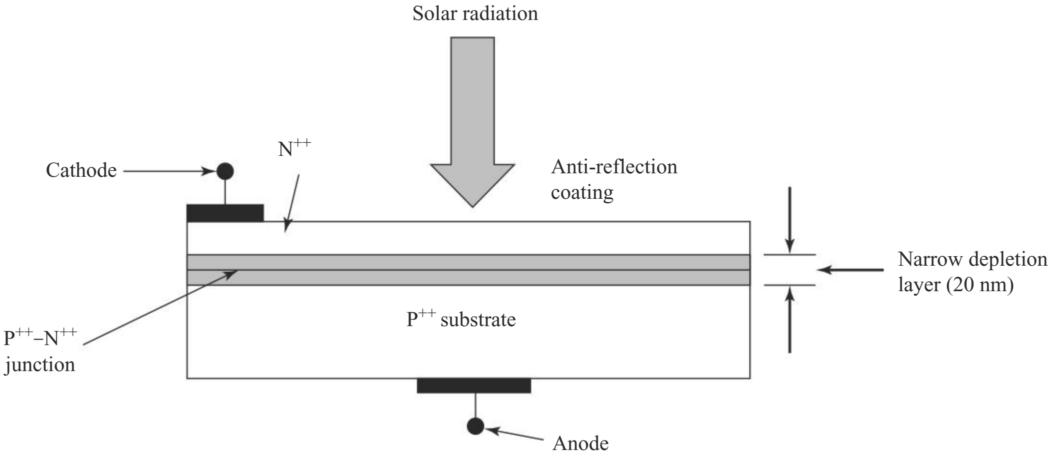


Fig. 4.37 *Side-view structure of a solar cell*

The solar cell consists of a heavily doped P-N (i.e., P^{++}-N^{++}) junction as shown in Fig. 4.37. The top layer forming the N^{++} region is usually made very thin to allow solar radiation to fall into the junction. The resulting junction will be extremely narrow (typically around 20 nm) because the doping level is extremely high, and may reach a degenerate doping level (about 10^{20} impurity atoms per cubic centimetre). Solar cells are designed to have large surface area to allow a large amount of light input into the device. Usually, the anode connection will be made from the bottom (P^{++} substrate) of the device and the cathode from the top (N^{++} layer). An anti-reflection coating on top of the top N^{++} layer is made to prevent light losses due to reflection from the surface of the solar cell. This is very essential because Si is a highly reflective material. Each solar cell typically produces an output of about 0.6 V. So, the total output voltage can be increased by connecting them in series, and to increase the current, the total resistance should be reduced which can be obtained by connecting several series of solar cells in parallel combination to get higher output power. In fact, they are available in the form of solar panels, and contain several solar cells in series and parallel combination. This is discussed now.

4.4.4 Theory of the Solar Cell

Looking at the structural details, the solar cell is a P-N junction diode with a window on the top for allowing optical irradiation. So the solar cell is a photodiode, but photodiodes just detect light and do not generate voltage. Then how can we claim that the same photodiode can generate electric voltage? To answer this question, the first feature is that the structure and doping level of this diode is quite different from those of other diodes. Let us now understand how these differences can convert the diode from a detector to a generator.

Explanation of the Photovoltaic Effect

We know that every P-N junction diode has a depletion-layer potential existing across its junction. For the silicon diode, it is about 0.6 V. But we cannot measure it directly. Why? To get an answer let us study the voltage drops in a P-N diode. If the voltage exists across the junction, we will get it as an output voltage across the cathode–anode terminals unless it is cancelled by another equal and opposite voltage. But we do not get this voltage across the anode–cathode terminals. This fact clearly indicates that there must be an equal and opposite voltage inside the diode that cancels the junction potential. On investigating further, we can see that such a potential does exist in every P-N diode.

Let us consider the junction potential (also known as contact potential) produced as the result of a junction formed by two different materials (i.e., the P and the N-type silicon- semiconductors). Each semiconductor region contributes an equal value (i.e., 0.3 V each) to the junction potential that gives the total junction potential to 0.6 V. To neutralise this potential, two more contact potentials exist. One is the junction formed as the anode−metal contact (i.e., semiconducting P-region with aluminum). The second junction formed as the cathode−metal contact (the N-region with aluminum). From the theory of Schottky-barrier junction (i.e., potential barrier between metal and semiconductor junction), we know that the junction potential between a metal and a silicon region is about 0.3 V. Therefore, the sum of these two contact potentials is 0.6 V, just enough to cancel the junction potential. Due to this cancellation of the junction potential, we are unable to measure it across the anode−cathode terminals.

Let us resolve these problems in the solar cell so that it will generate a voltage when illuminated by light.

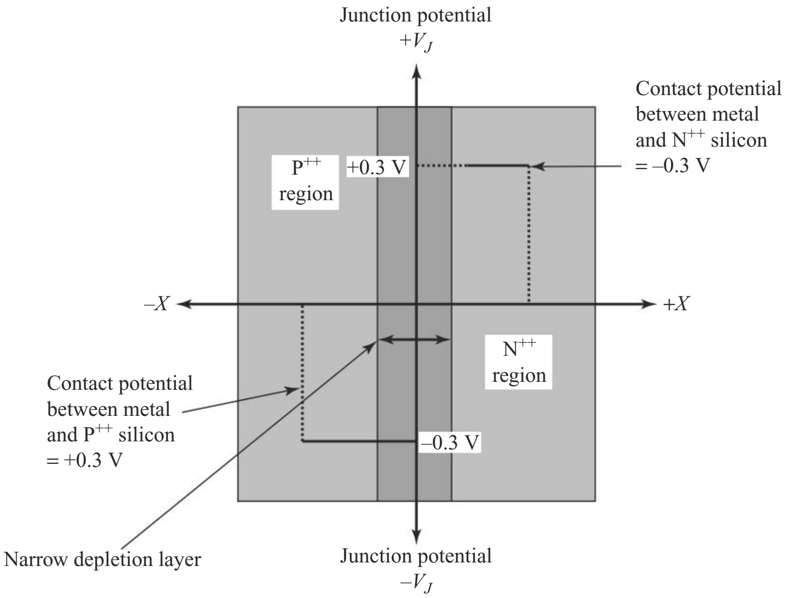


Fig. 4.38 *Contact potentials in a P-N junction*

Figure 4.38 depicts the situation described as above. A solar cell is a heavily doped (P^{++}-N^{++}) junction diode with a very narrow potential barrier. This narrow junction potential (depletion layer) is represented by the dark shading in Fig. 4.38. The *y*-axis in the figure represents potentials at various junctions (i.e., semiconductor− semiconductor and metal−semiconductor junction), and the *x*-axis represents the distance from the junction to the anode and cathode terminals. The figure displays +0.3 V as the portion of the junction potential in the N^{++} region, and −0.3 V as the portion of the junction potential in the P^{++} region. As explained earlier, these all-semiconductor contact potentials are counterbalanced by the metal−semiconductor contact potentials existing between the cathode-N^{++}, and the anode-P^{++} junctions, respectively. In Fig. 8.8, these contact potentials are shown by dotted lines. As a result of these cancellations, the net voltage across the anode and cathode terminals is zero.

Now, let the junction be illuminated by solar radiation. Due to solar radiation, we know that Electron–Hole Pairs (EHPs) are generated near the P-N junction. These optically generated EHPs tend to neutralise the immobile ions existing at the junction. Thus, in the P-region the immobile acceptor ions (–ve ions) are neutralised by the holes generated at the junction. Similarly, in the N-region the immobile +ve donor ions are neutralised by electrons generated at the junction. These neutralisations occur because of the slimness of the P^{++}-N^{++} junction.

Now, we are in a state, where the P^{++}-N^{++} junction potential is getting reduced or neutralised by the EHPs generated due to the solar radiation. The decrease in the junction potential depends on the number of EHPs generated. Even though a neutralisation of the semiconductor–semiconductor junction potential occurs due to solar-generated EHPs, such neutralisation does not occur at the metal-semiconductor junctions because in metal regions EHPs are not generated due to optical irradiation. Finally, the P-N junction potential gets reduced due to optical irradiation, but such a reduction does not occur in the terminal-contact potentials. This situation is shown in Fig. 4.39.

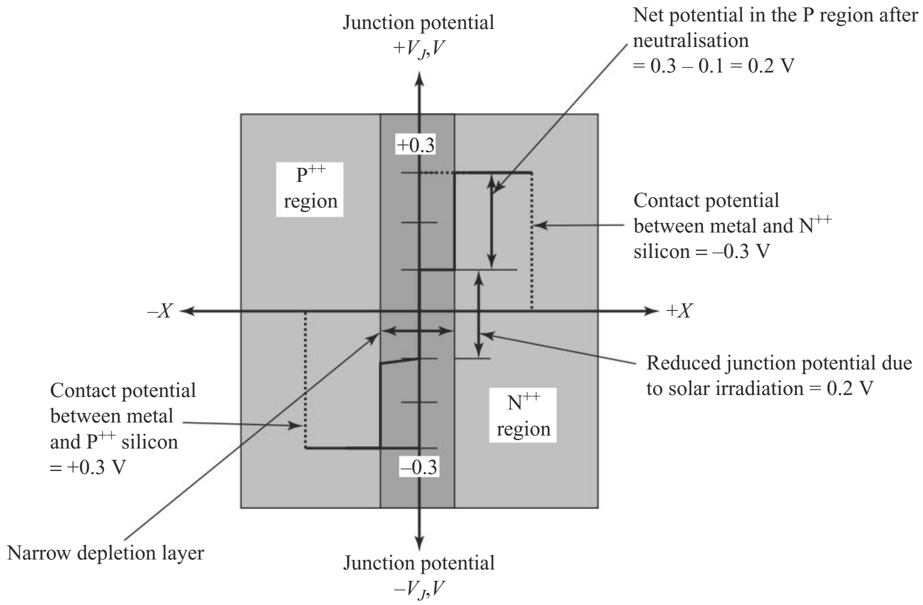


Fig. 4.39 *Solar-cell potential under radiation*

Figure 4.39 depicts a net potential produced across the anode–cathode terminals of the heavily doped P^{++}-N^{++} junction due to optical irradiation of the junction. It should be noted that the value of this potential is dependent on the amount of light falling onto the junction. The maximum value of this solar potential cannot overcome the normal value (i.e., 0.6 V) of junction potential of the silicon diode, because it is this potential that is neutralised by the optically generated EHPs. Therefore, we conclude that solar cells generate a net potential of about 0.6 V when irradiated by solar radiation.

In Fig. 4.39, the thick black horizontal lines at ± 0.1 V correspond to the reduced potentials at the P^{++}-N^{++} junction due to neutralisation by EHPs. The terminal potentials (at anode and cathode) are represented, as in Fig. 4.38, by dotted lines. The differences between these potentials and the junction potentials are shown by thick black vertical lines in Fig. 4.39. Note that a net potential difference of (0.2 V + 0.2 V = 0.4 V) 0.4 V is obtained in this case. One can see that this potential is positive (+ve) in the P side and negative (–ve) in the N side. This is shown by the thick dotted line in Fig. 4.40. Note that the maximum value of this potential can only be less than the junction potential of the silicon P-N junction (i.e., < 0.6 V). The solar potential cannot be exactly equal to the junction potential since practically total neutralisation of the immobile charge carriers in the junction region by the optically generated EHPs is impossible. Even if a large number of EHPs are generated by light, their number will still be less than the number of immobile ions generated due to heavy doping. As a result, some ions will remain at the junction, which are not neutralised.

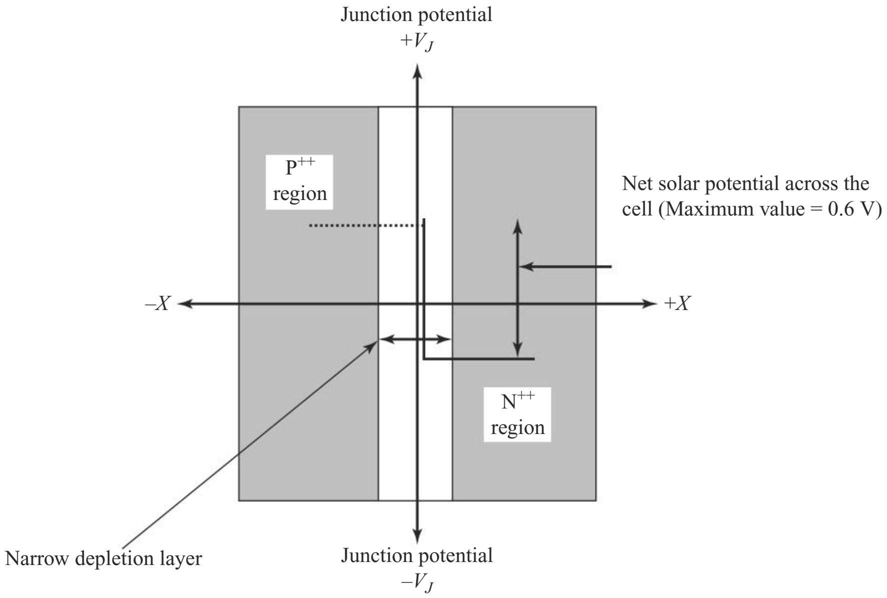


Fig. 4.40 *Net solar-cell potential under radiation*

4.4.5 Series and Parallel Operation

To get the desired voltage, several solar cells are connected in series, and connected in parallel to get the required current. Figure 4.41 shows the series–parallel connection of several solar cells. Three groups of solar cells are connected in parallel, with each group containing four solar cells connected in series. The total output voltage of the system is 2.4 V, assuming maximum 0.6 V for each cell.

The total current capacity of the unit depends on the material and construction used. Solar cells are in the form of solar panels and grouped under nonconventional energy sources.

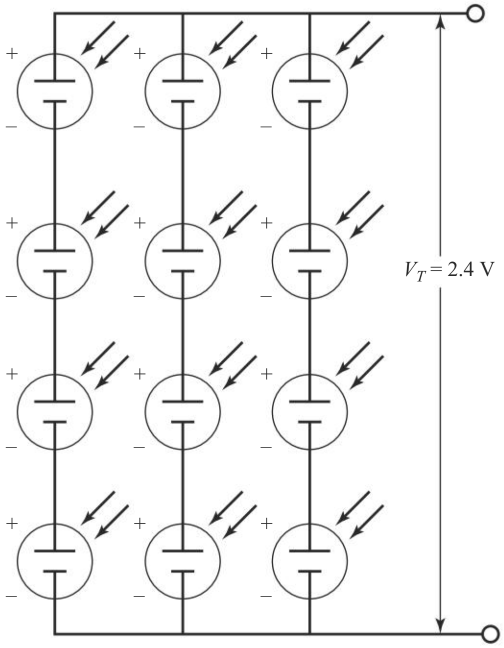


Fig. 4.41 *Solar cell connection in series and parallel*

4.4.6 P-N Junction Solar Cell

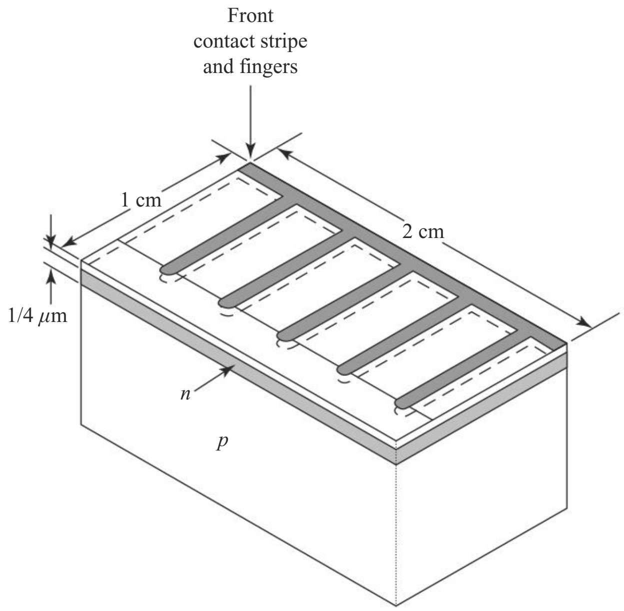
A schematic drawing of a P-N junction solar cell is shown in Fig. 4.42. It comprises of (i) a narrow P-N junction formed on the surface, (ii) a front ohmic contact stripe and fingers, (iii) a black ohmic contact that covers the entire back surface, and (iv) an antireflection coating on the front surface.

When the solar cell is exposed to the solar radiation, a photon that has energy less than the band gap E_g makes no contribution to the cell output. A photon having energy greater than E_g contributes an energy E_g to the cell output and an energy greater than E_g is wasted as heat.

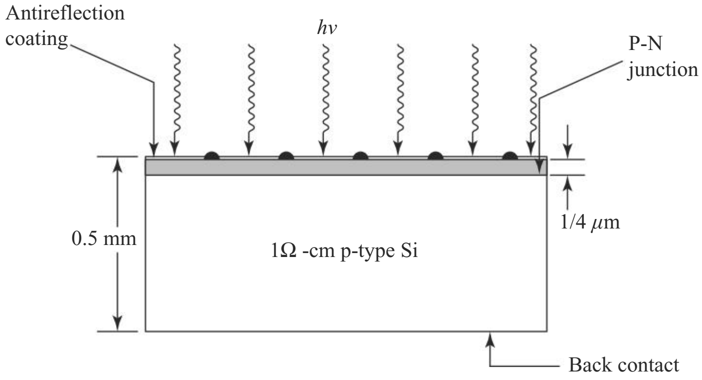
Consider a photon of light having wavelength λ incident on a semiconducting material like Si (P-N solar cell) having an energy band gap E_g. The energy of a photon is $E = hc/\lambda$, and it will be absorbed by the material if $E > E_g$, otherwise for $E < E_g$, they get transmitted without absorption. The absorbed energy of a photon excites valence electrons to the conduction band leaving holes in the valence band. The remaining $(E - E_g)$ energy will be lost as heat or phonons. Here, phonons are the energy quanta associated with lattice vibrations in solids.

The rate of carrier generation is a function of wavelength and intensity of incident radiation. Intensity of incident radiation corresponds to the number of photons. The number of photons corresponds to the number of carriers. The wavelength of a photon decides the energy of carriers. After a short time interval, the generated charge carriers recombine. The time rate of recombination r is given by the relation

$$r = \frac{n - n_0}{\tau} = \frac{p - p_0}{\tau} \text{ in } (\text{m}^{-3}\ \text{s}^{-1}) \tag{4.1}$$



(a) Top view of a silicon P-N junction solar cell



(b) Side view of a silicon P-N junction solar cell

Fig. 4.42 *Schematic representation of a silicon P-N junction solar cell*

where n_0 and p_0 are equilibrium concentration of electrons and holes whereas n and p refers to concentrations at any given instant of time. Now the fraction of minority carriers diffuse to the edges of the junction depletion region and are swept across the junction due to existing junction potential which has created an electric field there. So electrons are collected on the N-side and holes are collected on the P-side of a P-N solar cell.

This collection of electrons and holes on N- and P-sides of a solar cell produces light induced current (which is known as *photocurrent*) through a load resistance connected across the cell.

Let,

$$F_{ph} = \text{photon flux} = \frac{\text{number of photons incient}}{\text{area} \times \text{time}}$$

Assume 100% quantum efficiency, i.e., each absorbed photon generates one electron–hole pair. The pair generation rate G is given by

$$G = aF_{ph}(1 - R)e^{-ax} \tag{4.2}$$

where α is the absorption coefficient, in m^{-1}, R is the reflection coefficient, F_{ph} is the photon flux, and x is the distance where the pair-generation rate G is measured.

Under steady-state condition (i.e., rate of recombination and generation becomes equal), continuity equation for holes on N-side can be written as

$$\frac{1}{e}\frac{dJ_p}{dx} + \frac{p_e}{\tau} - G = 0 \tag{4.3}$$

where,

J_p = current density for holes = $eD_p\left(\frac{dp_e}{dx}\right)$

D_p = hole diffusion coefficient

p_e = hole concentration on N-side.

Therefore, Eq. (4.3) can be written as

$$\frac{1}{e}\frac{d\left[eD_p\left(\frac{dp_e}{dx}\right)\right]}{dx} + \frac{p_e}{\tau} - G = 0$$

$$D_p\frac{d^2p_e}{dx^2} + \frac{p_e}{\tau} - G = 0 \tag{4.4}$$

A similar way of expression for electrons on the P-side can be written as

$$D_e\frac{d^2e_p}{dx^2} + \frac{e_p}{t} - G = 0 \tag{4.5}$$

Solving these expressions and adding them gives the current passing through solar cell, which is given by

$$I = I_S(e^{v/vT} - 1) - I_L \tag{4.6}$$

where,

V is the bias voltage applied to collect charges generated by light,

I_S is the reverse saturation current,

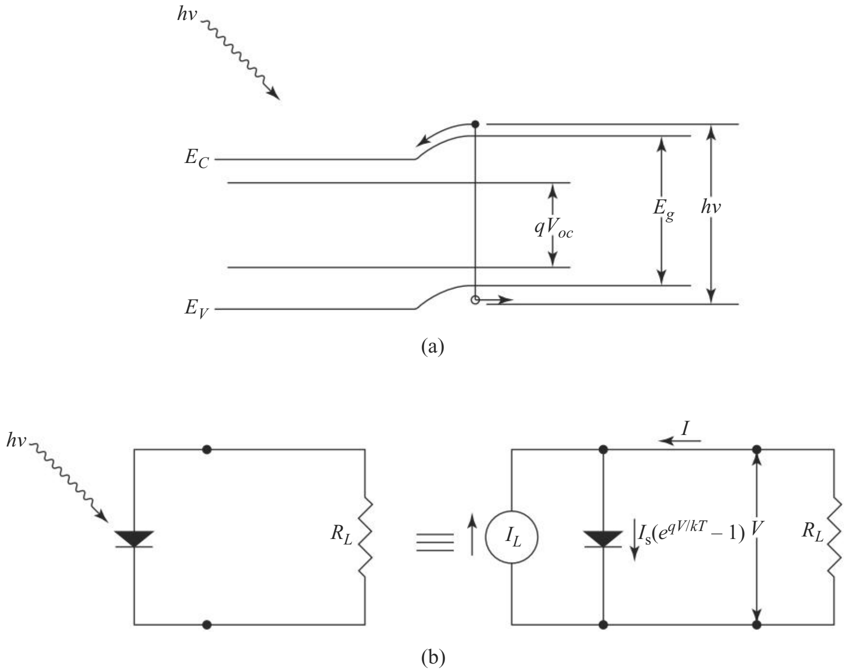
I_L is the photocurrent or light generated current passes through load and given by

$$I_L = eAG(L_p + L_n) \tag{4.7}$$

where, A is the cross-sectional area of the junction and L_P and L_N are diffusion lengths of the minority carriers on the P-side and N-side of the P-N solar cell respectively. The I-V characteristic curve is shown in Fig. 4.44. For different values of G, I-V characteristics are shown. The net current I is the difference between the injection current and light (photon) induced current.

Under the illuminated condition, the equivalent circuit of the solar cell is shown in Fig. 4.43(b). Photocurrent is the current flowing through the load R_L, and the shunt resistance R_{sh} can be attributed to the leakage currents. R_s, the series resistance, is due to the bulk resistance of the solar cell along with contact resistance.

To derive the conversion efficiency, we shall consider the energy band diagram of a P-N junction under solar radiation shown in Fig. 4.43(a).



(a)

(b)

Fig. 4.43 *(a) Energy band diagram of a P-N junction solar cell under solar irradiation; (b) Idealized equivalent circuit of a solar cell*

The equivalent circuit is shown in Fig. 4.43(b). As shown in the figure, a constant-current source is in parallel with the junction. The source I_L is as a result of the excitation of excess carriers by solar radiation, I_S is the diode saturation current, and R_L is the load resistance.

The ideal *I-V* characteristics of such a device are given by

$$I = I_S(e^{qV/kT} - 1) - I_L \tag{4.8}$$

And the saturation current density is given by the following relation.

$$J_S = \frac{I_S}{A} = qN_V N_C \left[\frac{\sqrt{D_p}}{N_D\sqrt{\tau_p}} + \frac{\sqrt{D_n}}{N_A\sqrt{\tau_n}} \right] e^{-E_g/kT} \tag{4.9}$$

In the above equation, A represents the area of a solar cell.

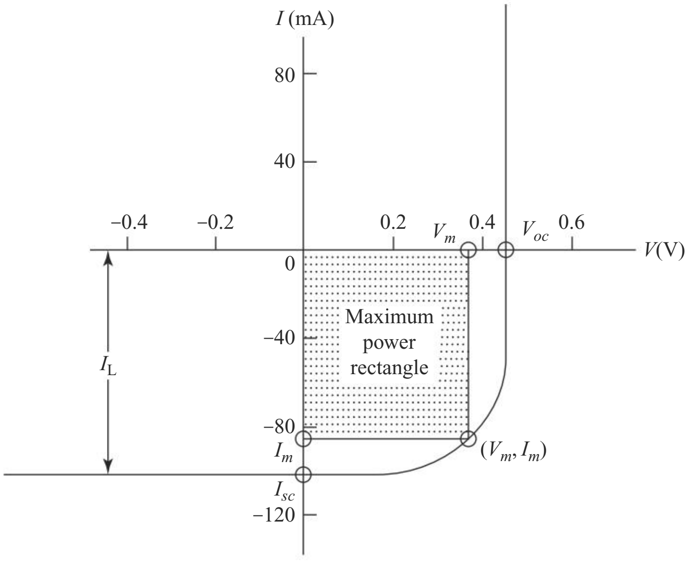
D_p, D_n represents the diffusion coefficient for holes and electrons.

N_D and N_A represent the donor and acceptor impurity densities respectively.

τ_p and τ_n represent the carrier lifetime for holes and electrons respectively.

N_C and N_V represent the effective density of states in the conduction band and valence band respectively.

A plot of Eq. (4.8) is given in Fig. 4.44(a). The curve passes through the fourth quadrant of the *I*-*V* characteristic curve. This indicates that the device is having negative conductance in the fourth quadrant. Therefore, if the operating point of the diode is selected in this quadrant, it could be used to deliver power to an external load. It acts as a converter of solar energy into the electrical energy. In the other three quadrants if the diode is operated, it will function as a sink of power or act like an ordinary resistor. The *I*-*V* curve is more generally represented by Fig. 4.44(b), which is an inversion of Fig. 4.44(a) about the voltage axis. By properly choosing a load (with power close to 80%), the product ($I_{sc}V_{oc}$) can be extracted, where I_{sc} is known as the *short-circuit current* equal to the photocurrent I_L, and V_{oc} is the open-circuit voltage of the cell. The shaded area in the figure represents the maximum-power rectangle. The quantities I_m and V_m correspond to the current and voltage respectively, for the maximum power output P_m ($= I_m \times V_m$) defined in Fig. 4.44(b).



(a)

Fig. 4.44 *(a) Current-voltage characteristics of a solar cell under illumination*

From Eq. (4.8), we obtain equations for the open-circuit voltage V_{oc} (and $I = 0$).

Taking log on both the sides and simplifying for V_{oc};

$$I = I_S(e^{qV/kT} - 1) - I_L$$

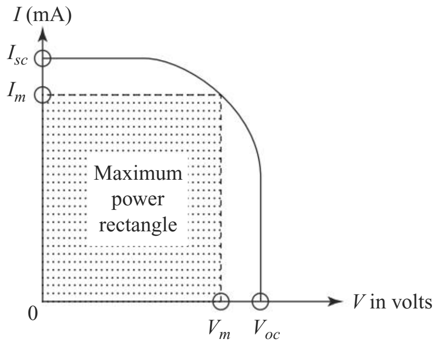
$$\Rightarrow 0 = I_S(e^{qV_{oc}/kT} - 1) - I_L$$

$$\Rightarrow I_S(e^{qV_{oc}/kT} - 1) = I_L$$

$$(e^{qV_{oc}/kT} - 1) = I_L/I_S$$

$$e^{qV_{oc}/kT} = (I_L/I_S) + 1$$

$$(qV_{oc}/kT) = \ln[(I_L/I_S) + 1]$$



(b)

Fig. 4.44 *(b) Inversion of Fig. 4.44(a) about the voltage axis*

$$V_{oc} = \frac{kT}{q}\ln[(I_L/I_S) + 1] \cong \frac{kT}{q}\ln(I_L/I_S) \tag{4.10}$$

The output power delivered by a solar cell to an external load is expressed as

$$P = IV = I_s V(e^{qV/kT} - 1) - I_L V \tag{4.11}$$

The condition for maximum power is $(dP/dV) = 0$ or we get,

$$V_m = \frac{kT}{q}\ln\left[\frac{(I_L/I_S) + 1}{\left(\frac{qV_m}{kT}\right) + 1}\right] \cong V_{oc} - \frac{kT}{q}\ln\left(\left(\frac{qV_m}{kT}\right) + 1\right) \tag{4.12}$$

$$I_m = I_s\left(\frac{qV_m}{kT}\right)e^{qV_m/kT} \cong I_L\left(1 - \frac{1}{qV_m/kT}\right) \tag{4.13}$$

The maximum output power P_m is given by

$$P_m = I_m V_m \cong I_L\left[V_{oc} - \frac{kT}{q}\ln\left(1 + \frac{qV_m}{kT}\right) - \frac{kT}{q}\right] \tag{4.14}$$

4.4.7 The Conversion Efficiency

The conversion efficiency of a solar cell is given by

$$\eta = \frac{\text{Maximum output power}}{\text{Input power}} = \frac{I_m V_m}{P_{in}} \tag{4.15}$$

$$\eta = \frac{I_m V_m}{P_{in}} = \frac{I_L\left[V_{oc} - \frac{kT}{q}\ln\left(1 + \frac{qV_m}{kT}\right) - \frac{kT}{q}\right]}{P_{in}} \tag{4.16}$$

or it can be given by

$$\eta = \frac{FF \cdot I_{sc}V_{oc}}{P_{in}} \tag{4.17}$$

where *Fill Factor (FF)* is defined as the maximum power rectangle to the rectangle of ($I_{sc} \times V_{oc}$).

$$FF = \frac{P_m}{I_{sc}V_{oc}} = \frac{I_m V_m}{I_{sc}V_{oc}} = 1 - \frac{kT}{qV_{oc}}\ln\left(1 + \frac{qV_m}{kT}\right) - \frac{kT}{qV_{oc}} \tag{4.18}$$

The ideal efficiency can be obtained from the ideal *I-V* characteristics defined by Eq. (4.8). The saturation current density is obtained from Eq. (4.9) for a given semiconductor. The short-circuit current I_{sc} ($= I_L$) is the product of q and the number of the available photons with energy $h\nu \geq E_g$ in the solar spectrum. Once I_s and I_L are known, the output power P and the maximum power P_m can be obtained from Eqs (4.11) to (4.14). The input power P_{in} is the integration of all the photons in the solar spectrum. The semiconductors with band-gap between 1 and 2 eV can all be considered solar cell materials.

4.4.8 Types of Solar Cells

1. P-N Homo-Junction solar cells
2. Hetero-Junction solar cells
3. Thin-film solar cells

It can also be broadly divided into crystalline (i.e., monocrystalline and polycrystalline) and amorphous solar cells. *Crystalline solar cells* are normally in the form of solar panels. The output of each cell is between 0.5 and 0.6 volts, and 36 cells are used to produce 20 volts of an open-circuit voltage. This is adequate to charge a 12-volt battery. The most common material used in solar cells is a single-crystal silicon. Its efficiency is about 25% as it is most sensitive to infrared light, and radiation in this region of the electromagnetic spectrum is relatively low in energy. Polycrystalline solar cells are made by a casting process in which molten silicon is poured into a mould and allowed to cool, and then sliced into wafers. The polycrystalline cells are cheaper than single crystal cells, with efficiency less than 20% due to internal resistance at the boundaries of the silicon crystals. Amorphous cells are made by depositing silicon onto a glass substrate from a reactive gas such as silane (SiH_4). Amorphous solar cell can be applied as a thin film to glass or plastic substrates. The advantages of thin-film cells are easier deposition and assembly on inexpensive substrates, the easy mass production, and the high suitability to large applications. Their efficiencies are presently only about 10% due to significant internal energy losses.

A number of other materials like gallium arsenide, copper indium diselenide and cadmium telluride can also be used to make solar cells. Note that solar cells are sensitive to different wavelengths of light as a function of the materials they are built from. Therefore, some cells are better performers outdoors while others are better performers indoors. Some solar cells have improved energy conversion efficiency by incorporating two or more layers of different materials with different wavelength sensitivities. Top layers are designed to absorb higher energy photons while allowing lower energy photons through to be absorbed by the layers beneath.

P-N Junction Solar Cell

The efficiency of the cell depends on the doping level of the P-region so that electrons generated in this region can diffuse to the junction before they recombine. So a proper match is required between the diffusion length of the electrons L_n, thickness of the P-region and the mean optical penetration depth. A large contact potential V_0 is required to obtain a large photovoltage which is achieved by heavy doping. Long lifetimes are needed for the charge carriers, but heavy doping reduces the lifetime of electrons and holes generated. The series resistance (R_s) of the solar cell should be small so that the power is not lost as heat due to Joule heat in the device itself. This is the major problem in solar cells, especially in the thin region at the surface. So the surface is partially contacted. A series resistance of few ohms can reduce the output voltage to appreciable levels. The resistance of the P-region of the diode of a solar cell is kept low by keeping area of the P-side of the diode large. The N-region is normally thin. This region is contacted at the edge and the current should be along the thin N-region to the contact. This produces large series resistance. The contact fingers distributed over the surface of the N-region will reduce this resistance. Much of the photon flux from the sun has less energy than the cell band gap, so they are not absorbed. For the high-energy photons that are absorbed, the recombination of electron–hole pairs at the surface takes place. This is the reason why the output from a solar cell is limited. A well-prepared Si solar cell is expected to give an efficiency of 10% for solar energy conversion and this gives an output of 100 W/m^2 of electrical power due to full illumination. This is not cost-effective. Also, the absorbed solar energy increases the temperature of the cell, thereby generating nonequilibrium concentration of charge carriers. The remedy is the use of GaAs and related compound with direct band gaps that can be used up to 100°C. The hetero junctions are very useful for good conversion efficiency and better for operating temperatures with reasonable efficiency. To reduce the surface recombination, the surface of the solar cell is coated with antireflection coating. The maximum output power is achieved by keeping a large–area junction located near the surface of the device. The thickness of the N-region must be less than the diffusion length in the N material.

Hetero Junction Solar Cells

These solar cells use two dissimilar semiconductor materials. The band gap of two materials will be such that $E_{g2} < h\nu < E_{g1}$. Incident photons are absorbed in the material on the top having a band gap E_{g2}. So the photocarriers are generated and diffuse into material with band gap E_{g1} which is beneath the first one. Due to differences in band energies of the two materials, a discontinuity occurs in the energy band diagram of the junction of the two materials. They are very useful for good conversion efficiency and suitable for higher operating temperatures. GaAs–AlGaAs hetero-junction cell is widely used because of good conversion efficiency even at higher temperatures.

Thin-Films Solar Cells

A Thin-Film Solar cell (TFSC), also called a Thin-Film Photovoltaic Cell (TFPV), is a solar cell that is made by depositing one or more thin layers (thin film) of photovoltaic material on a substrate. The thickness range of such a layer is wide and varies from a few nanometres to tens of micrometres.

Many different photovoltaic materials are deposited with various deposition methods on a variety of substrates. Thin-film solar cells are usually categorised according to the photovoltaic material used:

- Amorphous silicon (a-Si) and other thin-film silicon (TF-Si)
- Cadmium Telluride (CdTe)
- Copper indium gallium selenide (CIS or CIGS)
- Dye-sensitised Solar Cell (DSC) and other organic solar cells

Thin-film Silicon A silicon thin-film cell is a thin-film cell that uses amorphous (a-Si or a-Si:H), protocrystalline silicon, nanocrystalline (nc-Si or nc-Si:H) or black silicon. Thin-film silicon is opposed to wafer silicon (also called bulk or crystalline silicon).

Design and Fabrication The silicon is mainly deposited by Chemical Vapour Deposition (CVD), typically Plasma-Enhanced (PE-CVD), from silane gas and hydrogen gas. Other deposition techniques being investigated include sputtering and hot-wire techniques.

The silicon is deposited on glass, plastic or metal which has been coated with a layer of Transparent Conducting Oxide (TCO).

A P-i-N structure is usually used, as opposed to an N-i-P structure. This is because the mobility of electrons in a-Si:H is roughly 1 or 2 orders of magnitude larger than that of holes, and thus the collection rate of electrons moving from the P- to N-type contact is better than holes moving from P- to N-type contact. Therefore, the P-type layer should be placed at the top where the light intensity is stronger, so that the majority of the charge carriers crossing the junction would be electrons.

Thin-film solar cells are constructed using thin films of semiconducting materials deposited on a suitable substrate like glass, ceramic or metal. Metals and conductors form active substrates whereas insulating substrates are normally passive. Normally, thin-film deposition uses a variety of techniques like thermal evaporation, chemical vapour deposition, molecular beam epitaxy, sputtering, electroplating etc. The films deposited may be polycrystalline or amorphous. Thin-film solar cells are economical but having low efficiencies and stabilities compared to single-crystal solar cells.

Solar-cell arrays are spread over the surface of a satellite to supply power for the electronic equipment in satellites for a long time. This is the advantage over solid state batteries. They can also be attached to the main body of satellites in the form of paddles. They have about 10% efficiency with about 100 W/m^2 of electrical power under full illumination. To improve the power output, large-area Si cells are fabricated.

4.4.9 Advantages, Engineering Applications and Disadvantages

Advantages

- Solar cells are an inexpensive electrical source of energy. Once installed, they can work for years without any expenditure on maintenance.
- Solar cells do not use any fuels. So they are not dangerous.
- They produce enough voltage and current required for operation of equipment in satellites and rockets.
- They neither pollute atmosphere, nor produce noise when they are working, because they do not produce movements and vibrations when they are in operation.
- For small-power applications they can be used to replace chemically operated dry or wet batteries, where their small size is quite convenient.

Engineering Applications The applications found in different fields are rural electrification, water pumping, domestic supply, healthcare, lighting, ocean navigation aids, electric power generation in space, etc.

Disadvantages

- Solar cells are not normally used as optical detectors like photodiodes, because of their very slow response compared to photodiodes.
- To get large output voltage, solar panels occupy large space.
- Solar cells are very delicate. They are brittle, and can be easily broken into small pieces.
- Solar panels must be cleaned from time to time to get better results.
- Solar cells require uninterrupted solar light input. So, during night their operation stops. To get current at night, solar cells should be supported by storage batteries.

Solar cells, at present, furnish the most important long-duration power supply for satellites and space vehicles. They have also been successfully employed in small-scale terrestrial applications. The solar cell is considered a major candidate for obtaining energy from the sun, with high conversion efficiency, can provide nearly permanent power at low operating cost, and is virtually free of pollution. Recently, research and development of low-cost, flat-panel solar cells, thin-film devices, concentrator systems, and many innovative concepts have increased. In the near future, the costs of small solar-power modular units and solar-power plants will be economically feasible for large-scale use of solar energy.

PHOTODIODES AND PHOTOVOLTAIC CELLS

4.5

Photodiodes and photovoltaic cells are also special type of the *p*-*n* junction diode. A depletion layer is formed at the junction region by immobile negatively charged acceptor atoms in the *p*-type material and immobile positively charged donor ions in the *n*-type material, during the formation of the junction. The electric field due to these ions stops the motion of majority carriers but accelerate minority carriers across the junction. When a photon is incident on the device, electron-hole pairs are generated. In case of electron-hole pairs generated within the depletion region, the electric field acting across the region causes the pair to separate as shown in Fig. 4.45.

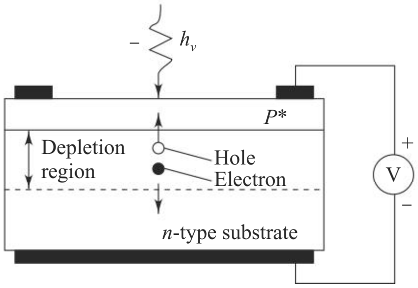


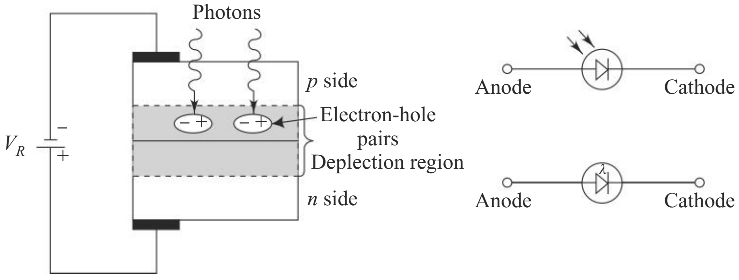
Fig. 4.45 *Separation of electron-hole pairs in a photovoltaic cell*

This charge separation can be utilized in two ways.

1. If the diode is short circuited, a current flows between p and n regions. This is known as **photoconductive mode** and diode must be in reverse bias condition to avail this mode, e.g., **Photo diodes**
2. If the diode is open circuited then potential difference across p and n regions appear. This mode is known as **photo voltaic mode** of operation, e.g., **solar cells**

4.5.1 Photodiodes

A semiconductor photodiode is essentially a reverse-biased p-n junction. The structure and symbol of a photodiode is shown in Fig. 4.46 (a) and (b).



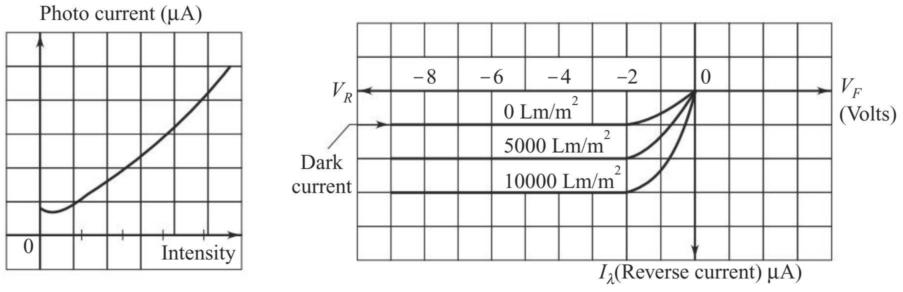
(a) (b)

Fig. 4.46 *(a) construction and (b) circuit symbol of Photodiode.*

When a reverse bias is applied across the junction as shown in the figure, the depletion layer widens as mobile carriers are swept to their respective majority sides. The motion of minority carriers forms the reverse leakage current of the diode. Therefore, even in the absence of light, a small leakage current exists in reverse bias condition. This leakage current is called *dark current*. The amount of dark current depends on the reverse bias voltage, the series resistance and the ambient temperature.

Now, if the diode is exposed to light, photons are absorbed mainly in the depletion layer and also in the neutral regions. A photon of energy $hv^{3*} E_g$ incident in or near the depletion layer of the diode will be capable to 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 are known as photocarriers. The electron–hole pairs generated in the depletion layer separate and drift in opposite directions under the action of the electric field. The electrons travel into the n-region and holes into p-region. Such a transport process causes an electric current in the external circuit in excess of the already existing dark current (reverse leakage current). The photocurrent created in the external circuit is always in the reverse direction, i.e., from the n to the p region. The number of electron–hole pairs will be dependent on the intensity of light (number of photons). These electrons and holes will be attracted towards the positive and negative terminals respectively of the external source, to constitute the photocurrent.

With increase in the light intensity, more number of electron hole pairs is generated and the photocurrent increases. Thus, the photocurrent is proportional to the light intensity. Increasing the level of illumination increases the flow of reverse current. It is, therefore, desirable that the depletion region be sufficiently wide so that a large fraction of incident light can be absorbed. Thus, optical excitation leads to an increase in the reverse-biased current. Hence, the diode can be used as a photodetector using a reverse bias voltage, as the measured photocurrent is proportional to the incident light intensity (Fig. 4.47a).



(a) (b)

Fig. 4.47 *(a) Output characteristics (b) V-I characteristics of photodiode*

The illumination (*I-V*) characteristic of a photodiode is shown in Fig.4.47 (b). It reveals that

- reverse current is independent of the reverse voltage
- reverse current increases with an increase in the intensity of light
- current produced in the diode with no incident light is called dark current which is a function of temperature only.

They are used in the cameras for sensing the light intensity, in the optical communication system, in the light intensity meters, in media players etc.

4.5.2 Photovoltaic Cells

Photovoltaic is the direct conversion of light into electricity at the atomic level. An unbiased *p-n* junction diode exposed to light develops voltage V_{oc} across it. This is known as *photovoltaic effect* and the diode is known as *photovoltaic cell*. Thus, a diode in a photovoltaic mode is a photovoltaic cell. It is an unbiased diode (Fig.4.45) illuminated by radiation. When the diode is illuminated by light, photons are absorbed mainly in the depletion layer and also in the neutral regions. A photon of energy hv $^{3}*$ E_g incident in or near the depletion layer of the diode will generate an electron–hole pair. These are separated by the internal electric field acting across the depletion region and drift in opposite directions. The electrons drift into *n*-region and holes into *p*-region, as illustrated in Fig.4.45.

They are widely used in solar-cell applications with appropriate design modifications.

PHOTOCONDUCTIVITY

4.6

When light or radiation is incident on an intrinsic semiconductor, the electrons are excited from the valence band to the conduction band leaving behind holes in the valence band which generates free electrons and holes in the material without leaving the material. This will result into the increase of free charge carrier concentration within the semiconductor. This phenomenon is known as *internal photoelectric effect*. This increase in free charge carriers leads to an increase in the conductivity of the semiconductor. Therefore, light-induced increment in the electrical conductivity is known as *photoconductive effect* or simply *photoconductivity*. The application of an electric field to the semiconductor causes the drifting of electrons and holes through the material and as a result, an electric current flows in the circuit.

HALL EFFECT AND HALL COEFFICIENT

4.7

If any non-insulator material strip carrying current I is placed in a transverse magnetic field B (as shown in Fig. 4.48) a voltage V_H (Hall voltage) develops across the surfaces which are normal to both the magnetic field and the current.

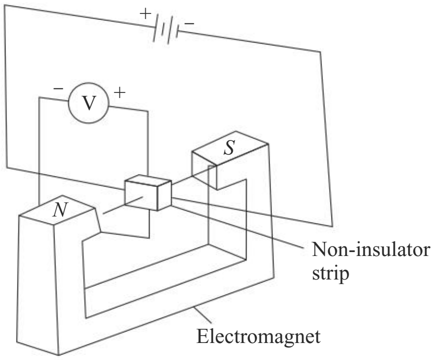


Fig. 4.48 *Hall effect arrangement*

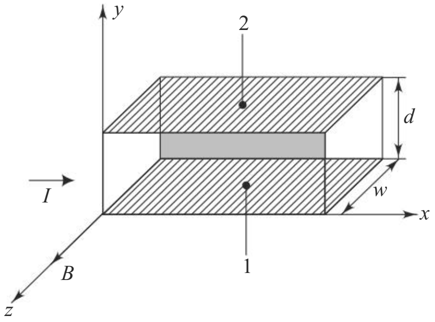


Fig. 4.49 *Hall effect*

Consider a rectangular strip of an n-type semiconductor material as shown in Fig.4.48. Let the current I flow through it along the positive X-direction and the magnetic field B act along the positive Z direction (i.e., B acts perpendicular to plane of paper in the outward direction). Hence, a Lorentz force F acts on the charge carriers along the negative Y direction. As the semiconductor is n-type, the majority charge carriers are electrons. This force due to the magnetic field has the magnitude,

$$F_L = Bev$$

where, v = drift velocity of electron and e = charge of electron

Thus, the electrons will be forced to move downward towards side 1 in Fig. 4.49 and side1 becomes negatively charged with respect to side 2. This causes electric field between these surfaces hence Hall voltage develops across these surfaces.

Let E be the electric field intensity due to Hall voltage V_H. Then the force due to this electric field balances the force due to magnetic field.

$$\therefore \qquad eE = Bev$$

$$\therefore \qquad E = Bv$$

But, $E = \frac{V_H}{d}$ where, d is the distance between surfaces 1 and 2

$$\therefore \qquad \frac{V_H}{d} = Bv$$

Current $I = neav$ and $V_H = Bvd$

$$\therefore \qquad v = \frac{I}{neA}$$

$$\therefore \qquad V_H = B\left(\frac{I}{neA}\right)d$$

where, n = carrier concentration and A = area of the strip

Thus, the Hall voltage is given by,

$$V_H = \frac{1}{ne}\frac{I}{A}Bd \text{ if } J = \frac{I}{A}$$

then,

$$V_H = \frac{1}{ne}BJd$$

From the expression it is clear that if the polarity of V_H is positive at the surface 2 (as in the above discussion) then the carriers must be electrons. But if the surface 1 has positive polarity with respect to the surface

2 then the carriers are positively charged holes and the semiconductor is *p*-type. Thus, measurement of Hall voltage helps to determine whether the given semiconductor is *p*- type or *n*-type.

If the volume charge density ρ is the charge per unit volume of the material, *w* is the width of the specimen in the direction of the magnetic field then, $\rho = ne$ and $A = wd$ so,

$$V_H = \frac{BI}{\rho\, w}$$

From the above equation, the charge density ρ can be determined if

V_H, *B, I* and *w* are measured, But $\rho = n\,.e$ so, we can determine the carrier concentration *n*. If **Hall coefficient** is defined as,

$$R_H = \frac{1}{ne} = \frac{1}{\rho} \quad \text{then,}$$

the Hall voltage in terms of Hall coefficient,

$$V_H = R_H = \frac{BI}{w}$$

$$R_H = \frac{V_H w}{BI}$$

$$R_H = \frac{Ewd}{BI}$$

$$R_H = \frac{E}{JB} \quad \text{because,} \quad J = \frac{I}{wd}$$

Thus, the Hall coefficient R_H of a material is defined as the electric field strength per unit surface current density due to unit transverse magnetic field. We know the conductivity of the material can be given as,

$$\sigma = ne\mu \quad \text{where } \mu = \text{mobility of the carrier}$$

$$\therefore \qquad \sigma = \frac{\mu}{R_H}$$

Thus, the mobility of a carrier can be determined with measurement of conductivity and Hall coefficient.

Applications The Hall effect is important in material science, semiconductor technology and engineering as it helps to determine the type of semiconductor, the sign of majority charge carriers, the majority charge carrier concentration, the mobility of majority charge carriers, and the mean drift velocity of majority charge carriers.

The Hall effect is widely used in various fields for a variety of applications. The Hall probe is employed for such applications. The Hall probe consists of a thin rectangular plate cut from a semiconductor crystal having high carrier mobility; Current-carrying electrodes are soldered to two opposite lateral faces of the plate. Hall electrodes are soldered to the remaining two lateral faces, which are used for measuring the Hall voltage developed in the plate. Hall probe in the form of a thin semiconductor film deposited onto a thin mica substrate is used sometimes.

One of the important applications of Hall effect consists in measuring magnetic fields. Hall voltage is proportional to the magnetic field intensity. Knowing the parameters of the Hall probe, and applied current, we can determine the intensity of the magnetic field. Hall probes can be used for static as well as variable magnetic fields. Hall probes measure variable magnetic fields up to a frequency of 10^{12} Hz. Another application of the Hall effect is to measure extremely large currents, where conventional ammeters cannot be used. The current being measured is used for generating a magnetic field and the Hall voltage developed in the probe is proportional to the magnetic field and hence to the current being measured.

Solved Problems

1 *A current of 0.25 A is passed through a metal strip of 0.2 × 10^{-3} thickness m and width same as thickness in a Hall effect experiment. For an applied magnetic field of 0.2 T, Hall voltage was measured to be 0.15 × 10^{-3}V. Calculate the carrier concentration.*

Solution Hall voltage is given by

$$V_H = R_H \frac{BI}{w} = \frac{1}{ne}\frac{BI}{w} = \frac{BI}{new}$$

$$n = \frac{IB}{V_H ew}$$

$$= \frac{0.25 \times 0.2}{0.15 \times 10^{-3} \times 1.6 \times 10^{-19} \times 0.2 \times 10^{-3}}$$

$$= 1.04 \times 1025 \text{ no. of carriers/m}^3$$

2 *The Hall coefficient (R_H) of a semiconductor is 3.22 × 10^{-4} m^{3}C^{-1}. Its resistivity is 9 × 10^{-3} ohm-m. Calculate the mobility and carrier concentration of the carriers. [Dec.2008, 4 Marks]*

Solution Given: $e = 1.6 \times 10^{-19}$C, $R_H = 3.22 \times 10^{-4}$ m^{3}C^{-1}

Hall coefficient

$$R_H = \frac{1}{ne}$$

$$\therefore \quad n = \frac{1}{R_H e}$$

$$= \frac{1}{3.22 \times 10^{-4} \times 1.6 \times 10^{-19}}$$

$$n = 1.94 \times 10^{22} \text{ /m}^3$$

Now

Conductivity: $\sigma = ne\mu$

Resistivity: $\rho = \frac{1}{\sigma} = \frac{1}{ne\mu}$

Mobility: $\mu = \frac{1}{ne\rho} = \frac{1}{1.94 \times 10^{22} \times 1.6 \times 10^{-19} \times 9 \times 10^{-3}} = 0.0357 \text{ m}^2\text{/Vs}$

3 *A rectangular semiconductor specimen of 1 × 10^{-3} thickness m is placed in the magnetic flux density of 0.5 Wb/m^2. A current of 1.5 mA is flowing through specimen in one direction. Calculate the resulting Hall voltage. Here, Hall coefficient of the material is 1 × 10^{-2} m^3/C. [June 2009, 5 Marks]*

Solution Given;

$$B = 0.5 \text{ Wb/m}^2$$
$$I = 1.5 \times 10^{-3} \text{ A}$$
$$w = 10^{-3} \text{ m}$$

$$R_H = \frac{1}{\rho} = \frac{1}{ne} = 1 \times 10^{-2} \text{ m}^3/\text{C}$$

Hall voltage:

$$V_H = \frac{BI}{\rho w}$$
$$= R_H \frac{BI}{w}$$
$$= 1 \times 10^{-2} \times \frac{0.5 \times 1.5 \times 10^{-3}}{1 \times 10^{-3}} = 7.5 \text{ mV}$$

4 *Calculate the energy gap in Ge when it is transparent to radiation of wavelength of 12000 Å. [h = 6.625 × 10^{-34} J·s; c = 3 × 10^8 m/s], [June 2009, 5 Marks]*

Solution Energy gap:

$$E_g = h\nu = \frac{hc}{\lambda}$$

Given:

$$\lambda = 12000 \times 10^{-10} \text{ m}, h = 6.625 \times 10^{-34} \text{ J·s}, c = 3 \times 10^8 \text{ m/s}$$

Energy gap:

$$E_g = \frac{6.625 \times 10^{-34} \times 3 \times 10^8}{12000 \times 10^{-10}}$$
$$= 1.66 \times 10^{-19}\, J \times \left(\frac{1 \text{ eV}}{1.6 \times 10^{-19} \text{ J}}\right)$$
$$= 1.03 \text{ eV}$$

5 *Calculate the open-circuit voltage and the output power at a voltage of 0.35 V for a solar cell having reverse saturation current 1 nA at 300 K, having photocurrent of 100 mA and cell area 4 cm^2 [thermal voltage at 300 K is 0.0258 V].*

Solution For open circuit voltage, V_{oc}, $I = 0$.

Thermal voltage $V_T = (kT/q) = 0.0258$ V

Photocurrent $I_L = 100 \text{ mA} = 100 \times 10^{-3}$ A

Reverse saturation current $I_s = 1 \text{ nA} = 1 \times 10^{-9}$ A

Voltage $V = 0.35$ V

$$V_{oc} = \frac{kT}{q} \ln\left(\frac{I_L}{I_S} + 1\right) \equiv V_T \ln\left(\frac{I_L}{I_S}\right)$$

$$V_{oc} = (0.0258)\ln\left(\frac{100\times10^{-3}}{1\times10^{-9}}\right) = 0.475\text{ V}$$

The output power, P

$$P = IV = I_s V\,(e^{qV/kT} - 1) - I_L V$$
$$P = I_S V\,(e^{V/VT} - 1) - I_L V$$

Here, I_s and I_L are reverse currents so use −ve sign for reverse current.

$$P = (-1\times10^{-9})(0.35)\,(e^{0.35/0.258} - 1) - (-100\times10^{-3})(0.35) = 3.48\times10^{-2}\text{ W}$$
$$P = 3.48\times10^{-2}\text{ W}$$

6 *Consider a silicon P-N solar cell of area 2 cm^2. The doping of the solar cell is $N_A = 1.7\times10^{16}$ cm^{-3} and $N_D = 5\times10^{19}$ cm^{-3}, effective density of states in conduction band and valence band are $N_C = 2.89\times10^{19}$ cm^{-3}, and $N_V = 2.66\times10^{19}$ cm^{-3} and the carrier lifetime for holes and electrons are $\tau_n = 10$ µs, $\tau_p = 0.5$ µs, $D_n = 9.3$ cm^2/s, $D_p = 2.5$ cm^2/s, and short circuit current $I_L = 95$ mA, energy band gap E_g for Si = 1.12 eV, thermal voltage at 300 K is 0.0258 V.*

(a) Calculate reverse saturation current density and hence, reverse saturation current. Also, calculate and plot the I-V characteristics of a solar cell for different values of V = 0.1, 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, 0.62, 0.63 V,

(b) Open-circuit voltage

(c) Determine the maximum output power of the solar cell, all at room temperature

Solution (a) current is given by

$$I = I_S\,(e^{qV/kT} - 1) - I_L$$

and the Saturation current density is given by

$$J_S = \frac{I_S}{A} = qN_V N_C\left[\frac{\sqrt{D_p}}{N_D\sqrt{\tau_p}} + \frac{\sqrt{D_n}}{N_A\sqrt{\tau_n}}\right] e^{-Eg/kT}$$

$$J_S = (1.6\times10^{-19}\times2.66\times10^{19}\times2.86\times10^{19}\left[\frac{\sqrt{9.3}}{1.7\times10^{16}\sqrt{10\times10^{-6}}} + \frac{\sqrt{2.5}}{5\times10^{19}\sqrt{0.5\times10^{-6}}}\right] e^{-1.12/0.0259}$$

$$J_S = (1.217\times10^{20})[5.67\times10^{-14} + 4.47\times10^{-17}]\,(1.658\times10^{-19})$$

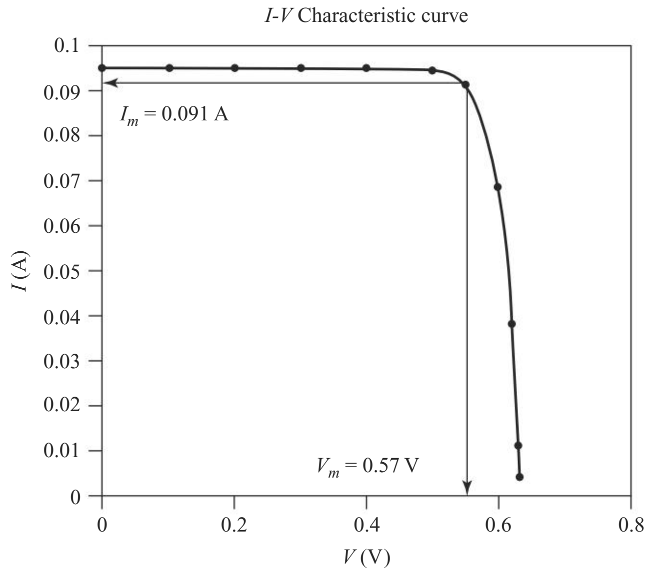
$$J_S = 1.1449\times10^{-12}\text{ A/cm}^2$$

Saturation current I_s is given by

$$I_S = J_S\times A = 1.1449\times10^{-12}\times2 = 2.289\times10^{-12}\text{ A}$$

From Fig. 4.49, we can find the value of V_{oc}, I_{sc}, I_m, V_m and can calculate P_m.

$$V_{oc} = 0.63\text{ V}$$
$$I_{sc} = I_L = 0.095\text{ A}$$
$$I_m = 0.091\text{ A}$$
$$V_m = 0.57\text{ V}$$
$$P_m = (I_m\times V_m) = (0.091\times0.57) = 0.0518\text{ W} = 51.87\text{ mW}$$



Voltage V in volts	Current I
0	0.095
0.1	0.095
0.2	0.095
0.3	0.095
0.4	0.094988
0.5	0.094446
0.55	0.09118
0.6	0.068666
0.62	0.037999
0.63	0.011138
0.632	0.004405
0.633	0.000839
0.634	−0.00287

(b) Open-circuit voltage V_{oc} is from

$$V_{oc} = \frac{kT}{q} \ln\left(\frac{I_L}{I_S} + 1\right) \cong V_T \ln\left(\frac{I_L}{I_S}\right)$$

$$V_{oc} = (0.0258) \ln\left(\frac{95 \times 10^{-3}}{0.0023 \times 10^{-9}}\right) = 0.63 \text{ V}$$

$$V_{oc} = 0.63 \text{ V}$$

(c) Maximum output power P_m from the graph is 57.87 mW

$$P_m = (I_m \times V_m) = (0.091 \times 0.57) = 0.0518 \text{ W} = 51.87 \text{ mW}$$

7 *A silicon P-N solar cell has a reverse saturation current I_s of 2 nA. Its short-circuit current is 150 mA. Find out the open-circuit voltage. If maximum output power is 51.87 mW, calculate Fill Factor (FF). Considering the solar power density to be of 1000 W/m^2, calculate the efficiency of the given solar cell. Given, thermal voltage = 0.022 V*

Solution

$$V_{oc} = \frac{kT}{q} \ln\left(\frac{I_L}{I_S} + 1\right) \cong V_T \ln\left(\frac{I_L}{I_S}\right)$$

$$V_{oc} = 0.026 \times \ln\left(\frac{150 \times 10^{-3}}{2 \times 10^{-9}}\right) = 0.47 \text{ V}$$

$$FF = \frac{I_m V_m}{I_L V_{oc}} = \frac{P_m}{I_L V_{oc}} = \frac{0.05187}{0.15 \times 0.47} = 0.7357$$

$$\eta = \frac{P_m}{P_{in}} = \frac{0.05187}{0.1} = 0.5187 = 51.87\%$$

8 *For a given solar cell, V_{oc}, I_{sc}, and FF are respectively given as 500 mV, 125 mA, and 80 %. Find out the maximum power generated by a solar cell and if the incident solar power density is 840 W/m^2, find the conversion efficiency of the cell. Consider the unit dimension of the solar cell.*

Solution

$$V_{oc} = 500 \text{ mV}$$

$$I_L = 125 \text{ mA}$$

$$FF = 0.8$$

Input power density = 840 W/m^2 = 840/10000 = 0.0840 W/cm^2,

So input power $P_{in} = 0.084$ W

$$\eta = \frac{FF \cdot I_L V_{oc}}{P_{in}} = \frac{0.8 \times 0.125 \times 0.5}{0.084} = 0.5952 = 59.52\%$$

9 *Solve the above problem if the above values are changed to V_{oc} = 600 mV, I_L = 160 mA, FF = 81.4, and incident solar power density = 840 W/m^2. Comment on the result obtained.*

Solution

$$V_{oc} = 600 \text{ mV}$$

$$I_L = 160 \text{ mA}$$

$$FF = 0.814$$

Input power density = 840 W/m^2 = 840/10000 = 0.0840 W/cm^2,

So input power $P_{in} = 0.084$ W

Maximum power generated is

$$P_m = I_L \times V_{oc} \times FF = (160 \times 10^{-3} \times 600 \times 10^{-3} \times 0.814)$$
$$= 0.078144 = 78.14 \text{ mW}$$

Power efficiency = (Maximum power generated/Incident power) = 0.078144/0.084 = 0.93 = 93%

Comment Looking at the above example, we conclude that the real solar power does not have an efficiency of more than 30%. So the given problem does not reflect a real solar cell.

10 *The band gap of Si is 1.12 eV and that of GaAs is 1.424 eV. What are the cutoff wavelengths for these materials so that they can be utilised for fabrication of a solar cell?*

Solution We know that the photon of light should have minimum energy E_g.

$$E_g = h\nu_{\min} = \frac{hc}{\lambda_{\max}} = \frac{hc}{\lambda_c}$$

$$\lambda_c = \frac{hc}{E_g} = \frac{6.62 \times 10^{-34} \times 3 \times 10^8}{E_g \times 1.6 \times 10^{-19}} = \frac{1.24}{E_g}\ \mu\text{m}$$

For Si, $\lambda_c = \dfrac{1.24}{1.12} = 1.10\ \mu$m

For GaAs, $\lambda_c = \dfrac{1.24}{1.424} = 0.87\ \mu$m

11 *A solar cell having a fill factor of 0.7, open-circuit voltage 0.45 V is exposed with a light with an input power of 500 mW on a cell area of 1 cm^2. Calculate its conversion efficiency. Given* $I_{sc} = 200$ *mA.*

Solution

$$FF = 0.7$$
$$V_{oc} = 0.45 \text{ V}$$
$$P_{in} = 500 \text{ mW}$$
$$I_{sc} = 200 \text{ mA}$$

Conversion efficiency,

$$\eta = \frac{P_m}{P_{in}} = \frac{I_m V_m}{P_{in}} = \frac{(I_L \cdot V_{oc})\, FF}{P_{in}}$$

$$P_m = (IL \cdot V_{oc})\, FF$$
$$= 200 \times 10^{-3} \times 0.45 \times 0.7$$
$$P_m = 0.063 \text{ W} = 63 \text{ mW}$$
$$\eta = \frac{P_m}{P_{in}} = \frac{0.063}{0.5} = 0.126 = 12.6\%$$

Multiple-Choice Questions

1. A semiconductor is a perfect insulator at
  (a) room temperature (b) high temperature
  (c) 0 K (d) ice temperature

2. Zener diode is used as
  (a) voltage follower (b) amplifier
  (c) voltage regulator (d) current regulator

3. Operation of a photodiode is in the
  (a) reverse biased condition
  (b) forward biased condition
  (c) unbiased condition
  (d) both the biased conditions

4. In a semiconductor material, the conduction band and valence band are
  (a) separated by very small band gap
  (b) overlap each other
  (c) separated by very large band gap
  (d) depends only on temperature

5. If the Hall coefficient is positive for a given material then the material is
  (a) an insulator
  (b) an *n*-type semiconductor
  (c) a *p*-type semiconductor
  (d) a metal

6. In conductors, with decrease in temperature, the Hall coefficient value
  (a) increases
  (b) decreases
  (c) stays constant
  (d) increases first and then decreases

7. Potential barrier opposes the flow of
  (a) majority carriers (b) minority carriers
  (c) electrons (d) holes

8. Limiting the frequencies of light on a solar cell
  (a) reduces the amount of electricity produced
  (b) increases the amount of electricity produced

(c) does not have much effect on the amount of electricity produced
(d) none of the above

9. Changing the number of cells in the solar array
(a) changes the voltage output
(b) changes the electrical output
(c) makes no difference in the output
(d) both (a) and (b) are correct

10. The value of the solar constant is
(a) 1.335 kW/m^2 (b) 1.533 kW/m^2
(c) 1.358 kW/m^2 (d) 1.353 kW/m^2

11. We get ______ kW solar power per square metre.
(a) 1.353 (b) 0.66
(c) 47 (d) 3.8×10^{26}

12. ______ is used to make a solar cell.
(a) Silicon (b) Silver
(c) Selenium (d) Germanium

13. A solar cell converts
(a) heat energy into electrical energy
(b) solar energy into electrical energy
(c) heat energy into light energy
(d) solar energy into light energy

14. The word photovoltaic comes from words meaning
(a) wind energy
(b) brightness
(c) light and electricity
(d) picture which moves

15. A PV module is
(a) dozens of photovoltaic cells connected together
(b) wired in series
(c) wired in parallel
(d) all answers a, b, c

16. In the shade,
(a) less light strikes the PV cells
(b) less current is generated in PV cells
(c) the PV cell is cooler
(d) all answers a, b, c

17. Improving the efficiency of a PV cell can be done by
(a) adjusting the light-facing angle all day
(b) placing coloured acetates on the cell
(c) heating the cell
(d) changing its direction to north

18. Developing solar energy is important because it
(a) does not produce pollution
(b) can be utilised in most regions of India
(c) reduces our dependency on imported energy
(d) all of the above

19. When planning your future home you will
(a) never consider photovoltaic systems
(b) research the cost of a PV system as a supplement to the grid
(c) work with local builders to find out if PV will be practical
(d) b and c

ANSWERS TO MCQs

1. (c) 2. (c) 3. (a) 4. (a) 5. (c) 6. (c) 7. (a) 8. (a)
9. (d) 10. (d) 11. (a) 12. (a) 13. (b) 14. (c) 15. (d) 16. (d)
17. (a) 18. (d) 19. (d)

Practice Problems

1. A zener diode (8.2 V at 25 °C) has a positive temperature coefficient of 1 %. What is the zener voltage at 60 °C? **[8.23 V]**

2. A silver strip of ribbon shape with 0.1×10^{-3} m thickness and 0.5×10^{-2} m width is used in the Hall effect experiment. If 0.8 T magnetic field is applied perpendicular to this silver strip and the 2 A current is passing through it, calculate the Hall voltage. Given: density of silver = 10.5 g/cm^3. **[0.3×10^{-7} V]**

3. Determine the concentration of holes in silicon single crystals having donor concentration of 1.4×10^{24} /m^3, when the intrinsic carrier concentration is 1.4×10^{18} /m^3. Find the ratio of the electron to hole concentration. **[1.4×10^{12} /m^3, 1×10^{12}]**

4. Calculate the open-circuit voltage and the output power at a voltage of 0.5 V for a solar cell having a reverse saturation current of 1 nA at a temperature of 300 K, given a photocurrent of 0.2 A with a cell area of 4 cm^2. The thermal voltage at 300 K is 0.026 V. **[0.497 V, 0.21 W]**

5. If output power of a solar cell is 0.123 W at a voltage of 0.4 V, calculate the reverse saturation current of the cell. If the photocurrent is 0.2 A, calculate the open-circuit voltage. **[1.5 nA, 0.49 V]**

6. For a given solar cell, the open-circuit voltage is 0.47 V, the photocurrent is 0.4 A and the reverse saturation current is 5 nA. Calculate the thermal voltage and hence the temperature at which the solar cell is working. **[0.0258 V, 300 K]**

7. Injection current of a solar cell is 0.5 nA and short-circuit current is 0.1 A. If fill factor (FF) is taken to be unity with unit cross-sectional area and input solar power per one cm^2 is 0.135 W, calculate its efficiency. **[36.74%]**

8. A solar cell has a saturation current of 2 nA and a short-circuit current of 0.32 mA. Find the open-circuit voltage. **[0.31 V]**

9. A typical solar cell has an efficiency of 8%, maximum output power of 0.034 W, open-circuit voltage of 0.38 V, and saturation current of 5 nA. Calculate the photocurrent, and incident power if the thermal voltage is 0.026 V. **[11.12 mA, 0.425 W]**

Short-answer Questions

1. Estimate the spacing between different energy levels in an energy band.

2. State applications of Hall effect.

3. Define the following terms:

(i) Valence band (ii) Conduction band, (iii) Forbidden energy gap

4. Explain the breakdown mechanisms in a Zener diode.

5. Briefly discuss the use of LEDs as display devices in instrumentation.

6. What is reverse saturation current in a diode?

7. Explain Hall effect and state its importance.

8. Explain Hall effect and state its applications.

9. What is Hall effect? How does this effect show whether holes or electrons predominate in a semiconductor?

10. What causes majority carriers to flow at the moment when *p*-region and *n*-region are brought together? Why does this flow not continue until all the carriers have recombined?

11. Can the voltage be measured across the depletion layer?

12. Why is reverse saturation current independent of reverse bias?

13. What is reverse breakdown in a semiconductor *p*-*n* junction diode?

14. What is meant by photoconductivity?

15. What is meant by photovoltaic emf?

16. What is the difference between a solar cell and a photodiode?

17. State the principle of working of a solar cell.

18. What happens to a solar cell working if the energy of a photon is less than the energy band gap?

19. State a few industrial uses of a solar cell.

20. What are the advantages of a solar cell?

21. State the few disadvantages of a solar cell.

22. Why is a solar cell a non-conventional energy source?

23. What types of solar cells are available in the market?

24. What is the efficiency of a solar cell for different types?

25. Define fill factor of a solar cell.

26. What do you mean by short-circuit current and open-circuit voltage?

27. Draw the *I*-*V* characteristics curve and briefly explain how the power can be extracted from the solar cell.

28. Briefly explain the mechanism of operation of a solar cell.

29. What are the materials used in fabrication of a solar cell?

30. What are the main sources of energy for a solar cell? What is intensity and wavelength range of a solar cell?

31. What is the advantage of a hetero-junction solar cell?

32. Draw a neat sketch of a structure of a solar cell.

33. Why is antireflection coating used in the fabrication of a solar cell?

34. What is the main difference in the structure between P-N junction diode and a solar cell? How are solar cells made different from it?

35. What is an output voltage of a typical solar cell you know?

36. What is the level of doping in a solar cell? What is the purpose to keep it like that?

Long-answer Questions

1. Explain in brief the formation of energy bands in solids on the basis of band theory of solids.

2. Draw the energy-band diagrams of metal, semiconductor and insulator and discuss it briefly.

3. Write a note on 'Classification of solids on the basis of band theory'.

4. What are the intrinsic and extrinsic semiconductors? State their types and explain how are they obtained?

5. Define Hall coefficient. Derive the expression for it.

6. Explain Hall effect. Derive an expression for the Hall voltage. How can mobility of charge measured using Hall effect experiment?

7. Explain in brief the working of *pn* junction of a semiconductor diode in forward biased and reverse biased condition. State the conclusion that can be drawn from this.

8. Draw the energy-band diagram of *pn* junction diode for the case of (i) forward biased, and (ii) reverse biased conditions.

9. Describe the principle, construction working and characteristics of a solar cell.

10. Discuss briefly the *I*-*V* characteristics of solar cell. State the merit and demerit of solar cell.

11. Explain in brief the operation of a Zener diode in the forward and reverse biased condition

12. Draw and explain the *V-I* characteristics of a Zener diode.

13. What is LED? Discuss its principle of working, construction, circuit symbol, merits and demerits and its various applications

14. What do you mean by a photodiode? Explain in brief the biasing arrangement and construction. Discuss its application.

15. Write a note on (i) solar cells (ii) LEDS (iii) photodiode.

16. What is varacter diode? Discuss its biasing condition and its V-C_t characteristic curve.

17. For a uniformly doped PN-junction solar cell, draw the energy band diagram and explain the working principle. Explain the various characterising terms involved in solar cell study.

18. State the expression depicting *I-V* characteristic of a solar cell and obtain the expression for open-circuit voltage and maximum power output and hence obtain an expression for efficiency of a P-N junction solar cell.

19. Describe the equivalent circuit of a solar cell. Explain the working of a solar cell.

20. Define fill factor and efficiency of a solar cell and explain the factors that affect them.

21. Explain the photovoltaic effect in detail giving suitable diagram.

22. Explain in detail the construction and theory of working of a solar cell and explain series and parallel operation.

Questions from Previous Exams

(i) Give two examples of pentavalent impurities. (GTU- Jan' 2009)

Phosphorus (P), Arsenic (As)

(ii) The Hall coefficient (R_H) of a semiconductor is 3.22 × 10^{-4} m^{3}C^{-1}. Its resitivity is 9 × 10^{-3} ohm-m. Calculate the mobility and carrier concentration of the carrier. (GTU- Jan' 2009)

$R_H = 3.22 \times 10^{-4}\ \text{m}^3\text{C}^{-1}$.

$\rho = 9 \times 10^{-3}$ ohm-m

μ =?

n =?

$R_H = 1/\text{ne}$

$n = 1/\ R_H\, e = 1/(3.22 \times 10^{-4})(1.6 \times 10^{-19}) = 1.94099 \times 10^{22}/\text{m}$

$\mu = \sigma R_H$

$\mu = R_H/\rho = 3.22 \times 10^{-4}\ \text{m}^3\text{C}^{-}1/9 \times 10^{-3}\ \text{ohm-m} = 0.35777\text{m}^2\text{V}^{-1}\text{s}^{-1}$

(iii) Explain how the materials are classified into conductors, semiconductors and insulators with the help of energy-band diagram. (GTU- Jan' 2009)

(Refer Section 4.2)

(iv) Explain the term Hall Effect. Derive the relation between Hall voltage and Hall coefficient. (GTU- Jan' 2009)

(Refer Section 4.7)

(v) Short notes (1) LED (2) Solar Cell (GTU- Jan' 2009)

(Refer Section 4.3.4, 4.4)

(vi) What are intrinsic semiconductors? **(GTU- Jan' 2010)**

It is pure form without any impurities **(GTU- Jan' 2010)**

(vii) What are the extrinsic semiconductors? Explain the term Hall Effect. **(GTU- Jan' 2010)**

(Refer Section)

(vii) Explain LED (principle and application) **(GTU- Jan' 2010)**

(Refer Section 4.3.4)

(viii) The Hall coefficient (R_H) of a semiconductor is $3.22 \times 10^{-4}\ \mathrm{m^3C^{-1}}$. Its resitivity is 9×10^{-3} ohm-m. Calculate the mobility and carrier concentration of the carrier. (Given $e = 1.6 \times 10^{-19}$C) **(GTU- Jan' 2010)**

$$R_H = 3.22 \times 10^{-4}\ \mathrm{m^3C^{-1}}.$$

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

$$\mu = ?$$

$$n = ?$$

$$R_H = 1/\mathrm{ne}$$

$$n = 1/\ R_H\, e = 1/(3.22 \times 10^{-4}\ )(1.6 \times 10^{-19}) = 1.94099 \times 10^{22}/\mathrm{m}$$

$$\mu = \sigma R_H$$

$$\mu = R_H/\rho = 3.22 \times 10^{-4}\ \mathrm{m^3C^{-}1/9} \times 10^{-3}\ \text{ohm-m} = 0.35777\mathrm{m^2V^{-1}s^{-1}}$$

(ix) Write a short note on (i) LED and (ii) photodiode. **(GTU- June' 2010)**

(Refer Section 4.3.4, 4.5)

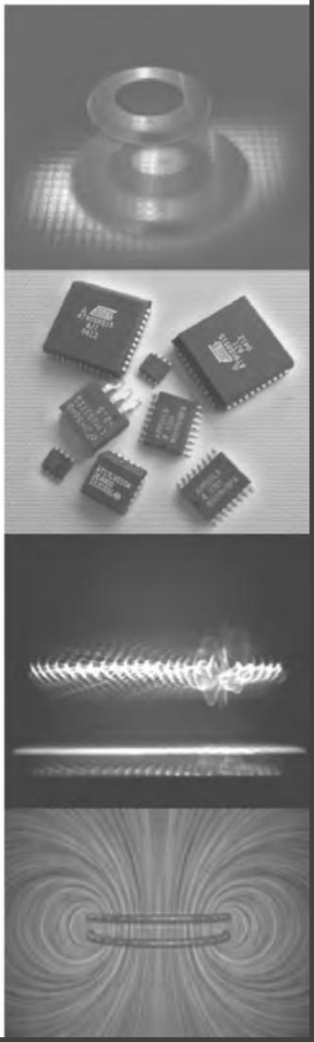
(x) Explain the Hall Effect and derive an expression of Hall coefficient. **(GTU- June' 2010)**

(Refer Section 4.7)

The McGraw-Hill Companies

Laser

5



Can we produce a light brighter than sunlight? Is it possible to send light through a very large distance without spreading? Do we have any source of light having complete coherence and perfectly monochromatic? All these characteristics you will find in LASER.

INTRODUCTION TO LASER

5.1

Lasers are devices that amplify light, just as transistors amplify electronic signals at audio, radio or microwave frequencies. Lasers have covered radiation at wavelengths ranging from the infrared range to ultraviolet and even the soft X-ray range.

In this chapter, we shall discuss the laser principles and devices in detail. A laser device consists of (1) laser medium like atoms, molecules, ions or semiconductor crystals; (2) pumping processes to excite these atoms, ions or molecules into higher quantum-mechanical energy levels; and (3) suitable optical feedback elements that allow the beam of radiation to either pass once through the laser medium or bounce back and forth repeatedly through the laser medium.

LASER is the acronym of Light Amplification by Stimulated Emission of Radiation. It is a light of special properties. Now we know that light is composed of ‘particles’ whose motion probability is determined by its wavelike behavior. The stimulated emission has the same frequency and phase as the incident radiation. So laser light would be very different from normal light. It is coherent. Two sources are coherent when they vibrate in the same phase or there is a constant phase difference. Coherent light sources have numerous applications. The light beams from lasers with coherence lengths of up to 10^{14} cycles will make many experiments and practical applications possible otherwise which would not be possible with the previously available light sources of coherence lengths extending almost up to 10^7 cycles.

Many scientists tried to find methods for amplified stimulated emission, but it was realised in 1960, about half a century after Einstein's prediction.

The first laser was invented by TH Maiman in May, 1960. It was a solid ruby laser. After the solid ruby laser, many kinds of lasers were invented—first, the uranium laser by IBM labs (in Nov., 1960), He–Ne Laser by Bell Laboratories in 1961, the semiconductor laser by Robert Hall at General Electric Labs in 1962, the Nd: YAG laser and CO_2 laser in 1964, the argon-ion laser in 1964, the chemical laser in 1965, and the metal vapour laser in 1966.

Einstein in 1917 predicted the phenomenon of stimulated emission—the basic principle involved in the lasing action. When an atom is in the excited state, it can make a transition to a lower energy state by emitting a photon. The emission process takes place in two different ways.

(i) Usually, the atom in the excited state cannot remain in the excited state for more than 10^{-8} second. After this interval, it emits radiation by jumping to a lower state even in the absence of any incident radiation. This occurs spontaneously. The rate of spontaneous emissions depends upon the number of atoms in the excited state.

(ii) In the second process, a photon of frequency n is incident on the atom in the excited state. This incident photon triggers the atom in the excited state to emit radiation. This is known as **stimulated emission of radiation.** The rate of stimulated or induced emission depends both on the intensity of the external fields and on the number of atoms in the excited state.

The stimulated emission phenomenon was first used by Ch. Townes in 1954 in the construction of **MASER,** the **microwave amplifier.** The MASER application was later extended to optical frequencies, which led to the development of the laser.

SPONTANEOUS AND STIMULATED EMISSION

5.2

We need to begin with the idea of atomic energy levels. We shall consider an idealised atom with two electron energy levels and one electron. The electron may be in either of the two energy levels.

If the electron is in the higher level 2, it may fall down into the lower level 1. In its doing this, it must give up an amount of energy equal to the energy difference between the two levels. This is the law of conservation of energy being applied. This energy is given up in the form of light.

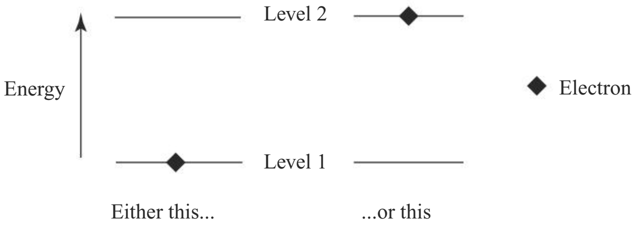


Fig. 5.1 *Idealised atom with two electron energy levels and one electron*

Light is also quantised and represented as groups of photons. Each photon carries one quantum of light energy. The amount of energy in a quantum depends upon the wavelength (colour) of the light.

$$E = \frac{hc}{\lambda} = hv \tag{5.1}$$

where E is the energy of photon, l is the wavelength, n is the frequency of light (photon), c is the velocity of light in vacuum, and h is the Planck's constant.

From Eq. (5.1), we infer that red light at 670 nm has a low energy and short wavelength such as blue light at 470 nm has a high energy per photon. Here, the wavelength of light is linked to the energy of a photon in a defined way. Thus, an electron in the idealised atom has given out a photon of defined energy, emitting light of a defined wavelength or 'colour'.

Some street lights contain sodium atoms which take electrical energy to move their electrons into higher levels and then fall back down to the lower level, giving light at 589 nm, the characteristic yellow colour of street lights. This process is known as **spontaneous emission.** The atom emits light spontaneously without external influences.

Other effects may occur if the atom is not isolated. Photons of the energy equal to the energy difference between upper and lower level may be utilised to move an electron from the lower level to the upper one. This process is known as **Absorption**, because the photon is absorbed in the process. If an atom with its electron in the upper level is exposed by a photon of the correct energy then it may cause the electron to fall to the lower level. This is known as stimulated emission which is very different from spontaneous emission. In the spontaneous process, the photon may travel in random direction and be emitted at any time. However, stimulated emission causes the emitted photon to travel in the identical direction to the passing photon and at the same time. The entire process is depicted in Fig. 5.2. Now we finally get to lasers.

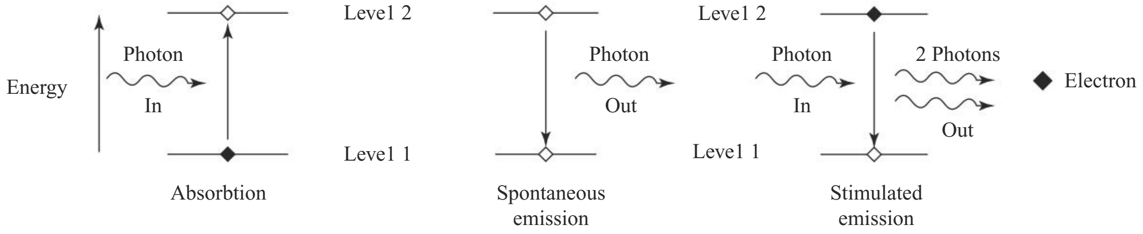


Fig. 5.2 *Absorption, spontaneous and stimulated emission*

Now let us consider a group of N atoms so that the three processes outlined above, all happen. Out of these, N_1 atoms have their electrons in the lower level and N_2 atoms have their electrons in the upper level. Let us look at the rate at which each process occurs.

Spontaneous Emission Out of N_2 atoms having electrons in the upper level, a certain proportion 'a' of the upper level electrons will emit in a given time. So, rate of spontaneous emission,

$$\frac{dN_2}{dt} \propto N_2 = aN_2$$

Absorption Out of N_1 atoms present in the lower level, a certain proportion 'b' of the lower level electron interacts with n number of incident photons in a given time.

$$\text{Rate of absorption} = b\,N_1\,n$$

Stimulated Emission Similarly, the rate of stimulated emission is given by the same form of equation with only consideration that the N_2 atoms are in the upper level and the electrons of these atoms will interact with n photons.

Rate of stimulated emission = $b\, N_2\, n$

The condition for the lasing action is that the last of these expressions is the largest, i.e., rate of stimulated emission should be greater than spontaneous emission. To increase the rate we should increase b, N_2 and n. Now 'a' and 'b' are constants depending on the particular atom used and are thus fixed once the material is fixed. So to make the rate of stimulated emission greater than the rate of spontaneous emission, we require n to be large, i.e., we need many photons in the laser. Also, for stimulated emission to be greater than absorption, we require N_2 to be greater than N_1, i.e., more number of atoms should have their electrons in the upper level than in the lower level, i.e., the population of the excited atoms should be more than that of the lower level. This is known as **population inversion** as it is normal for electrons to be in their lowest energy level. This can be seen from the fact that in the absence of any external influences, i.e., no photons, $n = 0$, the rate of stimulated emission will be zero and only spontaneous emission is possible, which will allow any electron that began in the upper level to fall to the lower level, but not vice-versa.

The main difficulty in producing a laser is creating this necessary population inversion in the populations of the two levels. It must be noted here that we are considering many atoms, each with a single electron, which may be in one of the two levels. We do not have many electrons in one atom since it would restrict their movements between levels by the rules governing how many electrons may occupy a single level of a particular atom.

Now assuming that we have a population inversion, $N_2 > N_1$, we can get the SE part of LA**SE**R. Now how do we amplify light using this? Consider a single photon entering a region with the atoms inside. This photon will pass by an atom with its electron in the upper level and cause it to emit a second photon traveling in the same direction by the process of stimulated emission. There are now two photons, each of which can cause stimulated emission in two more atoms to give four photons, and so on.

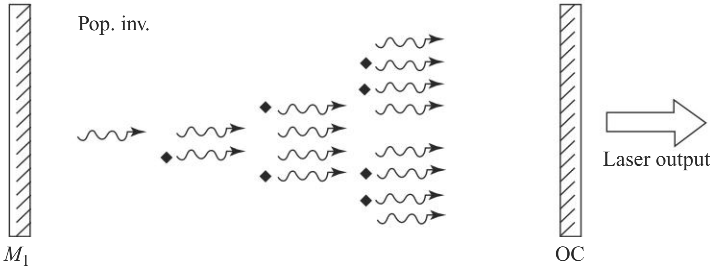


Fig. 5.3 *Process of population inversion*

Thus, we have amplification, which is also known as **gain.** The region containing the atoms is known as the **gain medium.** The final stage in a laser is to get this first photon to amplify. This is done by placing the gain medium between two mirrors. This forms what is known as a **laser cavity**.

A schematic diagram of a typical laser cavity is shown in Fig. 5.4. Mirrors M_1 and M_2 have intensity reflectivity very high, for mirror M_1 with almost ~ 99.5% or higher and M_2 which is also called Output Coupler (OC) with intensity reflectivity typically 98%-90%.

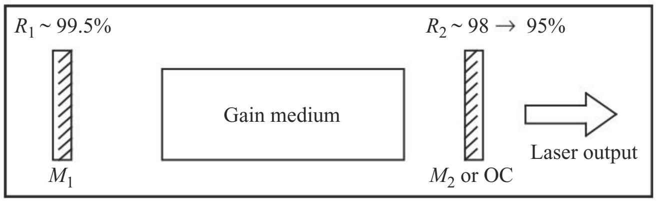


Fig. 5.4 *Typical laser cavity with mirrors on two sides*

To understand the process of light amplification, consider initially no light in the cavity. Therefore, spontaneous emission is the only possible process for the atoms to undergo transition and this duly occurs. As stated earlier, the spontaneously emitted photons may travel in any direction out of the gain medium, and most of them are lost from the cavity. However, out of the millions of photons emitted by the millions of atoms in any real medium, there is bound to be at least one photon which travels directly to one of the mirrors and is reflected back to the gain medium. This will act now as our first photon. As this photon passes through the gain medium, it produces stimulated emission of photons as described above and by the end of the gain medium there are, say, ten photons. In this process, the important part is that all these photons are traveling in the same direction as the first photon, and will be reflected back to the gain region by the other mirror. These ten photons now each produce stimulated emission of photons, and when they get out of the medium to the first mirror again, there are 100 which are reflected back to the gain medium again and are amplified to 1000, and so on. This process of multiplication of photons builds up in an avalanche-like manner, which is shown in Fig. 5.5. In this process, very high-amplitude electromagnetic waves are generated.

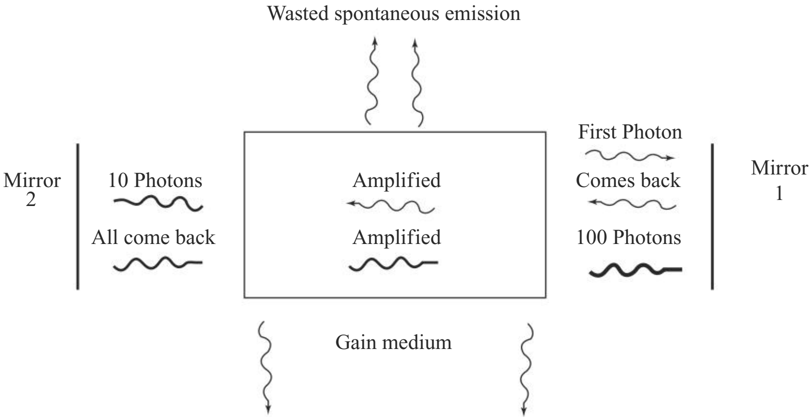


Fig. 5.5 *Complete process of ideal light amplification*

Thus, very rapidly we get a large amount of photons traveling back and forth in the cavity. In this idealised case no photons are lost from the steadily amplified beam so the photon number just goes on increasing. In any real laser cavity, some photons are always lost for many various reasons. One of these is quite deliberate. One of the mirrors is made to reflect only part of the light, and to allow the rest through. This is then the output beam of the laser and the 'leaky' mirror is referred to as the Output Coupler (OC). A steady state will be reached where the gain exactly replaces the photons lost from the cavity by the output coupler. There will be then a constant number of photons in the cavity at any given time. For example, for a laser with a gain of 1.12 (a realistic gain than 10 used earlier as the illustration) and an output coupler which reflects just 80% of the light we have the following figure.

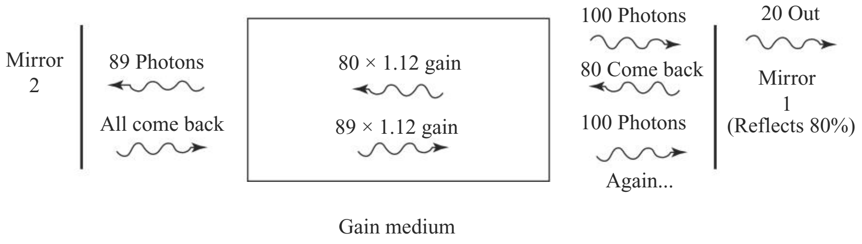


Fig. 5.6 *Real process of light amplification through stimulated emission of radiation at equilibrium*

Thus, the output beam has photons which are traveling in a fixed direction with a fixed wavelength (colour) defined by the energy levels of the electrons in the atoms of the gain medium.

Since the waves are traveling in the same direction, same phase and energy, constructive interference takes place and a coherent beam of light is produced. All the individual waves are now combined in phase that gives the resultant light as a single wave. The amplitude of the resultant wave is proportional to the square of the number of contributing atoms, i.e., N^2. The intensity ($I \quad N^2$) of the resultant beam is much higher than the incoherent beam because the number of contributing atoms is very large. Therefore, the intensity of light produced due to stimulated emission is much higher than the light produced due to spontaneous emission. The stimulated emission is the main process of the laser.

5.2.1 Relation Between Spontaneous and Stimulated Emission Probabilities (Einstein's Coefficients)

We know that an atom in an excited state emits radiation in two different ways: (i) spontaneous emissions, and (ii) stimulated or induced emissions. Stimulated emissions are produced by external effects but spontaneous emissions are not produced because of external effects. Stimulated emission arises because of the presence of external electromagnetic radiation of suitable frequency. The frequency of the emitted radiation is equal to that of the incident radiation in stimulated emission.

We shall now establish the relation between spontaneous and stimulated emission probabilities:

Let us consider an assembly of atoms in the cavity in thermal equilibrium at a temperature T containing radiation of frequency n and energy density (i.e., energy per unit volume of the cavity) $u(n)$. Let N_l and N_h be the number of atoms in the lower energy state l and higher energy state h respectively at any instant. The probability that the numbers of atoms in the state l absorb a photon and rise to the state h per unit time is

$$N_l P_{l\to h} = N_l B_{l\to h} u(v) \tag{5.2}$$

This is because the probability of transition from the lower state l to the higher state h, i.e., $P_{l\to h}$ is directly proportional to the energy density $u(n)$. Here, $B_{l\to h}$ is called the **Einstein's coefficient of absorption.**

The probability of emission is a sum of two parts, one which is independent of the radiation density (spontaneous emission of radiation) and the other proportional to it (stimulated emission of radiation). The probability that the number of atoms in the state h that drop to l, either spontaneously or under stimulation, emitting a photon per unit time is

$$N_h\,P_{h\to l} = N_h\,[A_{h\to l} + B_{h\to l}\,u(v)] \tag{5.3}$$

where,

$A_{h\to l}$ = Einstein's coefficient of spontaneous emission

$B_{h\to l}$ = Einstein's coefficient of stimulated (induced) emission

In thermal equilibrium, emission and absorption must balance. Thus,

$$N_l\,B_{h\to l}\,u(v) = N_h\,[A_{h\to l} + B_{h\to l}\,u(v]$$

or, $\quad u(v)\,[N_l\,B_{l\to h} - N_h\,B_{n\to l}\,u(v] = N_h\,A_{h\to l}$

$$u(v) = N_h\,A_{h\to l}/[N_l\,B_{l\to h} - N_h\,B_{n\to l}]$$

$$\therefore \qquad u(v) = \frac{A_{h\to l}/B_{h\to l}}{[(N_l/N_h)(B_{l\to h}/B_{h\to l}) - 1]} \tag{5.4}$$

Einstein proved thermodynamically that the probability of stimulated absorption must be equal to the probability of stimulated emission,

i.e.,

$$B_{l\to h} = B_{h\to l} \tag{5.5}$$

$$u(v) = \frac{A_{h\to l}/B_{h\to l}}{[(N_l/N_h) - 1]} \tag{5.6}$$

Now the equilibrium distribution of atoms among different energy states is given by Boltzmann's law which is applied to the system of gas molecules. This law is applied to a system of particles which are identical, distinguishable and having any spin. This is also applicable to solid crystalline materials because the positions of atoms are distinguishable. Now consider an assembly of atoms at an absolute temperature T in which the atoms are in different energy states. Let N_0 be the number of atoms per unit volume in ground state. Then the number of atoms N per unit volume in an excited state of energy E is given by the Boltzmann distributions law:

$$N = N_0\,e^{(-E/kT)} \tag{5.7}$$

where k = Boltzmann constant.

If N_h and N_l be the number of atoms per unit volume in the states of energies E_h and E_l then

$$\frac{N_l}{N_h} = e^{(E_h/E_l)/kT} = e^{hv/kT} \tag{5.8}$$

Consequently,

$$u(v) = \frac{A_{h\to l}/B_{h\to l}}{[e^{hv/kT} - 1]} \tag{5.9}$$

This is the formula for energy density of photons of frequency n in equilibrium with atoms in energy states l and h at a temperature T.

According to Planck's radiation formula,

$$u(v) = \frac{8\pi h v^3}{c^3}\,\frac{1}{[e^{hv/kT} - 1]} \tag{5.10}$$

Comparing Eqs (7) and (8), we have

$$\frac{A_{h\to l}}{B_{h\to l}} = \frac{8\pi h v^3}{c^3} \tag{5.11}$$

This is the required relation in the form of a ratio between the spontaneous emission and stimulated emission coefficients. The ratio is proportional to n^3. This shows that the probability of spontaneous emission rapidly increases with the difference in energy between the two states.

Here, it has been noted that if we had not assumed the presence of stimulated emission, we cannot derive the expression for $u(n)$, the energy density, which is similar to Planck's law. The stimulated emission was predicted by Einstein and was confirmed by quantum theory.

POPULATION INVERSION

5.3

Normally, it is a natural phenomenon that the number of atoms in the excited state is less than the number of atoms in the lower energy state, i.e., N_2, the population of higher energy level is less than the population N_1 of the lower energy level ($N_2 < N_1$). Now to make $N_2 > N_1$, i.e., the number of atoms N_2 more in the higher energy level than the number of atoms, N_1 in the lower energy level is known as population inversion or **inverted population.** This is also called the negative value of T. Here, the negative value of T is not a physical quantity but it is a convenient mathematical expression signifying the non-equilibrium state of the system. When the system has a number of states, then out of many states one of them may have a negative temperature with respect to the other state or states. So a system in which population inversion is achieved is called an **active system.** The method of raising the atoms from lower energy states to higher energy states is called **pumping.**

We know the processes of absorption, spontaneous emission and stimulated emission. The spontaneous emission process is independent of external factors and is not suitable to make population inversion. Einstein showed that the two processes (absorption and stimulated emission) are equally probable under normal circumstances. They can be acted upon from outside. Among the two processes, the relative number of atoms in the two states will decide which process predominates over the other. If $N_1 > N_2$, absorption predominates over emission. On the other hand, if $N_2 > N_1$, the stimulated emission dominates over the absorption. According to the Boltzmann distribution law

$$N_2 = N_1 \exp\left[-\frac{(E_2 - E_1)}{kT}\right] \tag{5.12}$$

For a positive value of the temperature T, if energies $E_2 > E_1$ then the number of atoms in the two states will be $N_2 < N_1$. This implies that the population of atoms in the lower energy state is more than in the higher energy state. Therefore, the photons incident on the atoms are very likely to be absorbed rather than to produce stimulated emission. Thus, the absorption process is predominant. To achieve the population inversion condition, i.e., stimulated emission to dominate over the absorption process, it is necessary to increase the population of the higher energy state. The situation is known as population inversion. The process of creating a population inversion is called pumping.

PUMPING AND ACTIVE SYSTEM

5.4

To attain population inversion, the atoms in the material must be continuously excited from lower energy levels to higher energy levels. To keep population in the higher energy levels, energy has to be supplied to the medium. This process is known as pumping. By heating the material, the state of population inversion cannot be achieved. To achieve this condition, one may think about heating the material. But by heating the material, the average energy of atoms is increased without the required situation, i.e., $N_2 > N_1$. Therefore, to achieve this condition, different methods are used.

5.4.1 Optical Pumping

In the method of optical pumping, optical energy in the form of photons is used to excite the atoms in the medium. For supplying optical energy, a discharge tube is employed to excite atoms from the ground state to a higher or excited state. The excited atom from the uppermost level spontaneously falls to the metastable state to create a state of population inversion with respect to the ground state. In doing so, some of the photons are released with a frequency and, hence, energy equal to the difference between higher state and metastable state. Obviously, the frequency of pumping photons must be higher than that of emitted photons, so that the atoms can be sent from the lower level to the upper level. This reduces the efficiency of the laser.

The pumping level should accommodate a range of energies, so it should be wide enough. If it has limited energy then it can be employed for only a specific frequency which limits the alternative of the source and a large amount of energy would be wasted. Hence, light sources emitting a range of wavelength are used to excite the atoms. Optical pumping is suitable for any medium which is transparent to light.

5.4.2 Electrical Pumping

Electrical pumping is used for some mediums which can conduct electricity without affecting laser activity. This method is suitable for gases. In a gaseous medium, the gas is ionised by a pulse of high electric voltage. The current flowing through the gas provides necessary energy to excite the atoms to the uppermost level. From this level, the atoms jump to the metastable state, creating a state of population inversion with respect to the ground or lower state.

Electrical pumping is also employed in case of a semiconductor laser medium. In this medium, instead of atoms, the charge carriers (electrons and holes) are excited and a state of population inversion is achieved in the junction region. In the junction area, electrons combining with holes develop laser. Therefore, in a semiconductor laser, electrical energy is directly converted to light energy.

In fact, energy (electrical, optical, chemical, whatever) would be pumped in to sustain the inversion, and a beam of light would be extracted. Such a device for *light amplification by stimulated emission of radiation* is called a laser. Figure 5.7 and 5.8 shows the principle of the **LASER**.

METASTABLE STATE

5.5

The atoms are energised to a higher energy level by absorbing energy from external sources like optical, electrical, etc. They remain in an excited state for a very short period of 10^{-8} second; after that, they spontaneously release their excess energy. For stimulated emission to occur, the atoms should remain excited long enough, typically for 10^{-4} second, as atoms are continuously lifted up to the excited state by pumping and a number of them rapidly undergo spontaneous transition to the lower level. Therefore, the state of Population Inversion (PI) is difficult to be achieved. To achieve that state (PI), the excited atoms should remain in the uppermost level till the condition $N_2 > N_1$ gets satisfied. So a metastable state is such a long-lived upper energy level from where the excited atoms do not return to the lower level instantly. In the metastable state, the excited atoms remain for an appreciable time for 10^{-6} s to 10^{-3} s, which is a very long time for an atom in the excited state to create a state of population inversion. The population of atoms in the metastable state is greater than that in the lower state, which leads to population inversion.

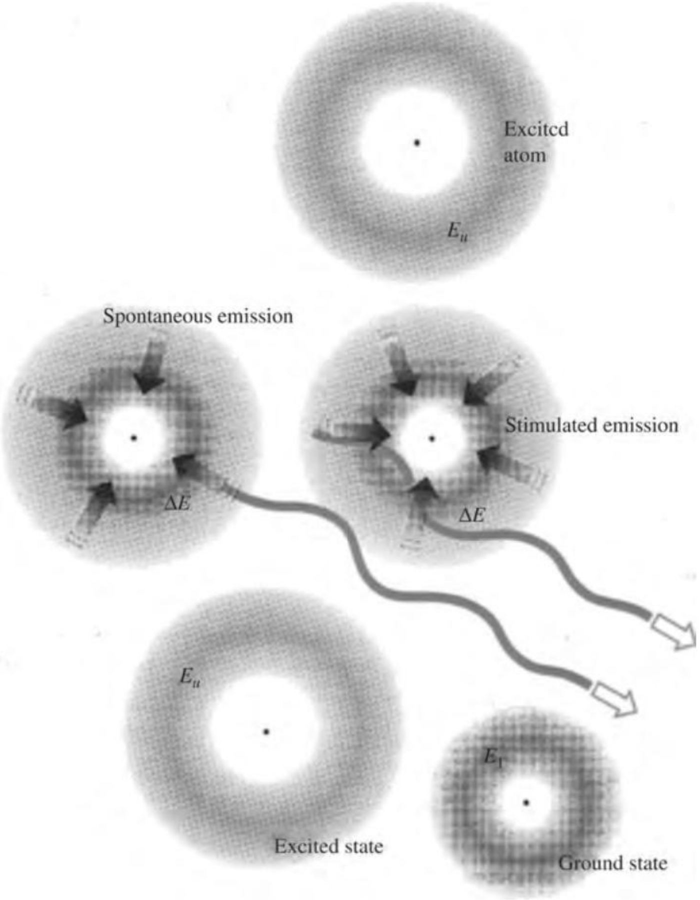


Fig. 5.7 *Figure depicting population inversion: five atoms in a group where one is in a ground state and remaining four are in an excited state. One atom (on the left) spontaneously returns to the groud state by emitting a photon, which causes a nearby atom to be stimulated into emitting an inphase photon og iys own.*

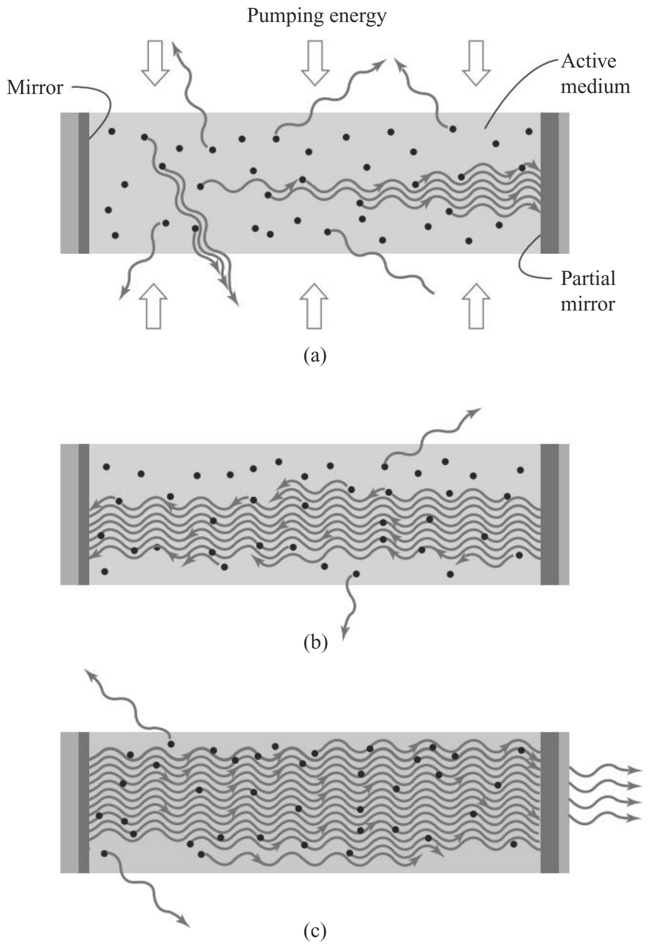
MAIN COMPONENTS OF LASER

5.6

There are three main components of a laser system: (1) active medium, (2) optical pumping system, and (3) optical resonator. Let us understand the function of each one in producing required laser light.

5.6.1 Active Medium

An active medium is defined as a medium which, when excited, attains the state of population inversion and causes light amplification. This implies that the whole process is performed in this medium and light amplification is achieved in this region. The active medium may be of any type—solid, liquid, or gas. In the active medium, only some atoms residing in a particular energy state are responsible for stimulated emission and light amplification. These atoms are called **active centres** and the remaining atoms support the active centres. The medium containing the active centres is called the active medium.



(a)

(b)

(c)

Fig. 5.8 *A schematic presentation of the laser. (a) Pumping of an active medium, initiation of the population inversion. (b) The reflected photons transverse the cavity, stimulating other atoms to emit photons and the beam builds up. (c) Almost all the medium contributes to the stimulating emission and laser action takes place.*

5.6.2 Optical Pumping System

Two basic mechanisms can produce population inversion in a laser, either by creating an excess of atoms or molecules in a higher energy state, or by reducing the population of a lower energy state.

To produce a population inversion in a laser medium, energy should be added to the system in order to excite the atoms or molecules into higher energy levels. If we add thermal energy to the medium, it is not sufficient (under thermodynamic equilibrium) to produce a population inversion, because heat only increases the average energy of the population, but does not increase the number of atoms or molecules in the excited state relative to that in the lower state. The ratio of the number of atoms at two energy levels (1 and 2) under thermodynamic equilibrium is given by Eq. (5.12) and can be written in the following way:

$$N_2 / N_1 = \exp[-(E_2 - E_1)/kT]$$

where N_1 and N_2 are the number of atoms in level 1 and level 2, with the energies E_1 and E_2 respectively, k is the Boltzmann constant, and T is the temperature in Kelvin. From the equation, at thermodynamic equilibrium, N_2 can be greater than N_1 only if the temperature is a negative number, referred to a population inversion as a **negative temperature**.

To acquire the required population inversion for laser activity, atoms or molecules must be selectively excited to specific energy levels. To attain the condition for population inversion, the atoms have to be raised to an excited state continuously. This requires energy to be supplied to the system. In optical pumping, a light-source discharge tube is used. In the electric discharge pumping method, ionisation in the medium occurs due to an electric field and the atom goes to the excited state. This method is used in gas lasers.

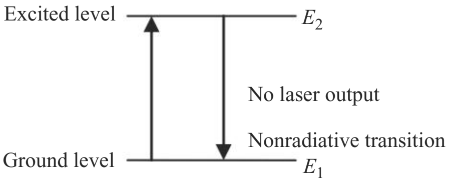


Fig. 5.9 *Two-level pumping scheme*

As the atoms have a large number of energy levels, two, three or four energy levels are eligible for the pumping process. The transition between two levels due to which stimulated emission occurs is called **lasing transition.** The uppermost level is called the **pumping level.** The two-level pumping schemes cannot lead to laser action. For population inversion, the population in the uppermost level must be greater than that of in the lower state. It is possible only when the upper level will populate faster than its depopulation. For this, the lifetime of spontaneous emission should be longer, i.e., the lifetime at upper level E_2 should be longer. The light source must be highly monochromatic. However, practically, more number of atoms cannot be excited to the upper level. The pumping excites the atoms from the ground state and induces the atoms for transition from upper to lower level. Hence, population and depopulation processes take place simultaneously. Therefore, in two-level schemes, the state of population inversion cannot be reached. The two level-pumping scheme is shown in Fig. 5.9.

The widely used pumping schemes are (a) three-level pumping scheme, and (b) four-level pumping scheme.

For most lasers, light and electricity are the excitation mechanisms of choice. Either light or electrons can furnish the energy necessary to excite atoms or molecules to selected higher energy levels, and the transfer of energy is not required to directly elevate electrons to a specific upper level of the laser transition. Some methods can be rather complex, but these often produce better-performing lasers. One often utilised approach excites an atom or molecule to a higher energy level than required, there after it drops to the upper laser level. Indirect excitation can be applied to excite atoms in a surrounding gas mixture, which then transfer their energy to the atoms or molecules responsible for producing the laser action.

As previously discussed, the amount of time spent by an atom or molecule in an excited state is critical in determining whether it will be stimulated to produce emission and participate in a cascade of photons, or lose its energy through spontaneous emission. Excited states commonly have lifetimes of only a few hundred nanoseconds before they release their energy by spontaneous emission, a period that is too short to likely undergo stimulation by another photon. A critical requirement for laser action to successfully take place is to have a longer-lived state that is suitable for the upper energy level. Such states do exist for certain materials, and are referred to as **metastable** states (see Fig. 5.10). The average lifetime before spontaneous emission occurs for a metastable state is of the order of a microsecond to a millisecond, quite a lengthy and sufficient period of time on the atomic timescale. With this much long lifetimes, excited atoms and molecules can develop significant amounts of stimulated emission. Laser action is only possible if the population builds up faster than it decays in the upper energy level, maintaining a population larger than that of the lower level. For laser applications, the atoms and molecules having longer spontaneous emission lifetimes are more suitable.

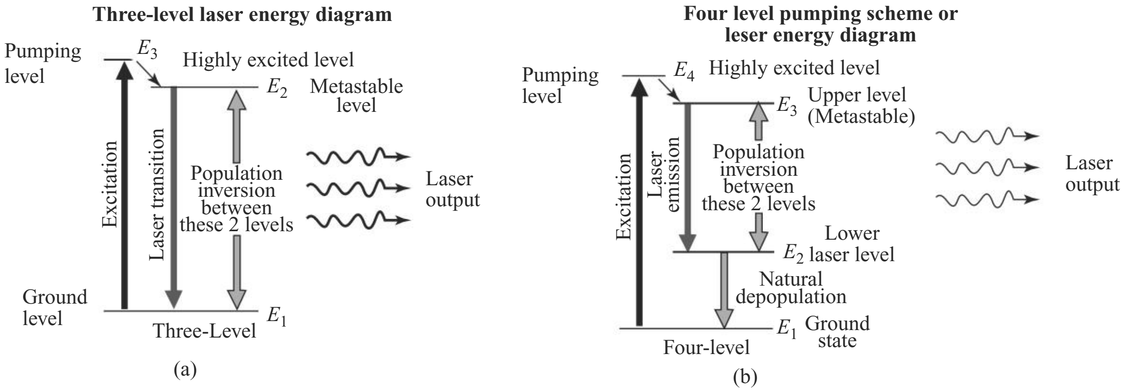


Fig. 5.10 *Three and four-level pumping scheme for laser production*

The simplest functional energy-level structure for laser operation is a three-level system, which is illustrated in Fig. 5.10(a). In this system, the ground state is the lower energy level E_1, and a population inversion is created between this level and a higher-energy metastable state with the energy E_2. Most of the atoms or molecules are initially excited to a short-lived high-energy state E_3 that is higher than the metastable level E_2 (ground level E_1 > metastable level E_2 > excited higher level E_3). From this state, they quickly decay to the intermediate metastable level E_2, which has a much longer lifetime (~order few *m*s to milliseconds) than the higher energy state. Because each atom's residence time in the metastable state is relatively long, the population tends to increase and leads to a population inversion between the metastable state E_2 and the lower ground state E_1 (which is continuously being depopulated to the highest level). Stimulated emission results from the fact that more atoms are available in the upper excited (metastable) state E_2 than in the lower state E_1 where absorption of light would most likely occur.

Although the three-level laser system works for all practical purposes, as illustrated by the first laser, a couple of problems restricts the effectiveness of this approach. The main problem occurs because the lower laser level is the ground level, which is the normal occupied state for most atoms or molecules. In order to produce the population inversion, a majority of ground-state electrons must be elevated to the highly excited energy level, which requires a significant input of external energy. In addition, the population inversion is difficult to maintain for an appreciable time, and therefore, three-level lasers must be operated in pulsed mode rather than continuously.

Lasers utilising four or more energy levels to avoid some of the problems mentioned above, and therefore are more commonly used. Figure 5.10(b) illustrates a four-level scheme of operation. The energy level structure is similar to that in the three-level system (ground level E_1 > lower laser level E_2 > metastable level or upper laser level E_3 > excited higher level E_4), except that after the atoms drop from the highest level E_4 to the metastable upper state E_3, in this case, they do not drop all the way to the ground state E_1, in a single step. Because the population inversion is not created between the ground state E_1 and the upper level E_3 (rather between E_3 and E_2), the number of atoms or molecules that must be elevated is drastically reduced in this model. If, for example, in a typical four-level laser system, if only 1% or 2% of the atoms or molecules reside in the lower laser level E_2 (which is above the ground state) then exciting only 2% to 4% of the total to the higher level E_3 will achieve the required population inversion. Another advantage of separating the lower laser level E_2 from the ground level E_1 is that the lower level atoms will naturally fall to the ground state. If the lower laser level has a lifetime that is much shorter than the upper level, atoms will decay to the ground level at a rate sufficient to avoid accumulation in the lower laser level. Many of the lasers designed under these restraints can be operated in a continuous mode to produce an uninterrupted beam.

5.6.3 Optical Resonator

An optical resonator plays a major role in the generation of the laser output, with high directionality to the laser beam and producing gain in the active medium to overcome the losses due to straying away of photons from the laser medium, diffraction losses due to definite sizes of the mirrors, radiation losses inside the active medium due to absorption and scattering, etc. In order to maintain laser action, one has to confine the laser medium and the pumping mechanism in a special way that should encourage stimulated emission rather than spontaneous emission. In practice, photons need to be confined in the system to allow the number of photons created by stimulated emission to exceed all other mechanisms. This is achieved by bounding the laser medium between two mirrors as shown in Fig. 5.11.

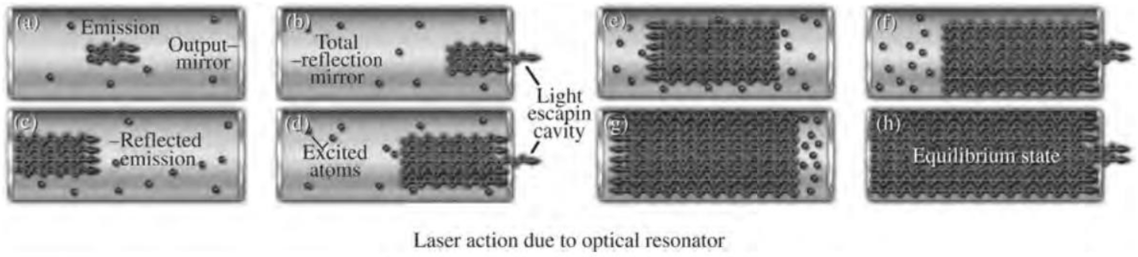


Fig. 5.11 *Optical resonator*

On one end of the active medium is the high reflectance mirror (100% reflecting) or the rear mirror, and on the other end is the partially reflecting mirror or the output coupler. The laser emanates from the output coupler, as it is partially transmissive. Stimulated photons can bounce back and forward along the cavity, creating more stimulated emission as they go. **In the process, any photons which are either not of the correct frequency or do not travel along the optical axis are lost.**

TYPES OF LASERS

5.7

There are many types of lasers available for research, medical, industrial, and commercial uses. Lasers are often described by the kind of lasing medium they use, e.g., solid state, gas, excimer, dye, or semiconductor.

(1) **Solid State** lasers have lasing material distributed in a solid matrix, e.g., the ruby or neodymium–YAG (yttrium aluminum garnet) lasers. The neodymium–YAG laser emits infrared light at 1.064 *m*m. (2) **Gas** lasers (helium and helium–neon, He–Ne, are the most common gas lasers) have a primary output of a visible red light. CO_2 lasers emit energy in the far-infrared, 10.6 *m*m and are used for cutting hard materials. (3) **Excimer** lasers (the name is derived from the terms *excited* and *dimers*) use reactive gases such as chlorine and fluorine mixed with inert gases such as argon, krypton, or xenon. When electrically stimulated, a pseudomolecule or dimer is produced and when lased, they produce light in the ultraviolet range. (4) **Dye lasers** use complex organic dyes like rhodamine 6G in liquid solution or suspension as lasing media. They are tunable over a broad range of wavelengths. (5) **Semiconductor** lasers, sometimes called diode lasers, are

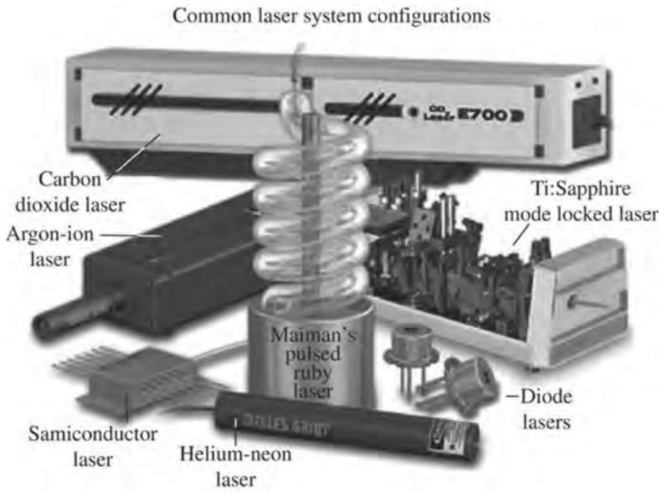


Fig. 5.12 *Some commonly available laser systems*

not solid-state lasers. These electronic devices are generally very small and use low power. They may be built into larger arrays, e.g., the writing source in some laser printers or compact disc players.

THE RUBY LASER

5.8

The first laser was created in 1961 by Theodore Maiman (1927) at the Hughes Research Laboratories. He used a rod of synthetic ruby as the lasing medium. The crystalline structure of ruby is similar to that of corundum, i.e., a crystal of aluminum oxide (Al_2O_3), in which a small part of atoms of aluminum (about 0.05 %) is replaced with ions of Cr^{+3}. The energy levels of Cr^{3+} that are important for the operation of a ruby laser are shown in Fig. 5.14. The ruby rod is illuminated by an intense impulse of light, which is generated by a helical xenon discharge lamp as shown in Fig. 5.13. The ends of the ruby rod are highly polished and silvered to serve as laser mirrors. The impulse of light creates the required population inversion of electrons in the ruby rod and due to the presence of mirrors, the laser generation is excited. The duration of the laser impulse is a little bit shorter than the pump impulse of the flash lamp.

When the flash tube is switched on, there is an intense burst of light lasting for a few milliseconds. Absorption of this light energy excites many of the chromium ions to the bands of energy levels indicated by E_3 and $E_3$9 in Fig. 5.14. During the relaxation of chromium ions, they give up the energy to the crystal in nonradiative transitions as they drop down to a pair of metastable states labeled E_2 in Fig. 5.14. These metastable states are about 1.79 eV above the ground state. The intense flash provides more number of photons of sufficient energy and thereby more atoms make the transition to the states E_2 via the process described as above than remain in the ground state. As a result, the populations of the ground state and the metastable states become inverted and lasing takes place if a cavity is provided. The cavity is formed by the silvered ends of the ruby crystals. One of the mirrors reflects almost all photons of light having a wavelength of 694.3 nm and the other end is only partially reflecting (99%) in order to leak the laser emission.

We should stress that the ruby laser works in the **pulsed mode** due to the high pump energy. Today many four-level lasers have been developed. Since they need far less power to create and to sustain the population inversion, they can work in the **continuous wave (cw) mode**.

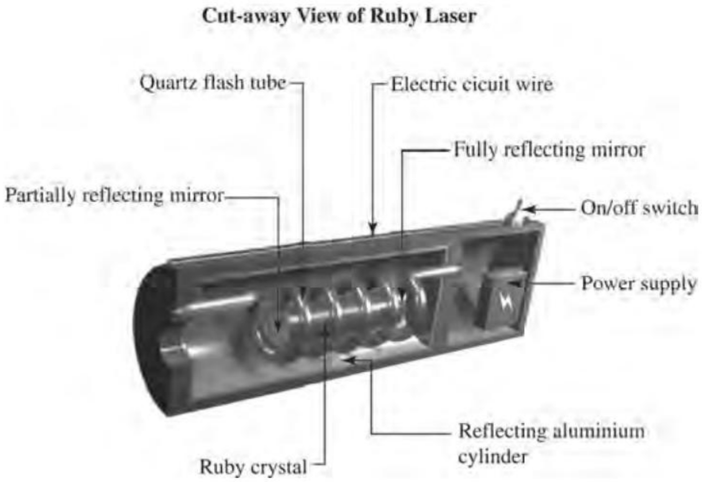


Fig. 5.13 *Ruby laser*

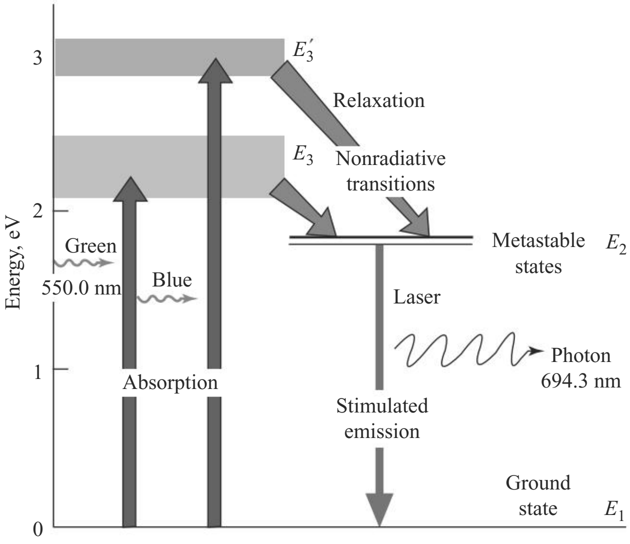


Fig. 5.14 *Energy levels of the ruby laser*

ND:YAG LASER

5.9 Neodymium doped Yttrium Aluminum Garnet (Nd:YAG) lasers were first invented by J E Geusic, H M Markos and L G Van Uiteit in 1964 at Bell Labs. It was developed by the same technology as the Ruby Laser. The gain medium is the YAG crystal which is doped with around 1% neodymium by weight. When they were first invented, the lasing medium for YAG lasers were stimulated optically by flash lamps. In the world today, YAG lasers have numerous applications in the medical and scientific field for processes such as Lasik surgery and laser spectroscopy.

Nd:YAG laser is one of the prominent solid state lasers. Solid state lasers use ions suspended in a crystalline matrix to produce laser light. The ions or *dopants* provide the electrons for excitation, while the crystalline matrix propagates the energy between ions. The two main classes of dopants in the laser medium are chromium (Cr^{3+}) for ruby lasers and neodymium (Nd^{3+}) for *Nd:YAG and Nd:glass* lasers. The Nd:YAG and Nd:glass lasers are in general very similar to each other in structure and lasing action. Excitation is achieved by any one of krypton or xenon flash lamps, and an output wavelength of 1.06 μm in the near infrared region of the spectrum is obtained. The Nd:glass laser uses a glass host material for the neodymium ions. The glass rods have the advantage that they can be grown to larger sizes more economically than the YAG crystals, but glass has a lower thermal conductivity which limits the pulsed operation of the Nd:glass laser. Therefore, Nd:glass lasers are used in applications which require low pulse repetition rates and high pulse energies (up to 100 joules per pulse). In general, the pulse operation of the neodymium lasers make them desirable for hole piercing and deep keyhole welding applications.

The host material in Nd:YAG lasers is a complex crystal of Yttrium-Aluminum-Garnet (YAG) with the chemical composition $Y_3Al_5O_{12}$. Figure 5.15 shows the schematic of a Nd:YAG Laser. The YAG crystal has a relatively high thermal conductivity, which improves thermal dissipation in the laser cavity, so continuous wave operation up to a few hundred watts is possible. When operated in a pulsed mode, high pulsing rates can be achieved, and average powers of up to 1 kW are available. The actual lasant is the Nd^{+3} ions which have been doped into a YAG crystal. Pure $Y_3Al_5O_{12}$ is a colorless, optically isotropic crystal that possesses a cubic structure characteristic of garnets. In Nd :YAG about 1% of Y^{3+} is substituted by Nd^{3+}. The radii of the two rare earth ions differ by about 3%. Therefore, with the addition of large amounts of neodymium, strained crystals are obtained—indicating that either the solubility limit of neodymium is exceeded or that the lattice of YAG is seriously distorted by the inclusion of neodymium. When doped with approximately 1% Nd, the crystal takes on a light blue color.

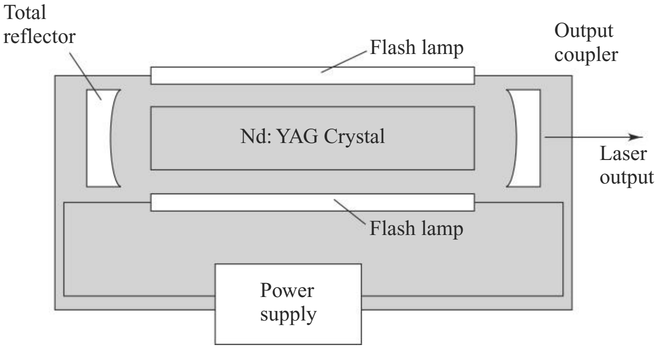


Fig. 5.15 *Schematic diagram of Nd:YAG Laser*

As shown in Fig. 5.15 and 5.16, the arrangement of laser rod Nd:YAG, twin elliptical reflector, excitation flashlamp or arc-lamp or laser diode for pumping purpose, high reflective dielectric coating end mirror on the left and partial reflector outcoupler on the right side is made such that on switching on the power, the flashlamp emits radiation and gets focused on the laser rod. The YAG crystal rods are typically 10 cm in length and 9 cm in diameter. The two ends are polished and silvered and constitute the optical resonator.

Generally, a cooling system is required for operation of ND:YAG lasers. With an efficiency of about 3%, a typical ND:YAG produces thirty times as much waste heat as laser output; this heat must be removed in

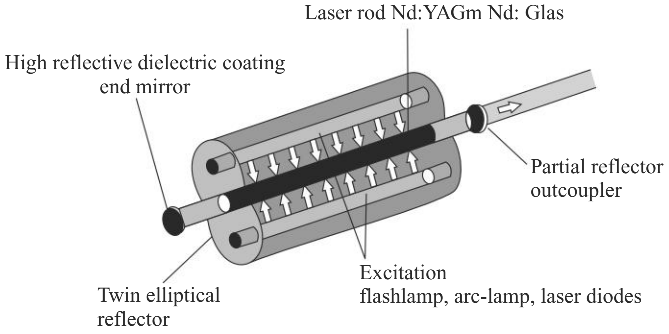


Fig. 5.16 *Showing the mechanism of lasing action in Nd:YAG laser*

order to ensure proper laser operation. Flooding the optical compartment with water can remove waste heat; however, the water absorbs a significant amount of the flashlamp energy, and the turbulence of the water can produce optical distortion and imaging problems. These problems can be overcome by flowing water over the outside of the optical cavity and by encasing the lasing rod and flashlamp with transparent cooling jackets. Furthermore, the flashlamp electrode assemblies should be cooled to prevent thermal loading, and deionized water should be used to prevent any electrical conduction through the water.

The output characteristics of a Nd:YAG laser can be altered by varying the pumping discharge waveform. For a high-pulse energy in a pulse time of one millisecond, a long lamp pulse is used to obtain a six to seven Joule pulse composed of many spiked oscillations within the pulse. Such a laser can operate at several hundred cycles per second and is useful in laser cuttings. Laser beam pulse frequency and shape can be tailored by using *Q switching,* where a shutter moves rapidly in and out of the path of the beam. In this manner, beam output is interrupted until a high level of population inversion and energy storage is achieved in the resonator. If the optical cavity is switched from no reflections (low Q) to near total reflection (high Q), the cycle can be optimized to build up the maximum population inversion before the pulse is generated (switched from low to high Q). This results in a beam pulse with a high energy (up to one joule) and a short pulse period (down to 10 ns), which is especially useful for deep hole drilling applications.

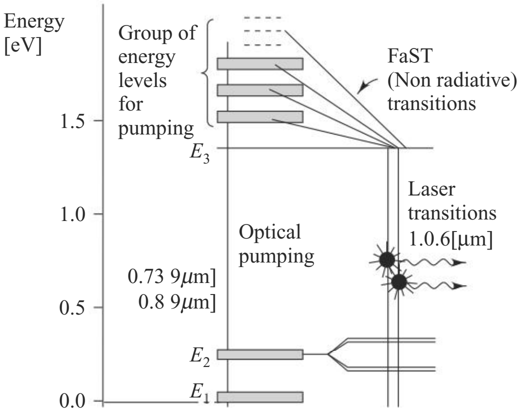


Fig. 5.17 *Simplified energy levels of Nd:YAG*

As can be seen from the energy level diagram in Fig. 5.17, Nd lasers are **four level lasers**. Nd ions have **two absorption bands**, and excitation is done by **optical pumping**, either by flash lamps for pulsed lasers, or by arc lamps for continuous wave lasers.

The flash lamp is switched on and the optical pumping excites the Nd ions from the ground state E_1 to higher group of energy levels E_4 by absorbing optical energy corresponds to wavelengths 0.73 μm and 0.8 μm.

From these excited energy levels, the Nd ions are transferring into the upper laser level indicated by E_3 by a non-radiative transition. Now there is a **stimulated emission** of photons from the upper laser level E_3 to the lower laser level E_2, and the wavelengths of the emitted photons are 1.06 μm. Then there is a transition from the lower laser level E_2 to a ground level E_1, which will be a non-radiative transition.

Applications

ND:YAG lasers are widely used in a variety of applications, including

1. Material processing, such as drilling and welding, cutting
2. Medical applications such as cataract surgery, gall bladder surgery
3. Laser ranging, in particular for laser range-finders and target designators used in a military context
4. Scientific applications like the study of inertial confinement fusion

GAS LASER

5.10

Gas lasers can be further divided into neutral atom, ion and molecular lasers, whose lasing mediums are neutral atoms, ions or gas molecules respectively.

5.10.1 The Helium–Neon Laser

Figure 5.18 shows a schematic sketch of an He–Ne laser. It consists of a gas tube containing 15% helium gas and 85% neon gas. A totally reflecting flat mirror is mounted at one end of the gas tube and a partially reflecting concave mirror is placed at the other end. The concave mirror focuses parallel light at the flat mirror and also acts as a lens that transmits part of the light so that it emerges as a parallel beam. The helium–neon (He–Ne) laser is a kind of neutral atom gas laser, the common wavelength of an He–Ne laser is 632.8 nm. It is tunable from infrared to various visible light frequencies. Pumping is done by a dc electrical discharge in the low-pressure discharge tube. First, the He atom is excited. Because the Ne atom has an energy level very near to an energy level of He, through kinetic interaction, energy is readily transferred from He to Ne; and the Ne atom emits the desired laser light. The typical power of an He–Ne laser is below 50 mW, hence it is widely used in holography, scanning, measurement, optical fiber communication, etc. It is the most popular visible light laser.

Population inversion is achieved somewhat differently in the continuous wave mode (cw) helium–neon (He–Ne) gas laser. The energy levels of helium and neon, which are important for operation of the laser, are shown in Fig. 5.19.

Helium has an excited energy state $E_{2,\mathrm{He}}$ at 20.61 eV above the ground state $E_{1,\mathrm{He}}$. Only helium atoms are excited to state $E_{2,\mathrm{He}}$ by an electric discharge. Neon has an excited state $E_{3,\mathrm{Ne}}$ that is 20.66 eV above its ground state. This is very close to the energy state of He, i.e., $E_{2,\mathrm{He}}$, just 0.05 eV above the first excited state of helium. The neon atoms are excited to the state $E_{3,\mathrm{Ne}}$ by collision with the excited helium atoms. The kinetic

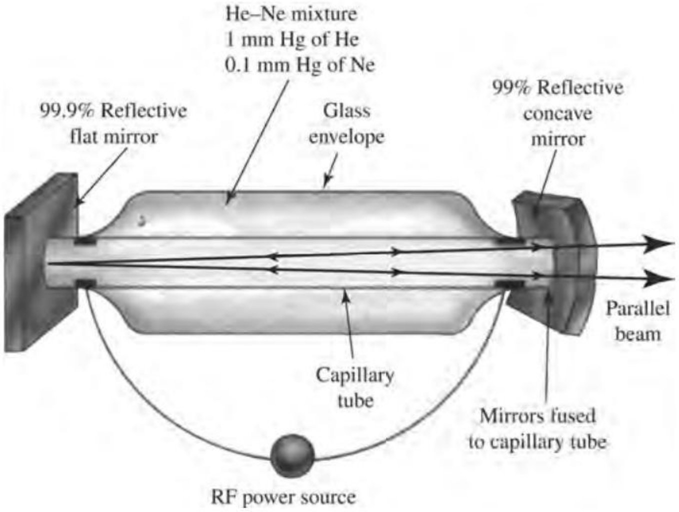


Fig. 5.18 *Schematic drawing of an He-Ne laser*

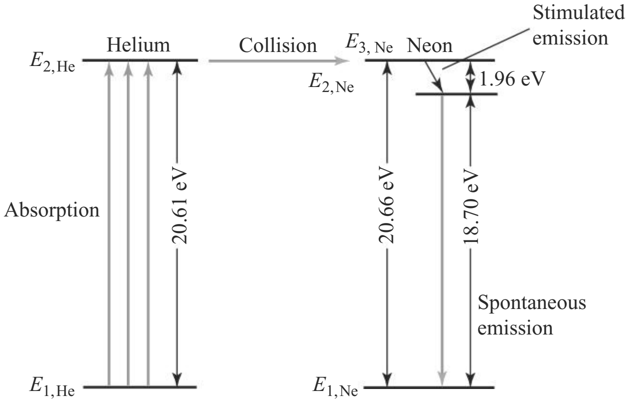


Fig. 5.19 *Energy levels of the He-Ne laser*

energy of the helium atoms provides the extra 0.05 eV of energy needed to excite the neon atoms. There is lower excited state of neon $E_{2,\text{Ne}}$ at 18.70 eV above its ground level and 1.96 eV below the excited state $E_{3,\text{Ne}}$. Since state $E_{2,\text{Ne}}$ is normally unoccupied, population inversion between states $E_{3,\text{Ne}}$ and $E_{2,\text{Ne}}$ is obtained immediately. The stimulated emission occurring between these states results in photons of energy 1.96 eV (632.8 nm), which produces a bright red light. After stimulated emission, the atoms in the state $E_{2,\text{Ne}}$ decay to the ground state by spontaneous emission. Note that there are **four energy levels** involved in the **He–Ne laser**.

5.10.2 CO_2 Laser

Lasing action in a CO_2 molecule was first presented by C Patel in 1964. He transmitted an electric discharge pulse through pure CO_2 gas in a laser tube, and got a small laser output. CO_2 is the gas in which the lasing process occurs, but other gas additives to the laser tube improve the total efficiency of the laser. The standard CO_2 laser includes in the active medium a mixture of CO_2 with N_2 and He. The optimal proportion of these 3 gases in the mixture depends on the laser system and the excitation mechanism. In general, for a continuous wave laser, the proportions are CO_2:N_2: He::1:1:8

CO_2 is a linear molecule, and the three atoms are situated on a straight line with the carbon atom in the middle. In Fig. 5.20, the three vibrational modes of CO_2 molecule are illustrated: symmetric stretch mode (v_1), bending mode (v_2), and asymmetric stretch mode (v_3).

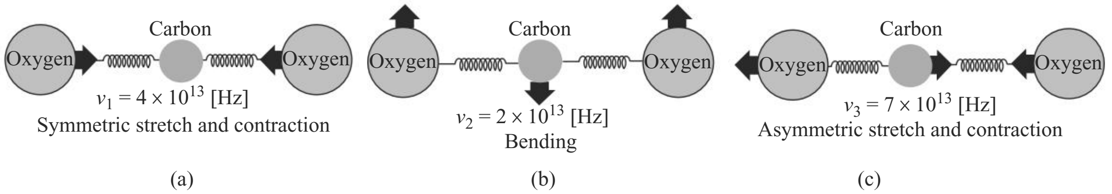


Fig. 5.20 *Different vibrational modes of CO_2*

Lasing Transitions in a CO_2 Laser Lasing transitions in a CO_2 laser occur when the molecule is going from higher energy level of the asymmetric mode into one of the other two, as can be seen in Fig. 5.18. The transition to the symmetric stretching mode corresponds to the wavelength of 10.6 *m*m. and the transition to the bending mode corresponds to the wavelength of 9.6 *m*m.

Each of the vibrational energy levels is subdivided into many rotational levels. Transitions can occur between vibrational energy levels with different rotational levels, so there are many lasing lines around the main vibrational transitions.

CO_2 Laser Operation Electric discharge is created in the laser tube. This discharge is produced by dc excitation. The energy of the accelerated electrons is transferred by collisions to the nitrogen molecules and to the CO_2 molecules. Here, nitrogen molecules help in the process of the excitation of the CO_2 molecules. The first vibrational energy level of the nitrogen molecule is very similar to the asymmetric stretching mode of the CO_2 molecule (see Fig. 5.21). Therefore, energy can be easily transferred from the excited nitrogen molecules to the CO_2 molecules. The purpose of adding Helium molecules in the gas mixture is to evacuate the lower laser energy level so that population inversion is maintained and to stabilise the electrical discharge by taking heat away from the lasing area. Gas pressure inside the CO_2 laser tube is 5–30 [Torr], of which 10% is CO_2 gas, 10% is N_2 and the 80% is He.

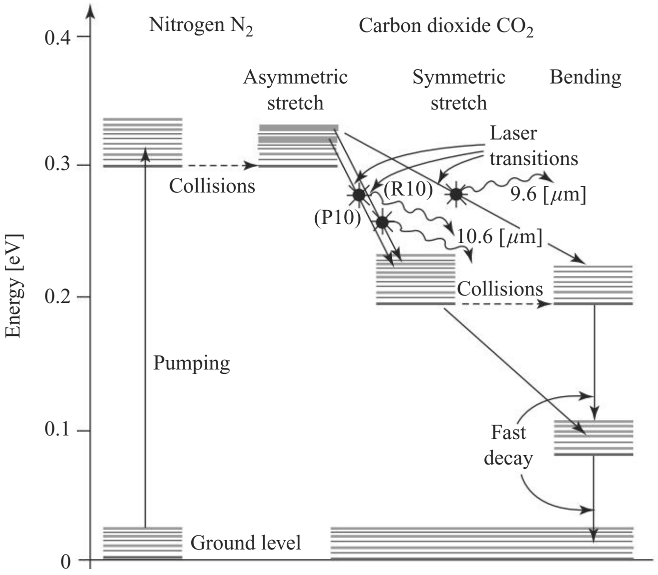


Fig. 5.21 *Energy-level diagram of CO_2 laser*

5.10.3 Types of CO_2 Lasers

There are different types of CO_2 lasers, based on the same physical principles. The difference between them is in their structure, excitation mechanism, and the output radiation. A few CO_2 lasers are described below.

Flowing CO_2 Gas Lasers In these lasers a fresh gas mixture flows continuously through the laser tube while lasing lasts. Flowing gas is used when the maximum power is needed out of the CO_2 laser. The gas flows along the tube and is released out into the atmosphere (since it is non-poisonous). These lasers are very simple, and require only small gas purity. Hundreds of watts can be achieved at the output of these lasers.

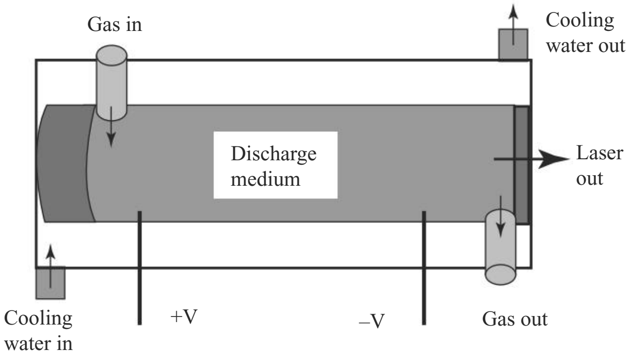


Fig. 5.22 *CO_2 laser tube*

Sealed-off CO_2 Laser In this type, the gas laser is filled with the appropriate mixture of gasses and sealed. High electric voltage is applied to electrodes at both ends of the gas tube. The accelerated electrons excite the gas molecules. The problem with sealed-off lasers is the dissociation of the CO_2 molecules into CO and oxygen with time. To reduce this effect, a catalysing agent is added to the gas mixture. This catalysing agent reverses the dissociation reaction and restores the CO_2 molecule which is required for lasing. The output power achievable is less than 200 watts. For higher output power, it is necessary to take away the heat generated inside the laser, and a flowing gas is needed. Sometimes, a gas reservoir is added to the sealed-off tube to allow some refreshment of the lasing gasses from the reservoir to the tube. The new class of sealed CO_2 lasers is made from metal with no glass tube. These lasers are not excited by direct high voltage that creates the electric discharge, but by a Radio Frequency (RF) voltage.

CO_2 Laser Inside a Waveguide When the laser tube diameter is reduced to a size of about 1 millimetre, a waveguide is made. The radiation inside the small diameter tube is confined to move along the tube, with low loss. Using ceramic tubes, very small CO_2 lasers can be made. An output of up to 50 watts of continuous wave radiation can be produced.

Transverse Flow CO_2 Gas Lasers When the gas flow inside the laser is perpendicular to the laser axis, it is possible to have a much larger flow since the distance is very short. Since cooling by gas flow is very efficient, it is possible to get very high power output out of these lasers. Both the gas flow and the electric discharge in these lasers are along the width of the laser. Thus, the distance between the electrodes is short, so electric discharge can be achieved even for gas at very high pressure (up to a few atmospheres). Transverse flow is used for very high-power CO_2 lasers.

5.10.4 Properties of CO_2 Lasers

1. Give high output power. Some commercial CO_2 lasers produce more than 10 kW continuously.
2. Output spectrum of laser is in the Infra-Red (IR) spectrum, i.e., between 9–11 μm.
3. Very high efficiency (up to 30%) is achievable.
4. It can operate both continuously or in a pulsed way.
5. Average output power is 75 W for slow flow of gas, and up to few hundred watts for fast gas flow.
6. Very simple to operate, and the gasses are non-toxic.

Summary of CO_2 Lasers According to Groups

- Gas laser
- Emit in the Infra-Red (IR) spectrum (l = 9–11 mm)
- Electrical excitation
- Continuous wave, although pulsed operation is possible
- Four-level laser

SEMICONDUCTOR LASER

5.11

All laser diodes are built from semiconductor materials, and all show electric properties which are characteristics of electrical diodes. For this reason, the diode lasers have other names such as *semiconductor lasers*—according to the composed materials,

Junction lasers—Since they are composed of a *p-n* junction,

Injection lasers—Since the electrons are injected into the junction by the applied voltage.

The diode laser was invented independently in 1962, at three different research laboratories in the US. The researchers succeeded in getting a coherent electromagnetic radiation from a forward biased diode (*p-n* junction) made from the semiconductor GaAs.

5.11.1 Laser Diode Construction

The basic structure of the layers of the simplest laser diode is shown in Fig. 5.23.

These layers of semiconductor materials are arranged such that at the *p-n* junction an active region is created, in which photons are created by the recombination process. On the top and bottom layers, a layer of metal allows connecting external voltage to the laser. The voltage is applied to metal contacts above and below the semiconductor layers. The sides of the crystalline semiconductors are cut to serve as mirrors at the end of the optical cavity.

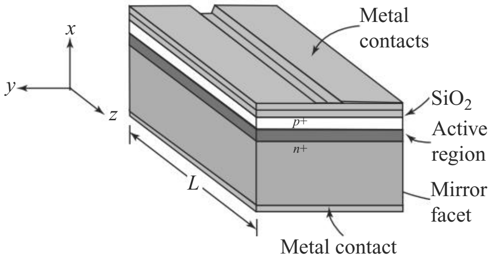


Fig. 5.23 *Basic structure of a laser diode*

5.11.2 Output Laser Radiation from a Diode Laser

Figure 5.24 depicts the shape of the electromagnetic laser radiation for a simple laser diode built of layers.

The radiation comes out of a rectangular shape of a very thin active layer, and spreads at different angles in two directions.

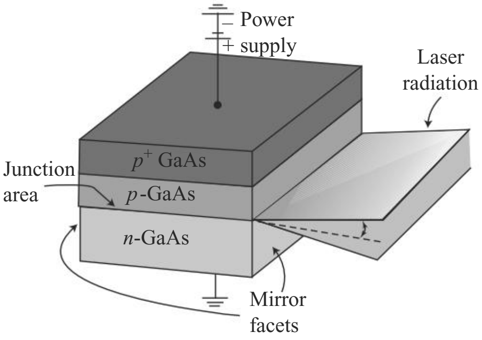


Fig. 5.24 *Radiation profile out of a simple laser diode*

5.11.3 Summary of Diode Lasers to this Point

- Charge carriers in a semiconductor laser diode are both the free electrons in the conduction band, and the positive holes in the valence band.
- In a *p-n* junction, electrons can 'fall' to the holes, which are lower energy states.
- Current passing through a *p-n* junction of a diode laser, causes both kinds of charge carriers (electrons and holes), to 'combine' together in the junction, through a process of recombination. Their energy is released in a form of photons of light. The energy of a photon is approximately equal to the energy of the energy gap.
- The energy gap is determined by the chemical composition of the diode laser, and by the crystal structure.

5.11.4 I-V Curve of Diode Laser

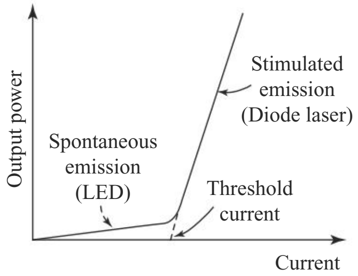


Fig. 5.25 *Output power from a diode laser as a function of input current*

If the condition of population inversion (which is required for the laser action) does not exist, the photons will be emitted by spontaneous emission. These photons will be emitted randomly in all directions, that is the basis of operation of a Light Emitting Diode (LED).

The condition for population inversion depends on pumping. By increasing the current injected through the *p-n* junction, we arrive at the threshold current which fulfills this condition. An example of the power output from a laser diode as a function of the injected current is shown in Fig. 5.25.

It is easily seen that the slope of this graph in a stimulated emission (laser) is far greater than the slope at spontaneous emission (LED).

The threshold current for lasing is determined by the intercept of the tangent to the graph at stimulated emission with the current axis (this point is very close to the point of change in the slope). When the current threshold is low; less energy will be wasted in the form of heat, and more energy will be transmitted as laser radiation (the laser efficiency increases). Practically, the important parameter is current density, which is measured in units of amperes per centimetre squared [A/cm^2] of the cross-section of the junction.

The name of each family of diode lasers is given by the type of materials near the active layer:

Homojunction Laser The entire laser is made from one substance, usually GaAs. In this simple structure, the emitted photons are not confined in the directions perpendicular to the laser axis. Thus, the laser is not efficient.

Single Heterostructure On one side of the active layer is a material with a different energy gap. Different energy gaps of the two layers cause a difference in the index of refraction of the materials, so we can build waveguide structures that confine the photons to a specific area. Usually, the second layer is made of a material which is similar to the first layer but has a lower index of refraction. Examples are GaAs and GaAlAs, typical materials used for adjacent layers in a diode laser.

Double Heterostructure On both sides of the active layer is another material which is different from the active layer and has a lower index of refraction (higher energy gap). For example, active layer of GaAs confined between two layers of Ga–Al–As. In this structure, light is confined into the active layer. Thus, these types of structures increase the laser efficiency, and make possible the operation of diode lasers at room temperatures.

5.11.5 Advantages of Diode Lasers

Very high efficiency (more than 20% of the input energy is emitted as laser radiation), high reliability, very long lifetime (estimated more than 100 years of continuous operation), very cheap price are the main advantages. Diode lasers are fabricated using mass production techniques used in the electronic industry, and there is a possibility to perform direct modulation of the emitted radiation by controlling the electric current through the *p-n* junction. The emitted radiation is a linear function of the current and can reach a modulation rate of tens of GHz. Other advantages are small volume, small weight, very low threshold current, low energy consumption, narrow spectrum band which can be a few kilohertz in special diode lasers.

CHARACTERISTICS OF LASERS

5.12

The main characteristics of laser radiations are (a) monochromaticity, (b) coherence, (c) high directionality, and (d) extreme brightness.

(a) Monochromaticity The laser light is almost perfectly monochromatic, much more monochromatic than that of any conventional monochromatic source. The optical resonator formed by two mirrors sustains only specific frequencies from the broad band. The mirrors convert the continuous emission band into a group of sharply defined discrete lines. By some suitable frequency-selective mechanism, a single line can be selected out from these lines.

(b) Coherence The beam of laser light is completely coherent, i.e., all waves are exactly in phase with one another. An interference pattern can be obtained also using beams from separate lasers. This is inherent in the process of stimulated emission which is responsible for laser action.

(c) High Directionality A laser emits light only in one direction. The width of a laser beam is extremely narrow and hence a laser beam can travel to long distances without spreading. They can be brought to an extremely sharp focus.

(d) Extreme Brightness Extreme brightness is the unique property of a laser beam. Due to high energy density and directional property, it can vaporise even the hardest metal; a laser beam can produce temperatures of the order of 104°C at a focused point. Further, a one milliwatt rediation of a helium–neon laser appears brighter than the sun.

5.13 USES OF LASERS

The most significant applications of lasers include telecommunications, medicine, manufacturing, the military, consumer goods and basic research.

(1) Telecommunications Telecommunications is one of the most important uses of lasers. Light from diode lasers is modulated by the signal that contains the information, which is then transmitted through a network of optical fibres. Future developments will most likely include tuneable lasers compatible with wavelength division multiplexing, a technique which will even further enlarge the bandwidth of fibre optic systems.

(2) Medicine Lasers are frequently used in surgery and medical diagnostics. Surgeons can now use (carbon dioxide) lasers as bloodless scalpels because the optical beam cauterises the incision as it is made. Lasers are also used to open up clogged arteries using a technique called **laser angioplasty.** Cancer treatment is a small but promising field that, today, is based on photodynamic therapy. In terms of the number of procedures, the most important growth areas are eyesight correction and cosmetic surgery. Cosmetic surgery also relies heavily on lasers for removal of pigments, tattoos, unwanted hair, spider veins, and skin resurfacing.

(3) Manufacturing The automotive industry relies greatly on high power CO_2 lasers for cutting and welding metals. Nd:YAG lasers are used in drilling holes, spot welding and marking. The printing industry uses a range of lasers, mostly ion lasers, but the trend is to use high-power diode lasers. The semiconductor industry is a major user of ultraviolet (excimer) lasers for photolithography. Photolithography plays an important role in the manufacture of integrated circuits with hundreds of millions of transistors on a semiconductor chip. The entertainment industry (i.e., laser light shows) is based on ion lasers.

(4) Military So far, (mercifully) lasers have been found to make poor weapons. On the other hand, they are extensively used in guiding missiles to their destination, range finders and other target designators. It is suggested that in future ‘star-wars’ applications lasers will be used in space-based weapons and other airborne systems.

(5) Consumer Goods DVDs, CD-ROMs, optical discs, etc, all rely on diode lasers. Big changes are coming with the development of laser diodes emitting in the blue and ultraviolet end of the spectrum. Shorter wavelengths (eg., blue and UV) will enable optical storage devices to pack more information per disc. Diode lasers are also at the heart of laser printers, barcode scanners, laser pointers, security devices, etc.

(6) Basic Research Lasers are the product of basic research, and researchers around the world are still developing new types of lasers. Physicists, chemists, biologists are major users of lasers, in areas such as spectroscopy, biology (e.g., the human genome program), laser fluorescence, holography, etc. Some experiments aiming to achieve controlled nuclear fusion, which one day may produce abundant and clean energy, also rely on lasers. Laser-induced fusion, is a technique in which light from lasers is used to raise the temperature of a deuterium and tritium pellet to 10^9°C.

(7) Chemical Applications Lasers can initiate or hasten certain chemical reactions which could not be possible in the absence of suitable photons. They can be utilised for investigating the structure of molecules. Raman spectroscopy is a method in which lasers have made so much impact that a separate branch named **Laser Raman Spectroscopy** has grown rapidly. By the use of lasers, the Raman spectrum can be obtained for much smaller samples and much faster too. Not only that, but by the use of lasers some interactions also arise due to high-intensity excitation, which provide additional information.

HOLOGRAPHY

5.14

A two-dimensional image of an object is the outcome of ordinary photography. To focus the image on the photographic plate, a lens system is used. In this case, focusing is only in a single plane and all other planes are out of focus. Therefore, a three-dimensional object is recorded in two dimensions. The intensity variations are recorded on the photographic plate while the phase distribution prevailing at the plane of the photographic plate is completely lost. Therefore, a photograph does not contain the three-dimensional character of the object. The quality of depth in a photograph is missing. This feature is recorded by holography. *Holography* is a Greek word where *holes* implies the whole and *graphy* implies the writing, i.e., holography implies 'complete recording'. Dennis Gabor (1948) of the Imperial College of Science and Technology, University of London, discovered the principle of holography. With the advent of lasers, Leith and Upatnicks in 1962 revised Gabor's idea in photography. This technique records and reproduces an image of the object without the use of lenses. The modification in holography is that the light waves reflected from the object are recorded instead of the image of the object, and the photographic record is called a **hologram**. We should note here that a holograph has no resemblance to the object in spite of it containing all information about the object in a kind of optical code. A three-dimensional image of the original object is formed when the holograph is illuminated by a coherent source of light. The image-formation process from a hologram is known as the **reconstruction process**. It involves a two-step process.

The first step involves the transformation of the object into a hologram. This is done by illuminating an object by coherent light which will produce interference fringes in a photographic emulsion. The second step involves the retransformation or reconstruction of the hologram into the image of the object which means a process involving reillumination of the developed interference pattern using light of same wavelength to produce a three-dimensional image of the original object.

5.14.1 Principle of Holography

When light is incident on a point object, the reflected wavefronts are spherical in nature with their centres at the point object. For a three-dimensional solid object, each point of the object will form its own reflected spherical wavefront. Therefore, the resulting wave pattern reflected from the entire object is highly complex. This wave pattern cannot be recorded on a photographic plate as such but it can be recorded by mixing the reflected wave pattern or object beam with a known wave called the **reference wave**. Interference will take place between the waves reflected from the object and the reference wave which modifies intensity at all points according to the phase relationship between the waves at those points. Thus, the phase variations in the wave pattern reflected from the object are converted into intensity variations.

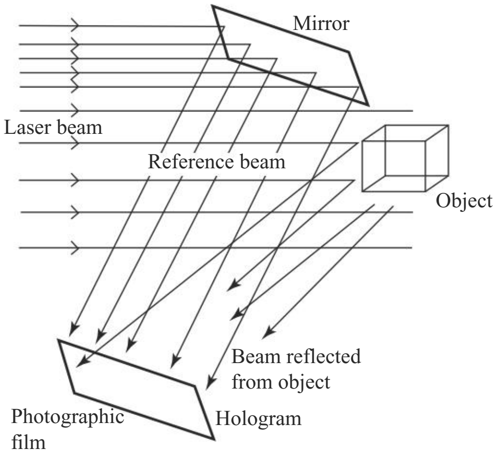


Fig. 5.26 *Construction of a hologram*

Construction of a Hologram Figure 5.26 shows the basic principle of making a hologram. The object is illuminated by a highly intense laser beam. The laser beam is divided into two parts by using a beam splitter not shown in the figure. One part of the beam is directly incident on the object, and the second part of the beam is incident on a mirror and reflected as a reference beam. The beam reflected from the object and the second beam, called the reference beam, after being reflected from the mirror is projected on the photographic plate. Thus, the photographic film is exposed simultaneously by both the reference beam and the beam reflected from the object. These two beams are highly coherent and highly directional. The resulting complex interference pattern is formed on the photographic film. After developing this photographic film, it will act as a hologram. The density at any point of the hologram depends both on the amplitudes and phases of the interference waves that reach the point.

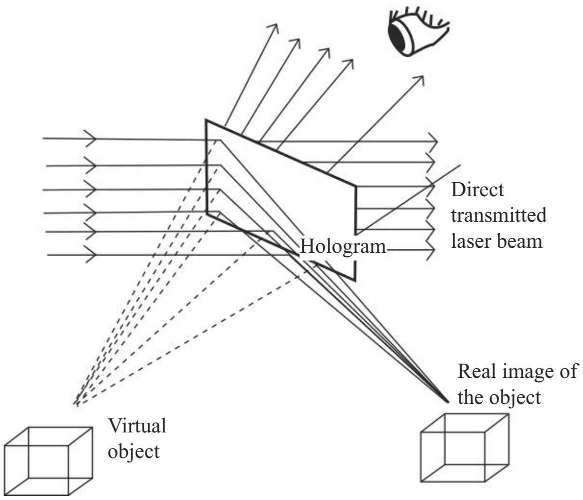


Fig. 5.27 *Reconstruction of image*

The structure of the hologram is like a diffraction grating—it has opaque and transparent regions very closely spaced. Thus, the film contains all the information needed to reproduce the wave field of the object.

Reconstruction of the Image To view the object, the image is reconstructed. The optical arrangement is simply reversed and the hologram is illuminated by a coherent light from the laser as shown in Fig. 5.27. Both real and virtual images are formed. When the viewer moves his eyes from side to side, the near parts of the object seem to move with respect to the far parts. Thus, when viewing the hologram, one sees the image depth, and hence a three-dimensional effect is obtained.

Here, the hologram acts as a complex grating; therefore, it diffracts the light incident on it. So zero-order diffraction takes place, in the direct direction of the beam giving no information because the laser beam passed through the hologram has only amplitude variation but no information regarding phase variation. In all other directions from the hologram, the diffracted waves of different orders carry both the information of intensity as well as phase variations. This way, reconstruction of the object wavefronts is made. Here, one of the diffracted beams forms real image while the other beam forms a virtual image as shown in Fig. 5.27.

Concluding Remark Laser will ultimately fulfill its promise in the huge field that is called analytical chemistry, and become increasingly prolific in diagnostic medicine. Spectral properties and spatial coherence will assure this as the lab shrinks to chip-scale. The future world of pharmagenomics will rely on lasers for genetic typing and perhaps for activation of the appropriate therapeutic course.

Lasers may also hold the key to the future of energy. Information from huge systems, such as the NIF and Laser MegaJoule and the rapid progress in tabletop high-repetition petawatt systems may converge to practical fusion power sources. Much has been accomplished in very few years, and the outlook is bright for numerous spin-offs to come. *It is clear, in any case, that laser is telling us again that science fiction of the past sometimes does become a scientific fact of the future.*

Solved Problems

1 *In an He–Ne laser system, the two energy levels of Ne involved in lasing action have energy values of 20.66 eV and 18.70 eV. Population inversion occurs between these two levels. What will be the wavelength of a laser beam produced? What will be the population of the metastable energy level with respect to the upper excited level at room temperature (27°C)?*

Solution Referring to the energy-level diagram of He–Ne laser, the energy levels involved are labeled as upper level $E_{3,\mathrm{Ne}}$, metastable level $E_{2,\mathrm{Ne}}$ and ground level $E_{1,\mathrm{Ne}}$.

Given, $E_{3,\mathrm{Ne}} = 20.66$ eV, $E_{2,\mathrm{Ne}} = 18.70$ eV $E_{1,\mathrm{Ne}} = 0$ eV

$$h\nu = E_2 - E_1 = \frac{hc}{\lambda}$$

$$h = 6.625 \times 10^{-34}\ \text{Js} = 4.14 \times 10^{-15}\ \text{eVs}$$

$$c = 3 \times 10^{8}\ \text{m/s}$$

$$k = 8.625 \times 10^{-5}\ \text{eV/K}$$

$$\lambda = \frac{hc}{E_2 - E_1} = \frac{hc}{E_{3,\mathrm{Ne}} - hc/E_{2,\mathrm{Ne}}}$$

$$\lambda = \frac{4.14 \times 10^{-15}\ \text{eVs} \times 3 \times 10^{8}\ \text{m/s}}{(20.66 - 18.70)\ \text{eV}}$$

$$= 6.3376 \times 10^{-7} = 633.76 \times 10^{-9} = 633.76\ \text{nm}$$

The laser output will have a wavelength of 633.76 nm.

$$\frac{N_2}{N_1} = \exp\left[-(E_2 - E_1)/kT\right]$$

$$kT = 8.625 \times 10^{-5}\ \text{eV}/K \times (273 + 27)K = 0.025\ \text{eV}$$

$$\frac{N_{3,\mathrm{Ne}}}{N_{2,\mathrm{Ne}}} = \exp\left[-\,(E_{3,\mathrm{Ne}} - E_{2,\mathrm{Ne}})/kT\right]$$

$$\therefore \quad \frac{N_{3,\mathrm{Ne}}}{N_{2,\mathrm{Ne}}} = \exp\left[-\,(20.66 - 18.70)/0.025\right] = \exp\left[-75.75\right] = 1.26 \times 10^{-33}$$

$$\therefore \quad N_{2,\mathrm{Ne}} = 7.89 \times 10^{32}\, N_{3,\mathrm{Ne}}$$

The above result implies that the metastable population is 10^{32} times the population in the excited states.

2 *Find out the separation between metastable and excited levels for two wavelengths of 9.6 mm and 10.6 mm emitted from a CO_2 laser source. Calculate the frequency and hence the energy of the light photons emitted. How many photons are required to be emitted per second to obtain a laser output power of 10 kW?*

Solution Given two wavelengths: 9.6 mm and 10.6 μm

$$h = 6.62 \times 10^{-34}\text{ Js} = 4.14 \times 10^{-15}\text{ eVs}$$
$$c = 3 \times 10^{8}\text{ m/s}$$
$$k = 8.625 \times 10^{-5}\text{ eV/K}$$

$$h\nu = E_2 - E_1 = \frac{hc}{\lambda} = \frac{1.2422 \times 10^{-6}\text{ eVm}}{9.6 \times 10^{-6}\text{ m}} = 0.129\text{ eV}$$

and

$$E_2 - E_1 = \frac{hc}{\lambda} = \frac{1.2422 \times 10^{-6}\text{ eVm}}{10.6 \times 10^{-6}\text{ m}} = 0.117\text{ eV}$$

So the estimated separation between the two required levels is 0.129 eV and 0.117 eV respectively. The energy of the photon is also respectively 0.129 eV and 0.117 eV. The frequency of the two different types of photons will be

$$\nu = \frac{c}{\lambda} = \frac{3 \times 10^{8}\text{ m/s}}{9.6 \times 10^{-6}\text{ m}} = 3.125 \times 10^{13}\text{ Hz}$$

$$\nu = \frac{c}{\lambda} = \frac{3 \times 10^{8}\text{ m/s}}{10.6 \times 10^{-6}\text{ m}} = 2.83 \times 10^{13}\text{ Hz}$$

To get 10 kW output power, 10000 J/s power is required

The energy of one photon is

$$h\nu = 6.625 \times 10^{-34} \times 3.125 \times 10^{13} = 2.07 \times 10^{-20}\text{ J}$$

Similarly, for other photon the energy in joules is

$$h\nu = 6.625 \times 10^{-34} \times 2.83 \times 10^{13} = 1.87 \times 10^{-20}\text{ J}$$

So, the number of photons per second required is

(10000 J/s)/2.07 × 10^{-20} = 4.83 × 10^{23} photons/second, and

(10000 J/s)/1.87 × 10^{-20} = 5.34 × 10^{23} photons/second.

So, approximately one mole of atoms are involved in the process in one second.

3 *A typical laser system is capable of lasing at infrared wavelengths. The light output at 3.124 mm is very prominent. What is the difference in the energy levels of the excited state and metastable state? What will be the energy of a photon emitted? What will be the frequency of the light emitted? If 1 mole of photons are emitted per second, what is the power of the laser output? Can you predict the type of the laser produced?*

Solution

$$\lambda = 3.124\ \mu\text{m} = 3.124 \times 10^{-6}\text{ m}$$

$$h\nu = E_2 - E_1 = \frac{hc}{\lambda} = \frac{1.2422 \times 10^{-6}\text{ eVm}}{3.124 \times 10^{-6}\text{ m}} = 0.398\text{ eV}$$

Energy difference between the metastable and excited states is 0.398 eV and hence the photon energy emitted is also 0.398 eV.

The frequency of the photon is

$$v = \frac{c}{\lambda} = \frac{3 \times 10^8 \text{ m/s}}{3.124 \times 10^{-6} \text{ m}} = 9.6 \times 10^{13} \text{ Hz}$$

The energy in joules is

$$hv = 6.625 \times 10^{-34} \times 9.6 \times 10^{13} = 6.36 \times 10^{-20} \text{ J}$$

One mole of photon

$$6.022 \times 10^{23} \times 6.36 \times 10^{-20} = 38299.92 = 38.3 \text{ kJ}$$

Therefore 38.3 kW of output power of laser beam is observed. Obviously, it will be a CO_2 laser light.

4 *A ruby laser has its metastable state at 1.79 eV from which stimulated emission produces laser light. Calculate the wavelength of the light. At room temperature when the population inversion is not achieved, calculate the ratio of the population of the atom in the metastable state to that in the ground state.*

Solution The difference in the energy levels

$$E_2 - E_1 = 1.79 \text{ eV}$$

$$E_2 - E_1 = \frac{hc}{\lambda} = \frac{1242.18}{\lambda},$$

where, energy is in J, wavelength in nm.

$$\lambda = \frac{1242.18}{E_2 - E_1} = \frac{1243.18}{1.79} = 693.95 \text{ nm}$$

Therefore, emitted light will have a wavelength of 693.95 nm.

Now the population ratio at temperature T for two different states having energy E_2 and E_1 is given by the relation

$$\frac{N_2}{N_1} = \exp\,[-(E_2 - E_1)/kT]$$

Taking room temperature = 27°C + 273 = 300 K

Thermal energy at room temperature = $kT = 8.625 \times 10^{-5}$ eV/K × 300 K = 0.025 eV

$$\frac{N_2}{N_1} = \exp\,[-(E_2 - E_1)/kT]$$
$$= \exp\,[-1.79/0.025] = \exp\,[-71.6] = 8.026 \times 10^{-32}$$

5 *A ruby laser emits light of 693.95 nm wavelength as calculated in the above problem. If 1 mole of Cr^{+3} ions are involved in population inversion process in a pulse, calculate the pulse energy in eV.*

Solution Given, $\lambda = 693.95$ nm;

$$E_2 - E_1 = \frac{hc}{\lambda} = \frac{1242.18}{\lambda} = \frac{1242.18}{693.95} = 1.79 \text{ eV}$$

1.79 eV is the energy of the single photon and 1 mole (= 6.022×10^{23} Cr^{+3} ions are involved in population inversion that produce a laser pulse.

Therefore, the energy is $6.022 \times 10^{23} \times 1.79$ eV $= 1.077 \times 10^{24}$ eV.

6 *A ruby laser emits light of 693.95 nm wavelength. The duration of pulses is 0.1 ns. Calculate the coherence length, bandwidth and line width.*

Solution Given, $\lambda = 693.95$ nm;

[The coherence length can be used for quantifying the degree of a temporal coherence as the propagation length (and thus propagation time) over which coherence degrades significantly.]

(a) Coherence length = (coherence time) × (velocity of light in vacuum)

$$L_{\text{coh}} = t_{\text{coh}} \times c$$

$$L_{\text{coh}} = 0.01 \times 10^{-9} \times 3 \times 10^{8} = 0.03 \text{ m}$$

(b) Bandwidth, $$\Delta v = \frac{1}{t_{\text{coh}}} = \frac{1}{1 \times 10^{-9}} = 10^{10} \text{ Hz}$$

(c) Line width, $$\Delta\lambda = \frac{\lambda^2}{c}\Delta v = \frac{(693.95 \times 10^{-9})^2}{3 \times 10^{8}} \times 10^{10} = 0.016 \text{ nm}$$

7 *A laser beam of 693.95 nm wavelength on earth is focused by a lens or mirror of 2 m diameter on to a crater on the moon. The distance of the moon is 3.82 × 10^8 m from the earth. How much larger will be the spot on the moon? Neglect the effect of earth's atmosphere.*

Solution Angular spread is given by

$$d\theta = \frac{\lambda}{d} = \frac{693.95 \times 10^{-9} \text{ m}}{2} = 3.469 \times 10^{-7} \text{ rad}$$

$$\therefore \quad d\theta = 3.469 \times 10^{-7} \text{ rad} \times \left(\frac{360^\circ}{2\pi}\right) = (1.98 \times 10^{-5})^\circ = 0^\circ 0' 0.07''$$

Area spread, i.e., area of the spot on the moon

$$A = (D\,d\theta)^2 = (3.82 \times 10^{8} \times 3.469 \times 10^{-7})^2$$

$$= 17570.56 \text{ m}^2 = 1.76 \times 10^{4} \text{ m}^2$$

This gives a circular area of radius 74.78 m.

8 *A laser having power of 75 mW, wavelength of 720 nm and an aperture 5 mm is focused with a lens of focal length of 0.1 m. Calculate the area and intensity of the image.*

Solution $$d\theta = \frac{\lambda}{d} = \frac{720 \times 10^{-9} \text{ m}}{5 \times 10^{-3}} = 1.44 \times 10^{-4} \text{ rad}$$

$$\text{Area spread} = (f\,d\theta)^2 = (0.1 \times 1.44 \times 10^{-4})^2 = 2.074 \times 10^{-10} \text{ m}^2$$

Intensity of the image is given by

$$I = \frac{\text{power}}{\text{areal spread}} = \frac{75 \times 10^{-3}}{2.074 \times 10^{-10}} = 3.616 \times 10^{8} \text{ W/m}^2$$

Multiple-Choice Questions

1. Laser light is produced due to
- (a) interference of light
- (b) spontaneous emission of light
- (c) light Amplification by stimulated emission of radiation
- (d) diffraction phenomenon

2. Which laser was invented first?
- (a) Semiconductor laser
- (b) Ruby laser
- (c) He–Ne laser
- (d) CO_2 laser

3. Which of the following is a gas laser?
- (a) CO_2 laser
- (b) Ruby laser
- (c) Semiconductor laser
- (d) Dye laser

4. Two different light sources are sending light and spots from both beams on a screen is obtained. Which of the following statements is true?
- (a) The broad and faint spot is a laser while the bright and narrow spot is an ordinary light.
- (b) The broad and bright spot is a laser while the narrow and faint spot is an ordinary light.
- (c) The broad and low intensity spot is an ordinary light, the narrow and highly intense spot is a laser light.
- (d) Actually both the spots will merge into each other and form an interference patterns.

5. A highly coherent beam is produced. The beam is
- (a) laser light
- (b) ordinary light
- (c) converged light from the sun
- (d) spark produced between the gap of two electrodes

6. Which of the following conditions is very essential for the production of laser light?
- (a) Spontaneous emission process
- (b) Stimulated emission process
- (c) Population inversion process
- (d) All of the above

7. Which of the following is not a pumping process?
- (a) Optical pumping
- (b) Electrical pumping
- (c) Chemical pumping
- (d) Thermal pumping

8. Which of the following scheme does not produce lasing action?
- (a) Two-level scheme
- (b) Three-level scheme
- (c) Four-level scheme
- (d) Five-level scheme

9. High-power laser light is obtained from
- (a) CO_2 laser
- (b) semiconductor laser
- (c) He–Ne laser
- (d) ruby laser

10. Pulsed laser light is produced from a
- (a) ruby laser
- (b) carbon laser
- (c) carbon dioxide laser
- (d) none of the above

11. Which of the following is not a laser property?
- (a) Coherence
- (b) Highly directional
- (c) Extreme brightness
- (d) Divergence

12. Production of laser does not include
- (a) active medium
- (b) two highly silvered optically flat mirrors
- (c) optical cavity
- (d) one highly silvered and other partially silvered optically flat mirror

13. Which source of light is brightest?
    (a) Sunlight
    (b) Laser light
    (c) Arc light
    (d) Flash light

Answers to MCQs

1. (c)	2. (b)	3. (a)	4. (c)	5. (a)	6. (d)	7. (d)	8. (a)
9. (a)	10. (a)	11. (d)	12. (b)	13. (b)			

Practice Problems

1. A laser system is capable of lasing at infrared wavelengths. If out of several possible wavelengths the most prominent wavelength is 750 nm and 1 mm, then calculate the difference between upper excited and lower metastable state in eV for this wavelengths. **[1.65 eV, 0.00124 eV]**

2. Calculate the wavelength of radiation for the CO_2 laser having an energy difference of 0.121 eV between two states involved in lasing action. **[102479.3 Å]**

3. The laser light emitted in ruby laser has a wavelength of 7000 Å and the total number of Cr^{+++} ions is 2.8×10^{19}. Calculate the energy of a photon and total energy available per laser pulse. **[1.77 eV, 7.92 J]**

4. Calculate the population inversion ratio for the two states involved in lasing action in a ruby laser at temperatures of 27°C and 227°C. Consider the emitted wavelength is 7000 Å. **[5.9×10^{-29}, 1.4×10^{16}]**

5. He–Ne laser produces light of 7000 Å wavelength at room temperature (= 27°C). Calculate the population inversion of two states responsible for lasing action. **[5.9×10^{-29}]**

6. Imagining that we chop a continuous laser beam which is assumed to be perfectly monochromatic having a wavelength of 623 nm into 0.1 ns pulses using a suitable shutter, calculate the resultant line width, bandwidth and coherence lengths.

[3 cm, 10^{10} Hz, 0.13 Å]

7. A laser beam of 8×10^{-7} m wavelength and 5×10^{-3} m aperture on the earth is sent on the moon. The distance of the moon is 3.82×10^{8} m from the earth. How much will be the angular spread of the beam and the areal spread when it reaches the moon? **[1.6×10^{-4} radian, 4.096×10^{9} m^2]**

8. A laser having power of 50 mW, wavelength of 720 nm and aperture of 5 mm is focused with a lens of 0.1 m focal length. Calculate the angular spread, area and intensity of the image. **[1.44×10^{-4} radian, 2.074×10^{-10} m^2, 2.411×10^{8} W/m^2]**

Short-answer Questions

1. How is a ruby laser favorable to lasing action? Why are the end faces of a ruby rod silvered?

2. What is the function of He atoms in an He–Ne laser?

3. Why is laser action not possible without population inversion between two atomic levels?

4. What are the other gases used in a CO_2 laser? What is the proportion?

5. Why are four-level lasers more efficient than three-level lasers?

6. What is an optical resonator cavity, and what is its role?

7. In lasers, the active medium of broad absorption band is required. Why?

8. How does metastable state act in laser media?

9. Why is a narrow tube used in an He–Ne laser? What happens if its diameter is increased?

10. Why is a laser more coherent and more monochromatic than ordinary light?

11. How is stimulated emission made to be more dominant than spontaneous emissions in a laser?

12. What is the function of mirrors in Ruby and He–Ne lasers?

13. What is coherence length, line width and bandwidth?

14. What do you mean by pulsed and continuous laser?

Long-answer Questions

1. What do you mean by lasers? Define spontaneous and stimulated emissions.

2. Define the following terms: (a) population inversion, (b) pumping, and (c) active system.

3. Explain the process of stimulated emission. Draw a neat diagram to represent the component of a ruby laser. Explain the operation.

4. Explain the operation of a gas laser with the essential components. How does stimulated emission take place with the exchange of energy between helium and neon atoms?

5. What are different uses to which laser beams are put?

6. Differentiate between spontaneous emission and stimulated emission in their mechanism and characteristics. Which one of them is maximised in a laser operation?

7. Explain the terms (i) stimulated emission, (ii) spontaneous emission, (iii) optical pumping, (iv) active medium, (iv) optical resonator, (vi) population inversion, (vii) metastable state.

8. Explain construction and working of the He–Ne laser with the help of an energy level diagram.

9. Explain construction and working of a semiconductor laser.

10. Explain in detail the construction and working of a CO_2 laser.

11. Deduce the relation between spontaneous and stimulated emission probabilities.

12. What are Einstein's coefficients? Derive Einstein's relation.

13. Explain the principle of optical pumping and stimulated emission of radiation. Discuss the properties of laser radiation and mention some of its applications.

14. Laser action is sometimes called 'inverted absorption'. Explain in what situation may A/B be small enough for laser action?

Questions From Previous Exams

(i) State the properties of LASER. **(GTU- Jan' 2009)**

It has four properties namely, coherence, high intensity, high directionality and high monochromatic.

(ii) What is the lifetime of charge carrier in metastable state? **(GTU- Jan' 2009)**

The lifetime of charge carrier in metastable state is 10^{-3} to 10^{-2} second.

(iii) State the full form of LASER **(GTU- Jan' 2009)**

Light **A**mplification by **S**timulated **E**mission of **R**adiation.

(iv) Describe the construction and working of Nd-YAG laser **(GTU- Jan' 2009)**

(Refer Section: 5.9)

(v) Derive the relation between Einstein's A and B coefficients. **(GTU- Jan' 2009)**

(Refer Section: 5.2.1)

(vi) Briefly explain stimulated emission. **(GTU- Jan' 2010)**

It is process in which there is emission of a photon whenever an atom transits from a higher energy state to lower energy state under the influence of an external agency, i.e., inducing photon

(vii) Explain (i) population inversion (ii) pumping (iii) optical resonator. **(GTU- Jan' 2010)**

Population Inversion It is the state of achieving more number of atoms in excited state compared to the ground state, i.e., $N_2 > N_1$.

Pumping It is a mechanism of exciting atoms from the lower energy state to higher energy state by supplying energy from an external source.

Optical resonator It is pairs of reflecting mirrors; of which one is being perfect reflector and other being partial reflector. It is used for amplification of photons thereby producing an intense and highly coherent output.

(viii) Describe the construction and working of Nd: YAG Laser with a suitable energy level diagram. **(GTU- Jan' 2010)**

(Refer Section: 5.9)

(ix) State the characteristics of LASER. Explain the method of construction and reconstruction of a hologram. **(GTU- Jan' 2010)**

(Refer Section: 5.12, 5.14)

(x) What is population inversion? **(GTU- June' 2010)**

Population Inversion It is the state of achieving more number of atoms in excited state compared to the ground state, i.e., $N_2 > N_1$.

(xi) The active medium in Nd: YAG laser is Nd^{+3} **(GTU- June' 2010)**

(xii) In carbon dioxide laser, the energy difference between two levels is 0.121 eV. Calculate the wavelength of radiation. **(GTU- June' 2010)**

Given 0.121 eV and hence 1 eV = 1.602×10^{-19} J

$$E_2 - E_1 = 0.121 \times 1.602 \times 10^{-19} \text{ J}$$
$$h = 6.626 \times 10^{-34} \text{ Js}$$
$$c = 3 \times 10^{8} \text{ m/s}$$
$$\lambda = hc/(E_2 - E_1)$$
$$\lambda = 1.0225 \times 10^{-5} \text{ m}$$

(xiii) Explain the construction and working of CO$_2$ laser with a suitable energy level diagram **(GTU- June' 2010)**

(Refer Section: 5.10)

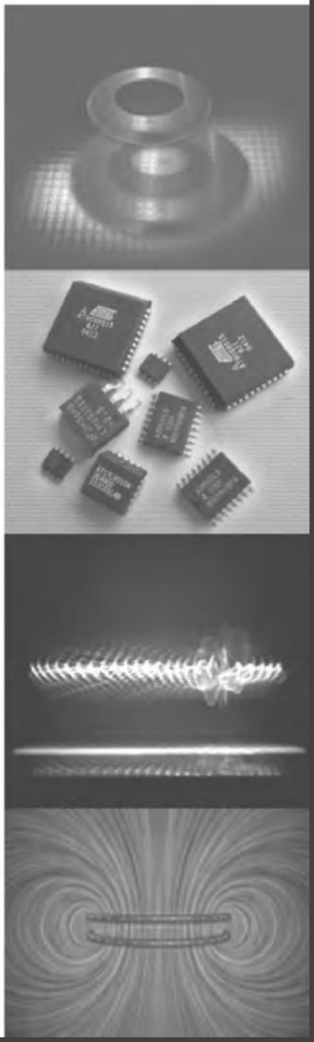
(xiv) Established the relation between Einstein's coefficients. **(GTU- June' 2010)**

(Refer Section: 5.2.1)

Optical-Fibre Communication

6

It is very comforting that today communication through telephone lines is very clear, without any disturbance, and even for a long-distance call, the sound clarity is perfect as a short-distance call. No waiting for calls and no network jam are other welcome changes. Even faults in communication lines nowadays are repaired so fast that we have forgotten the old days of long waiting for our calls to be connected, large amount of disturbance, cross-connections, traffic jams in communication lines and so on. This miracle happened due to the technology advancement that you will study in this chapter.



INTRODUCTION TO FIBRE OPTICS

6.1

Fibre Optics is the communication technology that works by sending signals down hairlike thin strands of glass fibres or plastic fibres. This technology was initiated about 30 years ago in UK and USA in 1976. By the early 1980s, fibre networks connected the major cities on each coast.

Fibre has replaced all the telephone copper cables, microwave and satellite links. CATV (Cable TV) offers phone and Internet service on the same fibre. Computers and LANs started using fibres and other applications like aircraft, ship and automobile data buses, CCTVs for security, even links for consumer digital stereo developed in the wake. Today, fibre optics is either the dominant medium or a logical choice for every communication system.

Fibre optics is a medium for carrying information from one point to another in the form of light, but is not electrical in nature like the copper form of transmission.

Fibre-Optic System A basic fibre-optic system consists of a transmitting device which generates the light signal; an optical fibre cable, that carries the light; and a receiver to accept the light signal transmitted. The fibre itself is passive and does not contain any active, generative properties.

Advantages of fibre-optic communication compared to metal base system

So the fibre optic system forms a communication system which uses light as the carrier of information from a source and propagated through a guided fibre cable (glass or plastic) to a destination.

For any comunication system, it is established that the information-carrying capacity of a connunication system is directly proportional to its bandwidth. This indicated that for a wide bandwidth, information capacity is more. The frequency range of light used is 10 kHz to 400,000 GHZ therefore, fibre-optic system has higher information-carrying capacity. Added advantage to this is extremely low power loss, i.e., about 0.2 dB/km or 0.5% of power loss over a distance of 1 km. Due to high information-carrying capacity and low attenuation, fibre are preferred in telecommunications, local area networks, computer networks, etc.

Optical fibre systems have many advantages over metal-based communication systems. These advantages are the following:

Long-Distance Signal Transmission The low attenuation and superior signal integrity found in optical systems allow much longer intervals of signal transmission than metal-based systems, while single-line, voice-grade copper systems longer than a few kilometres (1.2 miles) require in-line signal repeaters for satisfactory performance. It is common for optical systems to cover over 100 kilometres (km), or about 62 miles, with no active or passive processing (see Fig. 6.1). Emerging technologies promise even greater distances in the future. Figure 6.2 shows a fibre cable.

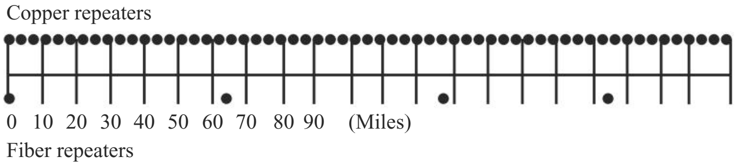


Fig. 6.1 *Copper and fibre repeaters comparison*

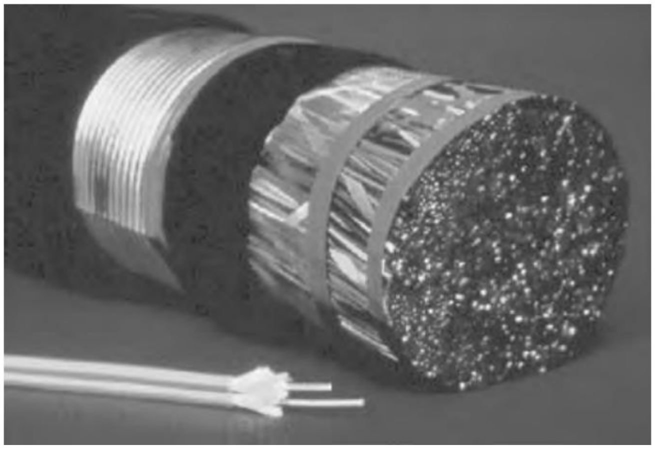


Fig. 6.2 *The optical fibre cable in the foreground has the equivalent information-carrying capacity of the copper cable in the background*

Large Bandwidth, Light Weight, and Small Diameter Today's applications require an ever-increasing amount of bandwidth and space constraints of many end-users. It is commonplace to install new cabling within existing duct systems. The relatively small diameter and light weight of optical cables makes such installations easy and practical, and saves valuable conduit space in these environments.

Long Lengths Long, continuous lengths and small diameters make it practical to manufacture and install much longer lengths than that needed for metallic cables: twelve-kilometer (12 km) continuous optical cable lengths are common. Some cable systems manufacturers make continuous single-mode cable lengths of up to 12 km, with a 96-inch reel size being the primary limiting factor.

Multimode cable lengths based on industry demand can be 4 km or more, although most standards require a maximum length of 2 km or less.

Easy Installation and Upgrades Long lengths of optical cables make installation much easier and less expensive. Optical fibre cables can be installed with the same equipment used to install copper and coaxial cables, with some modifications due to the small size and limited pull tension and bend radius of optical cables.

Optical cables can typically be installed in duct systems in spans of 6 km or more depending on the duct's condition, layout of the duct system, and installation technique. The longer cables can be coiled at an intermediate point and pulled farther into the duct system as necessary.

System designers typically plan optical systems to meet growth needs for a 15- to 20-year span. Although sometimes difficult to predict, growth can be accommodated by installing spare fibres for future requirements. Installation of spare fibres today is more economical than installing additional cables later.

Nonconductivity Another advantage is due to its dielectric nature. Since an optical fiber has no metallic components, it can be installed in areas with Electromagnetic Interference (EMI), including Radio Frequency Interference (RFI). Areas with high EMI include utility lines, power-carrying lines, and railroad tracks. All-dielectric cables are also ideal for areas of high lightning-strike incidence.

Security The dielectric nature of optical fibres makes it impossible to remotely detect the signal being transmitted within the cable. The only way to do so is by actually accessing the optical fibre itself which requires intervention that is easily detectable by security surveillance. These circumstances make fibres extremely attractive to governmental bodies, banks, and other organisations with major security concerns.

Designed for Future Applications Needs As bandwidth demands increase rapidly with technological advances, fibres will continue to play a vital role in the long-term success of telecommunications. Fibre solutions are less costly than copper.

A fibre optic cable functions as a 'light guide,' guiding the light introduced at one end of the cable through to the other end. The light source can either be a Light-Emitting Diode (LED) or a laser. The light source is pulsed on and off, and a light-sensitive receiver on the other end of the cable converts the pulses back into the digital ones and zeros of the original signals. Even a laser light shining through a fibre optic cable is subject to loss of strength, primarily through dispersion and scattering of the light within the cable itself. The faster the laser fluctuates, the greater the risk of dispersion. Light strengtheners, called repeaters, may be necessary to refresh the signal in certain applications.

There are three types of fibre optic cables commonly used, namely, **single mode**, **multimode** and **Plastic Optical Fibre** (POF).

Transparent glass or plastic fibres allow light to be guided from one end to the other with minimal loss. Note that fibre optic cables have become cheaper over time—an equivalent length of copper cable costs less per foot but is not adequate in capacity.

Fiber-optic cable connectors and the equipment needed to install them are still more expensive than their copper counterparts.

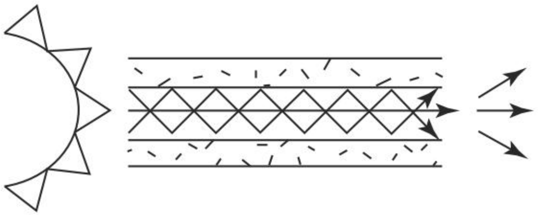


Fig. 6.3 *Signal transmission through total internal reflection*

We know that light waves cannot travel far in open atmosphere as the energy gets dissipated very rapidly. Hence, some kind of guiding channel is needed like a guiding electric current through a metal wire. Optical fibres provide the necessary waveguide for light. Fibre optics is a technology related to transportation of optical energy (light energy) in guiding media, specifically glass fibres (See Figures 6.3 and 6.4).

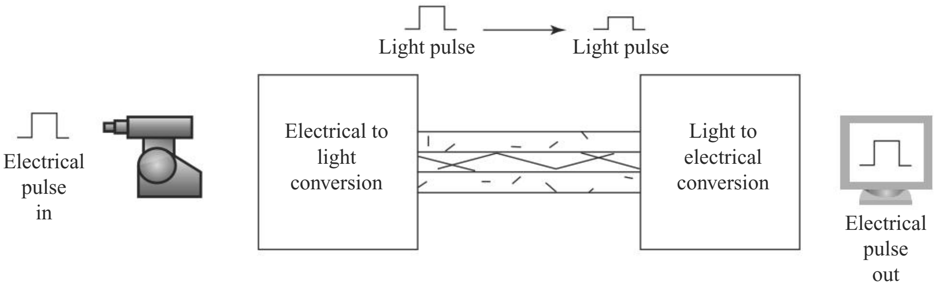


Fig. 6.4 *General view of signal communication in fibre optics*

PRINCIPLE OF OPTICAL FIBRE

6.2

Total internal reflection is the main principle of optical fibre communication. The phenomenon in which light is totally reflected from a denser-to-rarer medium boundary is known as **total internal reflection**. An optical fibre, which is a hair-thin cylindrical fibre of glass or any transparent dielectric medium, consists of many thousands of very long fine-quality glass/quartz fibres (Fig. 6.2). The fibres are coated with a layer of transparent material having a low refractive index, as shown in Fig. 6.5.

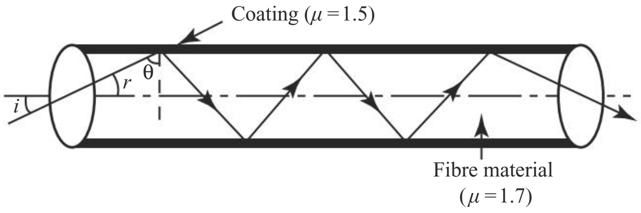


Fig. 6.5 *Total internal reflection principle in an optical fibre*

Light is incident at one end of the fibre making a small angle with the axis and passes through the fibre as explained below. Let i and r be the angles of incidence and refraction of the light ray with the axis, and θ be the angle at which the ray is incident on the fibre boundary so that $\theta = (90° - r)$. Let n_1, be the refractive index of the fibre and θ_C be the critical angle where $\theta_C = \sin^{-1}(1/n_1)$ and $\theta > \theta_C$. Then the ray is totally internally reflected. Therefore, the ray undergoes repeated total internal reflections until it emerges out of the other end of the fibre even if the fibre is bent. So the light ray is guided through the fibre from one end to the other end without any energy being lost due to refraction. The bundle of optical fibres is used for optical signal transmission and receiving electrical signals (which are converted to light by suitable devices).

It is essential that there must be very little absorption of light as it travels through long distances inside the optical fibre. This can be achieved by purification and special preparation of the material.

STRUCTURE AND CLASSIFICATION OF OPTICAL FIBRES

6.3

The fibres used for optical communication are wave guides made of transparent dielectrics whose function is to guide visible and infrared light over long distances. An optical fibre consists of an inner cylindrical portion of glass, called the **core**. The core that carries light is surrounded by another cylindrical shell of lower refractive index called the **cladding**. It helps to keep the light within the core through the phenomenon of total internal reflection.

The core and cladding are shown in Fig. 6.6. The range of the core diameter is 5 to 100 μm. The cladding diameter is usually 125 μm. Over the cladding, a soft plastic coating of about 250 μm diameter provides greater strength and protection to the fibre. Finally, strength members and the jacket provide the final protecting layers. The optical fibres are classified as (i) step-index optical fibres, and (ii) graded-index optical fibres.

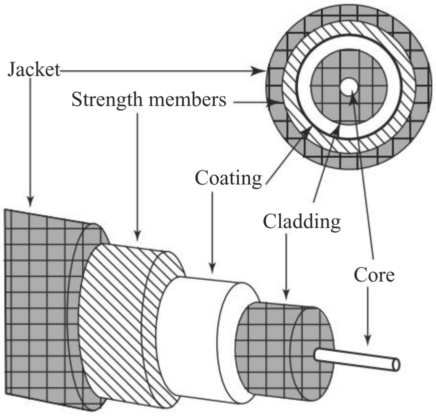


Fig. 6.6 *Structure of optical fibre*

Understanding the characteristics of different types of fibres helps in understanding the applications for which they are used. Operating a fibre optic system properly depends on knowing the type of fibre is being used. There are two basic types of fibres: Multimode fibres and single-mode fibres. **Multimode fibre** is best designed for short transmission distances, and is suited for use in LAN (Local Area Network: A communication link between two or more points within a small geographic area, such as between buildings) systems and video surveillance. **Single-mode fibre** is best designed for longer transmission distances, making it suitable for long-distance telephony and multichannel television broadcast systems.

Multimode fibre simply refers to the fact that numerous modes (*mode*: a single electromagnetic wave traveling in a fibre) or light rays are carried simultaneously through the waveguide. (*Waveguide*: a material medium that confines and guides a propagating electromagnetic wave. A waveguide normally consists of a hollow metallic conductor, generally rectangular, elliptical, or circular in cross-section in the microwave regime. This type of waveguide may, under certain conditions, contain a solid or gaseous dielectric material. In the optical regime, a waveguide used as a long transmission line consists of a solid dielectric filament (fibre), usually circular in cross-section. For integrated optical circuits, an optical waveguide may consist of a thin dielectric film. For the RF regime, ionised layers of the stratosphere and the refractive surfaces of the troposphere may also serve as a waveguide.) Modes result from the fact that light will only propagate in the fibre core at discrete angles within the cone of acceptance. This fibre type has a much larger core diameter, compared to a single-mode fibre, allowing for the larger number of modes, and a multimode fibre is easier to couple than a single-mode optical fibre. A multimode fibre may be categorised as a step-index or graded-index fibre.

6.3.1 Multimode Step-Index Fibre

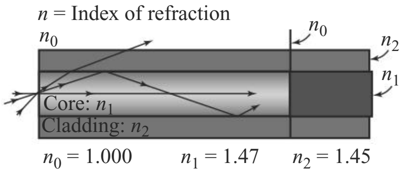


Fig. 6.7 *Total internal reflection in multimode step-index fibre*

Figure 6.7 shows how the principle of total internal reflection applies to a multimode step-index fiber. Because the index of refraction of the core is higher than the index of refraction of the cladding, ($n_1 > n_2$), the light that enters at less than the critical angle is guided along the fibre.

Three different light waves travel down the fibre. One mode travels straight down the centre of the core. A second mode travels at a steep angle and bounces back and forth by total internal reflection. The third mode exceeds the critical angle and refracts into the cladding. Intuitively, it can be seen that the second mode travels a longer distance than the first mode, causing the two modes to arrive at different times. This disparity between arrival times of the different light rays is known as **dispersion**, and the result is a muddied signal at the receiving end. However, it is important to note that high dispersion is an unavoidable characteristic of a multimode step-index fiber.

6.3.2 Multimode Graded-Index Fibre

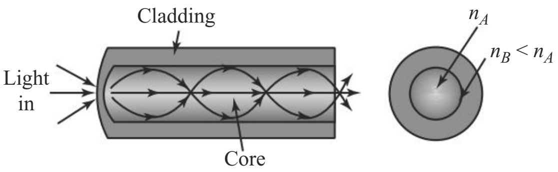


Fig. 6.8 *Multimode graded-index fibre*

Graded-index refers to the fact that the refractive index of the core gradually decreases farther from the centre of the core. The increased refraction in the centre of the core reduces the speed of some light rays, allowing all the light rays to reach the receiving end at approximately the same time, thereby reducing dispersion of light. Figure 6.8 shows the principle of a multimode graded-index fiber. The central refractive index in the core, n_A, is greater than that of the outer core's refractive index, n_B. As discussed earlier, the core's refractive index is parabolic, being higher at the centre.

As Fig. 6.8 shows, the light rays no longer follow straight lines; they follow a serpentine path being gradually bent back toward the centre by the continuously reducing refractive index. This lowers the arrival time disparity because all modes arrive at about the same time. The modes traveling in a straight line are in a higher refractive index, so they travel slower than the serpentine modes. These rays travel farther but move faster in the lower refractive index of the outer core region.

6.3.3 Single-Mode Fibre

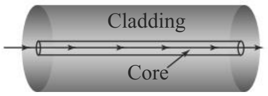


Fig. 6.9 *Single-mode fibre*

A single-mode fibre (Fig. 6.9) allows for a higher capacity to transmit information because it can retain the fidelity of each light pulse over longer distances, and it exhibits no dispersion caused by multiple modes. It also enjoys lower fibre attenuation than a multimode fibre. So, more information can be transmitted per unit of time. Like a multimode fibre, earlier a single-mode fibre was generally characterised as a step-index fibre, meaning the refractive index of the fibre core is a step above that of the cladding rather than graduated as it is in graded-index fibre. Modern single-mode fibres have evolved into more complex designs such as matched clad, depressed clad and other exotic structures.

The disadvantages of single-mode fibres are due to the smaller core diameter that makes coupling light into the core more difficult. The tolerances for single-mode connectors and splices are also much more demanding. (*Connector*: A mechanical or optical device that provides a demountable connection between two fibres or a fibre and a source or detector. *Splice*: A permanent connection of two optical fibres through fusion or mechanical means.)

Single-mode fibres, after a continuing evolution for several decades, are of three basic classes used in modern telecommunications systems. (i) The oldest and most widely deployed type is the Non Dispersion-Shifted Fibre (NDSF). These fibres were initially intended for use near 1310 nm. (ii) Due to very high dispersion at the 1550 nm wavelength, 1550 nm systems made the NDSF fiber undesirable. To resolve this shortcoming, fibre manufacturers developed the Dispersion-Shifted Fibre (DSF) that moved the zero-dispersion point to the 1550 nm region. Afterwards, it was discovered that while DSF worked extremely well with a single 1550 nm wavelength, it exhibits serious nonlinearities when multiple, closely spaced wavelengths in the 1550 nm were transmitted in DWDM systems. (iii) A new class of fibres, classified as NonZero-Dispersion-Shifted fibres (NZ-DSF) has been developed to resolve the problem of nonlinearities. The fibre is available in both positive and negative dispersion varieties and is fast becoming the fibre of choice in new fibre deployment.

One additional important variety of the single-mode fibre is the Polarisation-Maintaining (PM) fibre. All other single-mode fibres discussed are capable of carrying randomly polarised light. A PM fibre is designed to propagate only one polarisation of the input light. This is important for components such as external modulators that require a polarised light input.

Figure 6.10 shows the cross-section of a type of PM fibre. This fibre contains a feature not seen in other fibre types. Besides the core, there are two additional circles called **stress rods**. These stress rods create stress in the core of the fibre such that the transmission of only one polarisation plane of light is favoured. Single-mode fibres experience nonlinearities that can greatly affect system performance.

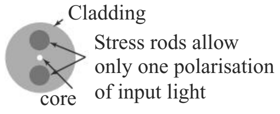


Fig. 6.10 *Cross-section of polarization-maintaining fibre*

THE NUMERICAL APERTURE

6.4

Let us consider a ray incident on the entrance aperture of the fibre making an angle i and an angle of refraction r with the axis as shown in Fig. 6.11, and let θ be the angle of incidence at the core–cladding interface for this ray.

Now the condition for total internal reflection to take place is given by

$$\sin\theta \geq \left(\frac{n^2}{n^1}\right) \tag{6.1}$$

$$\sin\theta = \sin(90° - r) = \cos r \tag{6.2}$$

$$\sin r = \sqrt{(1-\cos^2 r)} = \sqrt{(1-\sin^2\theta)} = \sqrt{1-\frac{n_2^2}{n_1^2}} = \sqrt{\frac{n_1^2-n_2^2}{n_1^2}} \tag{6.3}$$

We know from Snell's law that $(\sin i/\sin r) = n_1$ or $\sin r = (\sin i/n_1)$.

The condition of total internal reflection can be expressed as

$$\sin r \leq \sqrt{\frac{n_1^2-n_2^2}{n_1^2}} \tag{6.4}$$

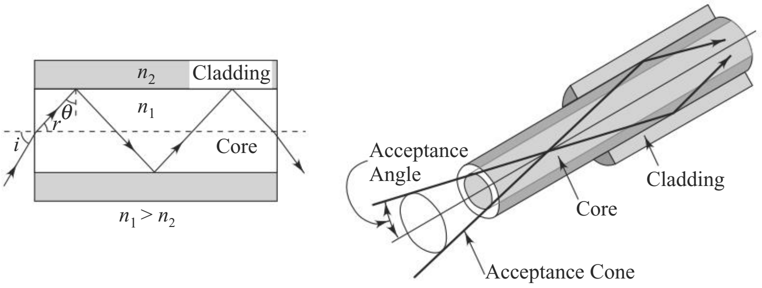


Fig. 6.11 *Numerical aperture and acceptance angle*

From Eqs (6.3) and (6.4) we can write

$$\frac{\sin i}{n_1} \le \sqrt{\frac{n_1^2 - n_2^2}{n_1^2}}$$

$$\therefore \qquad \sin i \sqrt{n_1^2 - n_2^2} \qquad (6.5)$$

Let i_m be the maximum angle of incidence (critical angle) for which total internal reflection may occur. So we have

$$\sin i = \sqrt{n_1^2 - n_2^2} \quad \text{for } n_1^2 - n_2^2 < 1$$

$$= 1 \text{ for } n_1^2 - n_2^2 > 1 \qquad (6.6)$$

Therefore, if a cone of light is incident on one end of the fibre, it will be guided through it provided the semi-angle of the cone is less than i_m.

sin i_m is known as Numerical Aperture (NA) of the fibre. **The numerical aperture represents a measure of the light-gathering power of the fibre**. For all practical situations, $n_1^2 < n_2^2 + 1$.

Thus, one defines the numerical aperture of the fibre by the equation

$$NA = \sqrt{n_1^2 - n_2^2} \qquad (6.7)$$

$$= \sqrt{(n_1 + n_2)(n_1 - n_2)}$$

$$\approx \sqrt{2n_1(n_1 - n_2)} \qquad [\because (n_1 + n_2) \approx 2n_1]$$

Now defining the fractional refractinve index change Δ as

$$\Delta = \frac{n_1 - n_2}{n_1}$$

$$n_1 \Delta = n_1 - n_2$$

$$\therefore \qquad NA = \sqrt{2n_1 \times n_1\Delta}$$

$$NA = \sqrt{2n_1^2 \Delta}$$

$$NA = n_1\sqrt{2\Delta}$$

$$\therefore \qquad NA = n_1\sqrt{2\Delta}$$

Acceptance Angle The other important term is the **acceptance angle.**

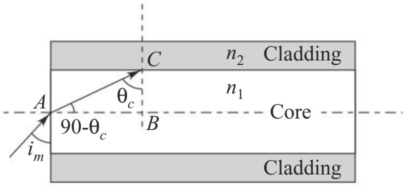


Fig. 6.12 *For derivation of Acceptance angle* i_m

from the Fig. 6.12 and by applying Snell's law at air–fibre interface A and core-cladding interface C we get,

$$n_0 \sin i_m = n_1 \sin (90° - \theta_C) \quad \Rightarrow n_0 \sin i_m = n_1 \cos \theta_C$$

$$n_1 \sin \theta_C = n_2 \sin 90$$

$$\sin \theta_C = \frac{n_2}{n_1}$$

$$n_0 \sin i_m = n_1 \sqrt{1 - \sin^2 \theta_C}$$

$$= n_1 \sqrt{1 - \left(\frac{n_2}{n_1}\right)^2}$$

$$\sin i_m = \sqrt{n_1^2 - n_2^2} \qquad [\because \text{ For air } n_0 = 1]$$

Acceptance angle is i_m given by

$$i_m = \sin^{-1} \sqrt{n_1^2 - n_2^2}$$

The above relation correlates the acceptance angle with the refractive indices of core and cladding. The light that travels within a cone defined by this acceptance angle is known as acceptance cone and the light cannot come out from the fibre.

FIBRE-OPTICS COMMUNICATION SYSTEM

6.5 Optical fibres are extensively used in the field of communication. A basic optical-fibre communication system consists of mainly three things. (i) A transmitter, which transforms an electric signal (information signal) to be transmitted into an optical system. (ii) An information channel (fibre transmission line), (iii) A receiver which converts the optical signal back to the original electric form which conducts the optical signal from the transmitter to receiver. These parts are shown in Fig. 6.13.

Fig. 6.13 *Basic fibre-optic communication system*

A more general block diagram is shown in Fig. 6.13.

The functions of the different parts are as follows.

Message Origin The first step is that the nonelectrical message is converted into an electrical signal. For example, a microphone is used for converting sound waves into current, and a video (TV) camera is used for converting images into currents.

Modulator The second step performs two main functions—the conversion of electrical message into proper format and thereafter impressing this signal onto the wave generated by the carrier source.

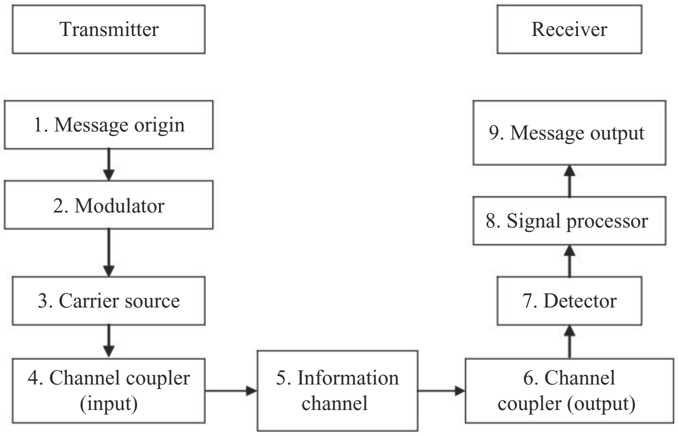


Fig. 6.14 *General block diagram showing a fibre-optic communication system*

The modulation formats are analog and digital. The analog signal is continuous and reproduces the form of the original message, e.g., when a microphone picks up a sound wave, it produces the same shape as the wave itself as shown in Fig. 6.15. So the format of the signal will remain unchanged.

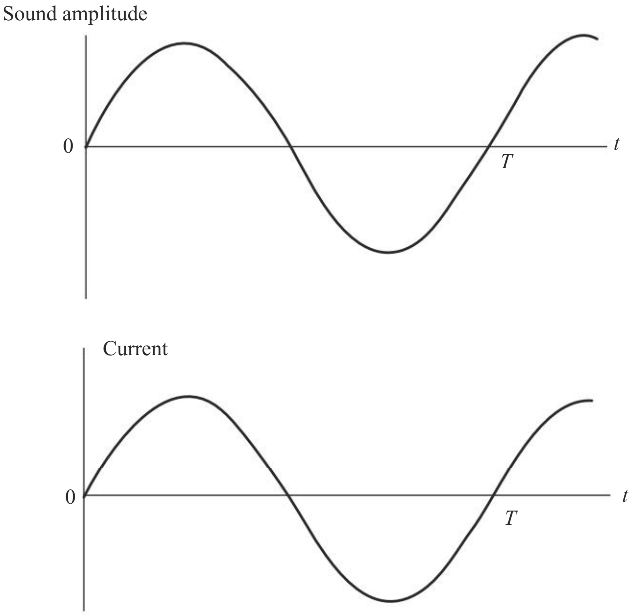


Fig. 6.15 *Analog signal*

The digital modulation involves transmitting information in discrete form. The signal is either ON or OFF as shown in Fig. 6.16.

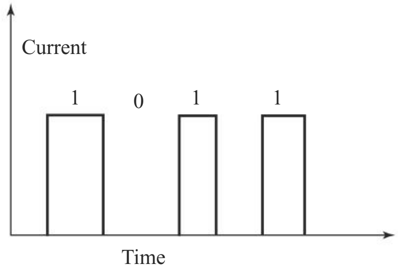


Fig. 6.16 *Digital signal*

Here, the ON state represents a digital 1, and the OFF state represents a digital 0. These states are the binary digits of the digital system. The data rate is the number of bits per second transmitted.

Carrier Source The carrier source generates the wave on which the information is transmitted. This wave is known as the **carrier wave**. We know that the carrier wave is produced by an electronic oscillator in radio-frequency communication. The fibre-optic communication system uses a Laser Diode (LD) or a Light-Emitting Diode (LED) known as an **optical oscillator**. It produces stable, monochromatic waves with sufficient power over large distances.

One should note that the information being transmitted is contained in the variation of optic power. This is called **Intensity Modulation** (IM).

Channel Coupler (Input) The channel power is used to feed power into the information channel, e.g., the channel coupler for a radio or television broadcasting system is an antenna. The antenna transfers the signals from the transmitter onto the information channel (here, atmosphere). Similarly, in a fibre optic system, the coupler transfers the modulated light beam from the source to the optical fibre.

Information Channel The information channel provides a path between the transmitter and receiver. In fibre-optic communication, a glass (or plastic) fibre works as the information channel. The information channels are of two types: (i) unguided, and (ii) guided channels. The atmosphere is an example of unguided channel. Guided channels are a variety of conducting transmission structures such as two-wire lines, coaxial cables and rectangular waveguides.

Channel Coupler (Output) For radio communication systems, an antenna acts as a channel coupler which collects the signal from the information channel and routes it to the receiver. In a fibre-optic communication system, the output coupler directs the light emerging from the fibre onto the light detector, for which a simple butt connection is used.

Detector A detector separates the information from the carrier wave. This process is known as **demodulation**. In a fibre-optic communication system, the optical wave is converted into an electric current by a photodetector which is proportional to the power in the incident optic wave. As the information is contained in the optic power variation, the detector output current contains this information.

Signal Processor Signal processing involves amplification and filtering of undesired frequencies in analog transmission. In a digital system, the processor may include a decision-making circuit in addition to amplifiers and filters.

Message Output The message output presented to a person may be either audio or visual. For hearing, the electrical signal is transformed into sound waves and for a visual image, a cathode ray tube is used. So, suitable transducers are used for specific purpose.

ADVANTAGES OF FIBRE-OPTIC COMMUNICATION SYSTEMS

6.6

(i) It has huge information-carrying capacity.
(ii) Smaller size and weight of the fiber-optic communication systems make them more suitable in space and aeronautical applications.

(iii) As fibre is made of glass (silica), the raw material is available in plenty on the earth.

(iv) The lower cost of cables per unit length compared to that of its metal counterpart increases its use significantly.

(v) Optical fibres (glass or plastic) are insulators. No electric currents flow through them, either due to transmitted signal or due to external radiation striking the fibre. Therefore, no effect of magnetic interruptions is possible. The optic wave within the fibre is trapped, so no leakage is possible during transmission which can interfere with signals in other fibres.

(vi) A fibre is well protected from interference and coupling with other communication channels.

(vii) Fibres will not pick up or propagate electromagnetic pulses caused by nuclear explosions. So it is safe as compared to existing systems.

(viii) For high-voltage lines, a wire communication link could short-circuit the lines by falling across them, causing considerable damage. A fiber-optic system does not face this problem.

(ix) Since the fibres do not radiate energy within them, so it is difficult for an intruder to detect the signals being transmitted. This provides a high degree of security and privacy.

(x) They can tolerate extreme temperature variations before deteriorating, e.g., temperatures approaching 800°C do not affect glass fibres.

(xi) Corrosion due to water or chemicals is minimal for glass.

(xii) There is no requirement of additional equipment to protect against grounding and voltage problems.

APPLICATIONS AND USES

6.7

(i) A fibre optic communication system is capable of handling a large number of channels because it has a large bandwidth that provides wide applications in communication.

(ii) The fibre-optic system maintains high privacy, so it is widely used in defence services.

(iii) Fibre-optic systems are particularly suitable for transmission of digital data like that generated by computers.

(iv) The military applications of fibre-optic communication systems are communications, command and control links on ships and aircraft, and data links for satellite earth stations.

(v) Fibre-optic systems are also used for signaling purposes.

(vi) Fibre-optic sensors are used to measure temperature, pressure, rotary and linear position and liquid levels. Hydrophones are used to measure acoustic disturbances in water. Gyroscopes measure rotational motion.

(vii) Cable television, space vehicles, ships and submarine cables use fibre-optic systems.

(viii) It has wide engineering applications like security and alarm systems, electronic instrumentation systems, industrial automation and process control.

(ix) For medical applications, it is possible to study the interior of lungs and other parts of the body that cannot be viewed directly.

Endoscopy (meaning to see within) is extensively used for diagnostic purposes. An endoscope consists of a bunch of optical fibres that carry light (laser beam) to the inside of the body. The light transfers the image of the inner parts of the body on the screen.

Fibre Endoscope The internal parts of the human body can be easily visualised by using a tubular optical instrument known as a fibre endoscope. The study of tissues and blood vessels far below the skin can be done with this instrument. Different types of endoscopes are used for the treatment of diseases and for surgery.

For example, a **bronchoscope** is used to check the presence of foreign bodies and/or infection in the trachea and larger airways. A **gastroscope** is used to check the presence of tumours, gastric ulcer and gastritis in the stomach. Similarly, a **cytoscope** is used to check the presence of tumours, inflammation and stones in the urinary bladder.

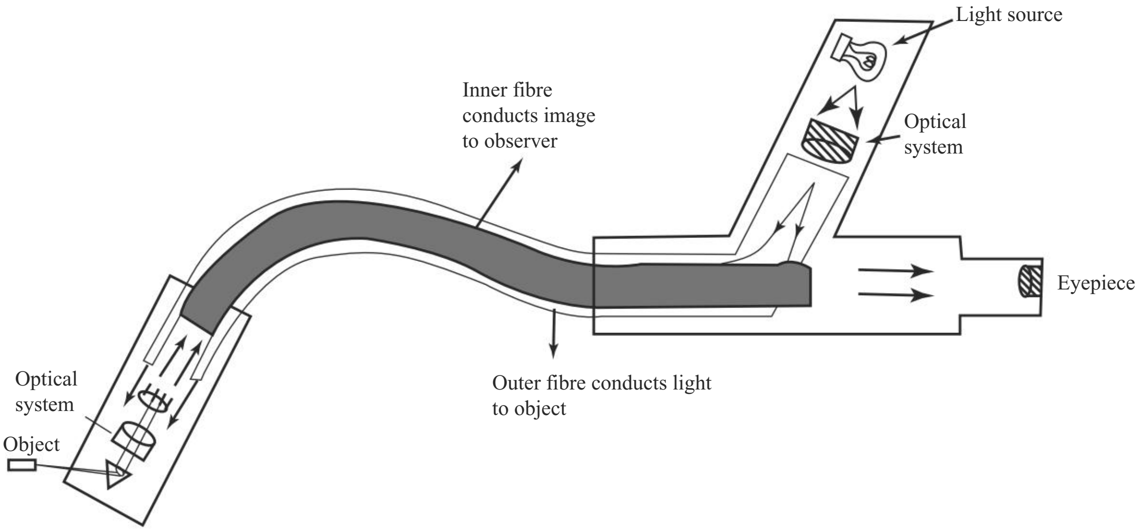


Fig. 6.17 *Fibre endoscope*

The schematic diagram of a fibre endoscope is shown in Fig. 6.16. The fibre endoscope is made up of a bundle of fibres. It consists of two fibres known as inner and outer fibres. At one end of the endoscope, an optical light source is attached with suitable optical arrangements to transmit the light into the outer fibre. Suitable optical arrangements are employed to collect the light beam and to view the object and are placed at the respective fibre ends. The function of the outer fibre is to carry the light and to illuminate the object under study. The function of the inner fibre is to collect the reflected light from the area under study, i.e., from the inner structure of the body.

When the light source is switched on, the optical arrangements fixed at the respective ends of the inner and outer fibres collect the light, transmit and illuminate the area under study and also help to view the object. If a telescope system is introduced at the internal part of the endoscope, it will improve the field of view and image quality.

Concluding Remark Clearly, progress in the technology of communication systems has been remarkable, rapid and impressive. Although fiber-optic systems serve as trunk lines that carry large numbers of voice and data channels between central telephone stations, industry specialists speak wistfully of the ‘last mile’—from the central station to our home. Today’s telephone system spans that last mile with conventional copper-wire equipment, which provides good voice connections but is still inadequate for transmitting large quantities of high-speed data. High-speed data lines for that last mile are available, and many businesses have them, but they are generally more expensive than is practical for home use today. Whatever new technology turns out to provide the last crucial link from individuals to the rest of the world, the research that gives rise to it will have come from scientists who probe beneath the immediate needs of any given industry, investigating seemingly unrelated processes to understand the fundamental nature of the world.

Solved Problems

1 *A step-index fibre has core glass and cladding glass with refractive indices of 1.68 and 1.5. Calculate the numerical aperture. Find the critical angle and acceptance angle for the entrance of light if the fibre is placed in air.*

Solution Given, $n_1 = 1.68$, $n_2 = 1.5$

The numerical aperture is given by

$$NA = \sqrt{n_1^2 - n_2^2} = \sqrt{(1.68)^2 - (1.5)^2} = 0.7565$$

Critical angle

$$i_m = \sin^{-1}\left(\frac{NA}{n}\right) = \sin^{-1}\left(\frac{0.7565}{1}\right) = 49.15^\circ$$

Acceptance angle = critical angle

2 *Calculate the NA the acceptance angle of the fibre having $n_1 = 1.48$ and $n_2 = 1.43$* [*Dec. 2008*]

Solution Given, refractive indices of core and cladding of the fibre is n_1 = and $n_2 = 1.43$

Numerical aperture

$$NA = \sqrt{n_1^2 - n_2^2}$$

$$\therefore \quad NA = \sqrt{(1.48)^2 - (1.13)^3} = 0.38$$

Acceptance angle

$$i_m = \sin^{-1}\sqrt{n_1^2 - n_2^2} = \sin^{-1}(0.38) = 22.4^\circ$$

3 *For a step-index fiber with a core refractive index of 1.48 and a numerical aperture of 0.649, calculate the refractive index of clad-glass material used. Also, calculate the maximum entrance angle of light if the fibre is placed in air.*

Solution Given, $n_1 = 1.48$, $NA = 0.649$

$$NA = \sqrt{n_1^2 - n_2^2}$$

$$(NA)^2 = n_1^2 - n_2^2$$

$$n_2 = \sqrt{n_1^2 - (NA)^2} = \sqrt{(1.48)^2 - (0.649)^2} = 1.33$$

So refractive index of the clad glass is 1.33.

Maximum entrance angle is

$$i_m = \sin^{-1}\left(\frac{NA}{n}\right) = \sin^{-1}\left(\frac{0.649}{1}\right) = 40.46^\circ$$

4 *If the acceptance angle for a given fibre is 68.16°, calculate the maximum entrance angle and numerical aperture. If the cladding glass has a refractive index of 1.52, calculate the refractive index of the core glass.*

Solution Acceptance angle = $2i_m = 68.16°$

So maximum entrance angle = $i_m = 34.08°$

Numerical aperture = $NA = \sin i_m = \sin(34.08°) = 0.56$

$\therefore$ numerical aperture = 0.56

$$NA = \sqrt{n_1^2 - n_2^2}$$

$$n_1 = \sqrt{(NA)^2 + n_2^2}$$

$$n_1 = \sqrt{(0.56)^2 + (1.52)^2} = \sqrt{2.624} = 1.62$$

$\therefore$ So refractive index of the core glass is $n_1 = 1.62$

Multiple-Choice Questions

1. Which part is not included in the fibre-optic communication system?

(a) Transmitter (b) Transformer

(c) Transmission channel (d) Receiver

2. Optical fibre works on the principle of

(a) multiple beam interference

(b) total internal reflection

(c) diffraction at the circular opening

(d) total internal refraction

3. Which one of the following is true if n_1 is the refractive index of core glass and n_2 is the refractive index of the clad glass?

(a) $n_1 < n_2$ (b) $n_1 = n_2$

(c) $n_1 > n_2$ (d) $n_1 \quad n_2$

4. The numerical aperture of a step index fibre when the core refractive index is n_1 and cladding refractive index is n_2 is given by

(a) $NA = \sqrt{n_1^2 - n_2^2}$

(b) $NA = (\sqrt{n_1} - \sqrt{n_2})^2$

(c) $NA = \sqrt{n_1^2 + n_2^2}$

(d) $NA = n_1^2 - n_2^2$

5. If i_m be the maximum angle of incidence for which total internal reflection occurs, then $\sin i_m$ is known as

(a) acceptance angle

(b) the value of the difference of refractive index of core and cladding

(c) numerical aperture

(d) measure of the efficiency of optical power

6. Periodic self-focusing of the optical rays occurs with the

(a) graded-index optical fiber

(b) step-index optical fiber

(c) single-mode optical fiber

(d) multi-mode optical fiber

7. The function of the cladding is

(a) to produce refraction of rays

(b) to produce total internal reflection of rays

(c) to produce transmission of rays

(d) to provide a firm support to the core glass

8. Which one from the following is not a function of the modulator?

(a) It converts the electrical message into proper format.

(b) It impresses the properly formatted electrical message onto the wave generated by the carrier source.

(c) Analog and digital modulation.

(d) Analog to digital and digital to analog conversion of the signal.

9. Optical fibers are basically

(a) insulators (b) conductors

(c) semiconductors (d) superconductors

10. Which of the following is not a part of the structure of optical fibers?

(a) Core (b) Cladding

(c) Primary coating and secondary coating

(d) A source of light

Answers to MCQs

1. (b) 2. (b) 3. (c) 4. (a) 5. (c) 6. (a) 7. (b) 8. (d)
9. (a) 10. (d)

Practice Problems

1. Calculate the numerical aperture of a step-index fibre when the core refractive index is 1.55 and the cladding refractive index is 1.33. **[NA = 0.795]**

2. Determine the numerical aperture of a step-index fiber when the core refractive index is 1.5 and the cladding refractive index is 1.48. Find the maximum angle for entrance of light if the fibre is placed in air. **[(i) NA = 0.24413 (ii) 14.13°]**

3. If the maximum angle of incidence is 45° for entrance of light for a fibre placed in air, calculate the numerical aperture. What is the acceptance angle? **[(i) NA = 0.7071 (ii) 90°]**

Short-answer Questions

1. If by fabrication mistake the glass of cladding and core are interchanged, what will happen to the transmission of optical energy?

2. What are the three main parts of a communication system?

3. What is fibre optics?

4. Explain briefly the principle of optical fiber communication.

5. What is the function of the cladding?

6. Why should clad glass have less refractive index than core glass?

7. What is the main difference between the step-index optical fiber and graded-index optical fibre?

8. What is graded-index optical fibre?

9. What is step-index optical fibre?

10. Draw the path of rays in the graded-index optical fibre.

11. Draw the path of rays in the step-index optical fibre.

12. What is single-mode fibre?

13. What do you mean by multi-mode fibre?

14. Define numerical aperture. What does it represent?

15. What is the specific use of a mono mode fibre?

16. State the names of different parts of a fibre optic communication system.

17. What is the function of a modulator?

18. What is the function of a carrier source?

19. What is the function of a channel coupler?

20. What are the types of information channels?

21. Briefly discuss about detectors.

22. How do optical fibres offer high degree of security and privacy?

Long-answer Questions

1. Explain the principle of the optical fibre as a wave-guide for light.

2. Describe the structure of a typical optical fibre used in practice. Describe the step-index and graded-index optical fibres.

3. What is Numerical Aperture (NA) of an optical fibre? What does the numerical aperture signify?

4. Describe schematically the basic elements of a fibre-optic communication system.

5. What are the advantages of using fibre-optic communication system?

6. Enumerate some applications of fibre-optic communication system.

Questions From Previous Exams

(i) Define total internal reflection. **(GTU-Jan′ 2009)**

The light approaches the wall of the fibre with an angle of incidence F that is greater than the critical angle; the incident light not get refracted but reflects inside the denser medium.

(ii) State the main components of optical fibre communication system. **(GTU-Jan′ 2009)**

The important components of a fibre optic communication system are transmitter, fiber optic cable and receiver.

(iii) Discuss the advantage of optical fibre communication system over the conventional coaxial communication system. Refer Section: 6.1 **(GYU-Jan′ 2009)**

(v) Calculate the *NA*, the acceptance angle of the fibre having $n_1 = 1.48$ and $n_2 = 1.43$. **(GTU-Jan′ 2009)**

Given, $n_1 = 1.48$ and $n_2 = 1.43$

$$NA = \left(n_1^2 - n_1^2\right)^{1/2}$$

$$NA = 0.3872$$

Acceptance angle $i_m = \sin^{-1} NA,$

$$i_m = \sin^{-1}(0.3872) = 22°46,$$

(vi) What is Kevlar? **(GTU-Jan′2010)**

It is a yarn type of material having high tensile strength. It gives an additional strength to cable.

(vii) Describe the construction of fibre optic cable and compare the advantage of fibre optic cable over metallic cable. Refer Section: 6.3 and 6.1 **(GTU-Jan′ 2010)**

(viii) Give the conditions to be satisfied for total internal reflection for an optical fibre. **(GTU-Jan′ 2010)**

The refractive index n_1 of the core must always be greater than the refractive index n_2 of the cladding.

The light must approach the wall with an angle of incidence φ that is greater than the critical angle φ_C.

(ix) An optical fibre fore and its cladding have refractive indexes of 1.545 and 1.495 respectively. Calculate the critical angle φ_C, acceptance angle $\varphi_{in(max)}$ and numerical aperture.
(GTU-Jan′ 2010)

Given

$n_1 = 1.545$

$n_2 = 1.495$

Critical angle $\theta_C = \sin^{-1} \theta_C = n_2/n_1$

$\theta_C = \sin^{-1} (0.967)$

$\theta_C = 75° 14'$

Acceptance angle $i_m = \sin^{-1} NA = \sin^{-1}(0.3898) = 22° 56'$

Numerical aperture = $NA = \left(n_1^2 - n_2^2\right)^{1/2} = 0.3898$

(x) Define fibre optic system **(GTU-June′ 2010)**

The important components of fibre optic communication systems are transmitter, fiber optic cable and receiver.

(xi) An optical fibre has refractive index of core and cladding is 1.514 and 1.48 respectively. Calculate the acceptance angle and the fractional index change. **(GTU-June′ 2010)**

Given, $n_1 = 1.514$, $n_2 = 1.48$

$\Delta = (n_1 - n_2)/n_1 = 0.022$

Acceptance angle $i_m = \sin^{-1} NA = \sin^{-1}(0.316) = 18° 25'$

Numerical aperture = $NA = (n_1^2 - n_2^2)^{1/2} = 0.316$

(xii) Describe the construction of fibre optic cable. Refer Section: 6.36 **(GTU-June′ 2010)**

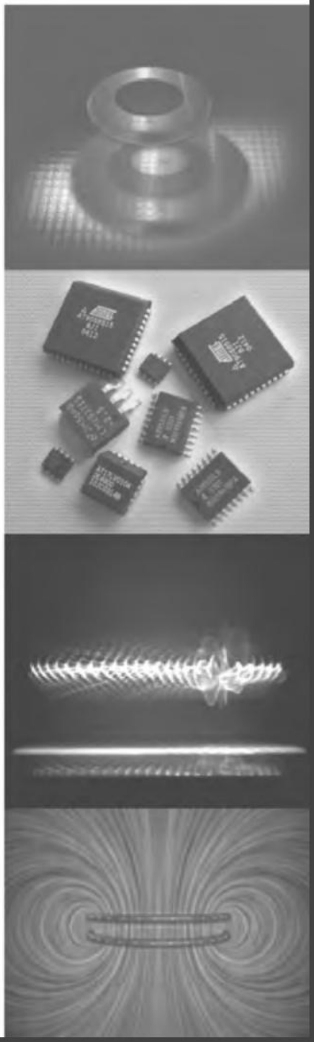
(xiv) Give the difference between step index fibre and graded index fibre. Refer Section: 6.3
(GTU-June′ 2010)

The McGraw·Hill Companies

Conducting Materials

7

Often We wonder, why does wood not conduct heat as well as electricity but metals conduct both? Why do some metals have better conductivity than others? What can we do to improve their conductivity? To learn the answers to these questions, you must read this chapter. The knowledge of the behaviour of electrons in solids is one of the keys to understand the behaviour of materials. The electron theory of solids is capable of explaining optical, electrical, magnetic and thermal, properties of materials. It provides the important fundamentals to the technology of today and of the future.



INTRODUCTION

7.1 Solids differ in their physical properties like electrical, thermal, optical, magnetic and mechanical. It would be interesting to understand the reasons for this difference and thereby to use them for specific purposes. Knowledge of these properties is used in technological advancement.

For understanding the electronic properties of materials, three approaches are well established. (a) The laws of physics which were eventually discovered were empirically derived .They considered only macroscopic quantities and inter-related experimental data, e.g., Ohm’s law, Newton’s laws (no oversimplification while their interpretation). (b) In order to refine the understanding of the properties of materials and these laws, atomistic principles were introduced, e.g., the classical electron theory. (c) For a further refinement to resolve the shortcomings of classical theory, ‘Quantum theory’ was developed.

In this chapter we will mainly consider the electrical and thermal properties of materials. Metals and alloys are together known as conducting materials, and because of their striking properties they are widely used in many applications. Ohm gave the well-known law, $V = RI$, but only after the discovery of the electron by J J Thomson. Drude and Lorentz applied the kinetic theory of gases to derive Ohm's law based on classical theory. Sommerfeld further refined it to remove the drawbacks of the classical theory by introducing the quantum theory.

BRIEF SURVEY OF MATERIALS

7.2

The electrical conductivity σ of different materials spans about twenty-five orders of magnitude as shown in Fig. 7.1. This is the largest known variation in a physical property. It is generally accepted that in metals and alloys, the electrons, particularly the outer or valance electrons, play a significant role in electrical conduction. Therefore, it is most appropriate to make use of the electron theory to understand the electrical conduction properties of these materials.

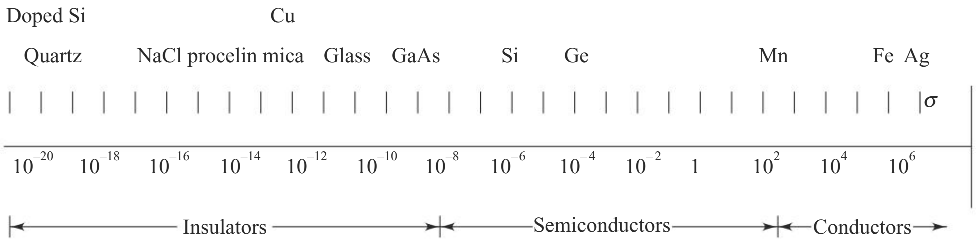


Fig 7.1 *Room temperature conductivity (Ohm cm)$^{-1}$ of various materials.*

CONDUCTION IN METALS

7.3

The first step towards an understanding of electrical conduction is to postulate a free 'electron gas' or 'plasma', consisting of the valance electrons of the individual atoms in a crystal. Then ***classical free electron theory given by Drude–Lorentz*** may be considered as follows.

(a) Free Electron Model

- Conducting materials like metals consist of valence electrons. These valence electrons are responsible for electrical and thermal conduction in the bulk state of matter.
- The valence electrons become *free* in conducting materials and keep moving about randomly within the material just like molecules in a gas. Therefore, they are referred as **free electrons** or **conduction electrons.**
- The collection of free electrons is referred as **free electron gas (Fig. 7.2).**
- The movement of free electron is restricted to the boundaries of the material.
- These free electrons experience resistance during acceleration (motion) produced by the application of field on the conduction electrons, which brings these electrons to settle into a state of constant velocity, proportional to this applied field.

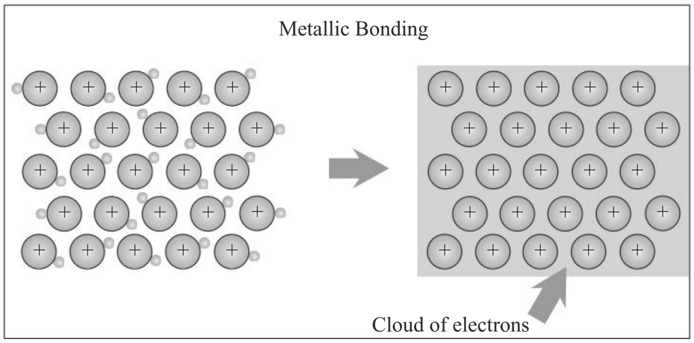


Fig. 7.2 *Free electron gas in a metal*

(b) Assumptions of the Classical Free-Electron Theory

- The motion of electrons in conducting materials obeys the classical Maxwell–Boltzmann distribution.
- The kinetic energy of electrons can be given by $\frac{1}{2} mv^2 = 3/2\, kT$.
- The free electrons in the electron gas wander freely in the metal without any mutual interaction.
- The free electrons collide with positive ions in the lattice and also among themselves. These collisions are elastic and, therefore, there is no loss of energy.
- In the absence of an electric field, the electrons move in random directions so that their individual velocities cancel and no net velocity results; hence, no current flows.
- When an electric field is applied, the free electrons are accelerated with an electric force ***eE***. They move in opposite direction to the electric field while maintaining their random motion.
- Monovalent metals such as sodium contribute one electron per atom to the plasma (electron gas).

The number of atoms, n, per cubic centimetre will be

$$n = N_a\, \delta / M \tag{7.1}$$

where, N_a = Avogadro's number,
δ = density of material,
M = atomic mass of the element

Typically, 10^{22} to 10^{23} free electrons /cm^3 for a monovalent atom is available.

(c) Drift Motion of Electrons

The free electrons keep moving randomly in all directions throughout the lattice structure of the material due to thermal energy. The *average* ***thermal speed*** *is very high* (~10^6 m/s).The direction of motion of each free electron changes on every collision and the electron moves along a zigzag path. But their directions of motion are so randomly distributed that the *average* ***thermal velocity*** *of the electron is zero*. Therefore, there is no flow of current due to the thermal motion of electrons.

When electric field is applied by connecting a battery across the conductor, the equilibrium condition is disturbed. The electric field accelerates the electrons. They move in a direction to that of electric force on electrons (in the opposite direction to applied electric field). *This directional motion of electrons due to the* *applied electric field is called drift motion*. Whenever an electron encounters a collision, it loses this drift velocity completely. After that, the electron gets accelerated once again and loses its velocity at the next collision. The process thus goes on repeating and the electron moves on an average with a mean drift velocity υ_d. Thus, **drift velocity** is defined as *the average velocity with which free electrons get drifted towards the positive end of the conductor under the influence of an external electric field.* Typically, it is of the order of 10^{-3} m/s.

BASIC TERMINOLOGY OF ELECTRICAL CONDUCTIVITY

7.4

- **Ohm's Law** When an electric field is applied across the conductor, electric current flows through it and it is directly proportional to the voltage (potential difference)**.** This law is valid at constant temperature and under identical physical conditions of the conductor, i.e.,

$$V = R\,I \tag{7.2}$$

where, V = applied voltage in **volts**, I = current in **amperes**, R = resistance of the conductor in **ohms**

- ***Resistance R*** It is a geometry (shape and size) and property-dependent factor of the material.

$$R = \frac{\rho l}{A} = \frac{l}{\sigma A} \tag{7.3}$$

where, l is the length of the conductor in **m**, A is the cross-sectional area of the conductor in $\mathbf{m^2}$, ρ is the electrical resistivity in **ohm-metre** and σ is the electrical conductivity which is the reciprocal of the resistivity ρ. Its unit is $\mathbf{(ohm\text{-}metre)^{-1}}$ or $\mathbf{mho\text{-}m^{-1}}$, mho is called as **siemens *S*.**

- ***Current Density J*** It is defined as the current flowing per unit cross-section of the conductor, when the cross-sectional plane is held normal to the flow of the electric current. Its unit is $\mathbf{A/m^2}$

$$J = \frac{I}{A}$$

- ***Electric Field E*** It is the applied potential drop **(voltage)** per unit length of the conductor of uniform cross-section. Its unit is $\mathbf{Vm^{-1}}$**.**

$$E = \frac{V}{l}$$

- ***Electrical Conductivity σ*** We can obtain the microscopic form of Ohm's law ($J = \sigma E$) and hence it defines electrical conductivity σ.

We know that, $V = IR$ and $R = \dfrac{l}{\sigma A}$

So, $V = \dfrac{lI}{\sigma A}$

Or, $\dfrac{I}{A} = \sigma \dfrac{V}{l}$

Or, $J = \sigma E$

∴ $\sigma = J/E$

Thus, electrical conductivity can be defined as the current density in a conductor per applied electric field.

- **Mobility** μ The drift velocity υ_d is proportional to the applied electric field.

  ∴ $$\upsilon_d \, \alpha \, E$$

  ∴ $$\mu_d = \mu E$$

  ∴ $$m = \frac{v_d}{E}$$

  Thus, we define electron mobility of the conductor as the drift velocity of the electrons per unit electric field. It indicates the ease with which electrons move in a conductor. In metals, it is of the order of 10^{-2} **m^2/Vs**
- ***Relaxation Time*** τ It is the duration in which the drift velocity of an electron decays to $1/e$ times of its initial velocity. It shows the time taken by the electrons in a conductor to return from non-equilibrium condition to equilibrium condition, after the electric field is switched off. It is also the average time lapsed between two successive collisions. It is inversely proportional to υ_d. Its value is of the order of 10^{-14}s.

ELECTRICAL CONDUCTIVITY σ

7.5

7.5.1 Relation with Relaxation Time

In the absence of an electric field, motion of the free electrons is random just like molecules of a gas in a container, i.e., without any particular direction. But, when an electric field is applied to a conductor, they move in such a manner as to drift slowly in the opposite direction to that of the applied electric field, with an average velocity known as υ_d.

When an electric field is applied, the free electrons in any conductor experience a force eE. It is the cause for drift motion and acceleration of electrons in a conductor.

$$F = eE$$

$$ma = eE$$

or, $$a = \frac{eE}{m}$$

Consider an electron that has just collided with an ion core. The collision disturbs the tendency of drift of the electron in the presence of an electric field. In the successive collision, its velocity changes and, therefore, we will consider average drift velocity υ_d.

$$\upsilon_d = a\,\tau$$

$$v_d = \frac{eE\tau}{m}$$

Now mobility is $$\mu = \frac{v_d}{E}$$

$$\frac{v_d}{E} = m = \frac{e\tau}{m}$$

If n is the charge carrier density or concentration or number of *charge carriers per unit volume* in a conductor of length l, with cross-sectional area A then the current flowing through the conductor is,

$$I = \frac{\text{Total charge flowing through the conductor}}{\text{Time}}$$

$$I = \frac{neAl}{t} v_d$$

$$I = neAv_d$$

but, $$J = \frac{I}{A}$$

∴ $$J = \frac{neA}{A} v_d$$

∴ $$J = nev_d$$

∴ $$J = ne\frac{eE\tau}{m}$$

∴ $$J = \frac{ne^2\tau}{m} E$$

But, $$J = \sigma E$$

∴ $$\sigma E = \frac{ne^2\tau}{m} E$$

Therefore, $$\sigma = \frac{ne^2\tau}{m} \tag{7.4}$$

$$\sigma = \frac{ne^2\tau}{m} = \frac{ne}{1}\frac{e\tau}{m} = ne\mu$$

Resistivity, $$\rho = \frac{1}{ne\mu}$$

Thus, **electrical conductivity σ is directly proportional to the relaxation time and also directly proportional to carrier concentration.**

7.5.2 Relation with Temperature

The mean free time or relaxation time ρ can also be expressed in terms of mean free path λ and average thermal velocity v as,

$$\tau = \frac{\lambda}{v}$$

The kinetic energy of an electron based on kinetic theory is given by,

$$\frac{1}{2}mv^2 = \frac{3}{2}k_BT$$

where k_B is the Boltzmann constant

∴ $$m = \frac{3k_BT}{v^2} \quad \text{But, } \sigma = \frac{ne^2}{m}\tau$$

∴ $$\sigma = \frac{ne^2\tau v^2}{3k_BT}$$

$$\therefore \qquad \sigma = \frac{ne^2}{3k_B T v}\lambda v^2$$

$$\therefore \qquad \sigma = \frac{ne^2}{3k_B T}\lambda v \tag{7.5}$$

Thus, it is clear that the **electrical conductivity σ of the conductor decreases with increase in temperature.**

THERMAL CONDUCTIVITY

7.6

Heat conduction can be described as the transfer of heat (thermal) energy from the hotter part to the colder part of a material. Let us understand the mechanism involved and hence obtain the expression for thermal conductivity of conducting materials.

- Heat transfer in solids may be provided through free electrons as well as through phonons (lattice vibrations).
- Since there are no free electrons in insulators, heat transfer must be through phonons only. But, in metals and alloys, the heat transfer is mainly due to free electrons [contribution from phonons (elastic wave energy due to lattice vibrations in crystalline solids) is neglected in these types of materials].
- To maintain charge neutrality of the material, equal numbers of electrons travel from hot to cold and from cold to hot parts.
- The hot parts of the conductor possess and transfer a high energy, whereas the cold parts possess and transfer lower energy.
- The heat transfer from hot to the cold parts of the conductor is directly proportional to the difference in the energies of the electrons and, hence, the temperature gradient $\frac{dT}{dx}$.
- Time rate of heat flow per unit cross-section area through which it flows is directly proportional to the temperature gradient $\frac{dT}{dx}$.

Thus,

$$\frac{\Delta H}{A.\Delta t} \alpha \frac{dT}{dx}$$

where, $\frac{\Delta H}{A.\Delta t} = \phi_t$ = thermal flux, i.e., time rate of heat flow per unit cross-sectional area .

$$\therefore \qquad \phi_t = -K\frac{dT}{dx}$$

where K is the proportionality constant called the **thermal conductivity**.

The negative sign in the above equation indicates that the heat flows from the hot part to the cold part of the conductor. Thus, thermal conductivity K of any solid can be defined as the quantity of heat flowing per unit time through a unit cross-sectional area and maintaining a unit temperature difference across the body.

7.6.1 Expression for Thermal Conductivity K

Consider a thermal conducting material with uniform cross-section of area A as shown in Fig.7.2. Let the surface P be at a higher temperature T, and the surface Q be at a lower temperature $T - dT$. Let the distance of separation between the surfaces be λ (mean free path). The electrons conduct heat from P to Q. During collision, the electrons near P lose their kinetic energy whereas the electrons near Q gain the energy. Let the density of electrons be n, the average thermal velocity is v.

Now on the basis of kinetic theory of gases (for electron gas), the average kinetic energy of an electron at surface P is $\frac{3}{2} k_B T$.

Similarly, the average kinetic energy of an electron at surface Q is $\frac{3}{2} k_B (T - dT)$.

Therefore, the excess kinetic energy transferred by electrons from P to $Q = \frac{3}{2} k_B dT$

Electrons have equal probability to move in all the six directions ($\pm x$, $\pm y$, $\pm z$). So, the number of electrons crossing unit area in unit time from P to $Q = \frac{1}{6} nv$

Therefore, the excess energy transferred from P to Q per unit area in unit time

$$= \frac{1}{6} nv \times \frac{3}{2} k_B dT = \frac{1}{4} nv k_B dT$$

Similarly, the deficiency of energy carried from P to Q per unit area in unit time

$$= -\frac{1}{4} nv k_B dT$$

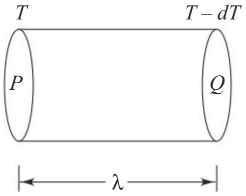


Fig. 7.3 *Thermal conducting material with uniform cross sectional area A*

The net energy transferred from P to Q per unit area per unit time is the rate of heat transfer H. Then,

$$H = \frac{1}{4} nv k_B dT - \left(-\frac{1}{4} nv k_B dT\right)$$

$$H = \frac{1}{2} nv k_B dT \qquad (7.6)$$

According to the definition of thermal conductivity,

$$K = \frac{H}{\left(\frac{dT}{dx}\right)}$$

$$\therefore \qquad K = \frac{H}{\left(\frac{dT}{\lambda}\right)} \tag{7.7}$$

But in this case $dx = \lambda$

Substituting for H from Eq. (7.6) Eq. (7.7), we get,

$$K = \frac{\frac{1}{2} nvk_B}{\frac{dT}{\lambda}} dT \tag{7.8}$$

$$\therefore \qquad K = \frac{1}{2} nvk_B \lambda \tag{7.9}$$

Thus, **the thermal conductivity K is directly proportional to carrier concentration of the material, thermal velocity and mean free path**.

Temperature drop per unit heat flow $R_T = \frac{dT}{H}$ is known as **thermal resistance. It depends on both, the geometry as well as material.** It is also expressed in terms of thermal conductivity K, length of the sample l and area of cross-section A as $RT = \frac{l}{KA}$ in kW^{-1}.

WIDEMANN–FRANZ LAW

7.7

It is now well established that free electrons are responsible for both thermal and electrical conduction in the conducting materials such as metals and alloys. Then, it is reasonable to believe that the ratio of their conductivity $\frac{K}{\sigma}$ should be universal constant (same for all conductors). In1853, **Widemann and Franz** observed experimentally and established the relation between these two conductivities. This law states that *The ratio of thermal to electrical conductivity of a conducting material is proportional to the absolute temperature and it is constant at a given temperature*.

That means,

$$\frac{K}{\sigma} \alpha T$$

$$\frac{K}{\sigma} = LT \quad \text{where } L \text{ is called the } \textbf{\textit{Lorentz number.}}$$

Its value is $1.22 \times 10^{-8}\ \text{W}\Omega\ \text{K}^{-2}$

We know that $\sigma = \frac{ne^2 \lambda v}{3k_B T}$ and $K = \frac{1}{2} nvk_B \lambda$

$$\therefore \qquad \frac{K}{\sigma} = \frac{nvk_B \lambda}{2 \frac{ne^2 \lambda v}{3k_B T}}$$

$$= \frac{3}{2}\frac{k_B^2}{e^2}T$$

$$= LT$$

where, Lorentz number can be obtain by substituting the values of constants.

$$L = \frac{3}{2}\left(\frac{k_B}{e}\right)^2 = \frac{3}{2}\left(\frac{1.38 \times 10^{-23}}{1.602 \times 10^{-19}}\right)^2 = 1.12 \times 10^{-8}\ \text{W}\Omega\text{K}^{-2}$$

This classical value is only half of the experimental value. Using quantum theory, its value is obtained to be 2.45×10^{-8} WΩ K^{-2}, which agrees well with the experimental value.

This indicates that the assumption made in classical calculations that all the free electrons of a metal participate in thermal conductivity is not justifiable and requires some modification. The quantum theory takes care of this and is most suitable for elimination of such discrepancies between theory and experimental results. Table 7.1 shows the electrical and thermal conductivities of some metals at 300 K. Theoretical calculations suggest that the Lorentz number L which depends only on the universal constants k_B and e, should be the same for all metals. This conclusion suggests that the electrical and thermal conductivities are closely related, which is to be expected, since both electrical and thermal current are carried by the same agent electrons. Experimental values of L at 0°C and 100°C given in Table 7.1 are in good agreement with the theoretical predicted value.

Table 7.1 *Electrical and thermal conductivity of some metals at 300 K*

Sr. No.	Metal	$\sigma\ (\Omega m)^{-1} \times 10^7$	$K\ (Wm^{-1}K^{-1})$	Lorentz Number (L) $(W\Omega K^{-2}) \times 10^{-8}$ 0°C	100°C
1	Au	4.47	295	2.35	2.4
2	Ag	6.22	423	2.31	2.37
3	Cu	5.82	387	2.23	2.33
4	Al	3.65	230	2.01	2.06
5	Cd	1.3	102	2.42	2.44
6	Pb	0.48	35	2.47	2.56
7	Zn	1.69	130	2.31	2.33

SUCCESS AND SHORTCOMINGS OF FREE-ELECTRON THEORY

7.8

Success of Free-Electron Theory

- It explains thermal and electrical conductivities of conductors.
- Widemann–Franz law can be verified.
- Optical properties can be explained.
- Ohm's law can be verified.

Shortcomings of Free-Electron Theory

- It is not able to explain the electrical conductivity of semiconductors and insulators.
- The phenomena such as photoelectric effect, Compton effect and black-body radiation could not be explained by this theory.
- According to this theory, the Lorentz constant should be same at all temperatures for any conducting materials, but it does not hold true at low temperatures.
- The Theoretical value of the molar specific heat of a gas at constant volumeand electronic specific heat of metals is not in agreement with their experimental value.
- Mean free path of electrons in metals is of the order of 3 Å. However, the experimentally measured value is of the order of 50 Å. This suggests that electrons pass a long distance through lattice without collisions which is contrary to the assumption of this theory that the origin of resistivity is due to frequent collisions of electrons with the lattice ions.
- It cannot explain the classification of materials into conductors, semiconductors and insulators.

List of Important formulae for Problems

1. The number of atoms, n, per cubic centimetre will be

$$n = N_a \delta/\mathrm{M}$$

where, N_a = Avogadro, number,
δ = density of material,
M = atomic mass of the element

2. *Resistance* $R = \frac{\rho l}{A} = \frac{l}{\sigma A}$

   where, is the length of the conductor in **m**, A is the cross-sectional area of the conductor in $\mathbf{m^2}$, ρ is the electrical resistivity in **ohm-metre** and σ is the electrical conductivity which is reciprocal of resistivity

3. Mobility of electrons, $\mu = \frac{e\tau}{m}$, $\mu = \frac{v_d}{E}$
4. Electrical conductivity, $\sigma = \frac{ne^2\tau}{m} = n\,e\,\mu$, $\sigma = \frac{ne^2}{3k_BT}\lambda v$
5. Resistivity, $\rho = \frac{1}{ne\mu}$, $\rho = \frac{m}{ne^2\tau}$
6. Thermal conductivity K,

$$K = \frac{1}{2} nvk_B\lambda$$

7. Thermal resistance $R_T = \frac{l}{KA}$
8. Widemann–Franz Law, $\frac{K}{\sigma} = LT$
9. Current density, $J = nev_d$

Solved Problems

1 *Consider a copper wire having cross-sectional area of 10^{-5} m^2 and carries a current of 1 A. Calculate the drift velocity of electrons in a copper assuming each copper atom contribute one electron to the free electron gas. [Given: $\delta_{Cu} = 8.969 \times 10^3$ kg m^{-3} At.wt. of Cu = 63.54 kg/kmol]*

Solution The number of atoms, N_O, per cubic centimetre will be

$$n = N_a \delta/M$$

$$n = \frac{N_a \delta}{M}$$

$$= \frac{6.02 \times 10^{26} \times 8.969 \times 10^3}{63.54}$$

$$= 8.5 \times 10^{28} \text{ electrons/m}^3$$

Current density is $J = nev_d$

Drift velocity $v_d = \frac{J}{ne}$

$$= \frac{I}{neA}$$

$$= \frac{1}{8.5 \times 10^{28} \times 1.6 \times 10^{-19} \times 1 \times 10^{-5}}$$

$$= 7.3 \times 10^{-6} \text{ m/s}$$

2 *An aluminium conductor has electron concentration of 18.06×10^{28} m^{-3}. If the drift velocity of electrons in a conductor is 0.625 m/s, what is the current density? Hence calculate the mobility of the electrons. [Given: electrical conductivity of aluminium is 3.65×10^7 $\Omega^{-1}m^{-1}$.*

Solution Given: $n = 18.06 \times 10^{28}$ m^{-3}, $v_d = 0.625$ m/s, $\sigma = 3.65 \times 10^7$ Ω^{-1}m^{-1}

Current density J:

$$J = nev_d$$

$$= 18.06 \times 10^{28} \times 1.6 \times 10^{-19} \times 0.625$$

$$= 1.806 \times 10^{10} \text{ A/m}^2$$

Mobility μ: $\sigma = n e \mu$

$$m = \frac{\sigma}{ne}$$

$$= \frac{3.65 \times 10^7}{18.06 \times 10^{28} \times 1.6 \times 10^{-19}}$$

$$= 1.26 \times 10^{-3} \text{ m2 V}^{-1}\text{s}^{-1}$$

3 *The mobility of electrons in a copper is $4.16 \times 10^{-3} m^2/Vs$, and the drift speed is $1.75 \times 10^{-3} m/s$ if electric field E is applied. Calculate the strength of the field.*

Solution Electric Field from the following relation:

$$v_d = \mu E$$

$$E = \frac{v_d}{\mu}$$

$$= \frac{4.16 \times 10^{-3}}{1.75 \times 10^{-3}}$$

$$= 2.38 \text{ V/m}$$

4 *The electrical conductivity of Zn is $1.69 \times 10^7 \, \Omega^{-1} m^{-1}$ and conduction electrons per m^3 is $13.1 \times 10^{28} m^{-3}$. Calculate the relaxation time of the conduction electrons.*

Solution

$$\sigma = \frac{ne^2\tau}{m}$$

$$\tau = \frac{m\sigma}{ne^2}$$

$$= \frac{9.1 \times 10^{-31} \times 1.69 \times 10^7}{13.1 \times 10^{28} \times (1.6 \times 10^{-19})^2}$$

$$= 4.58 \times 10^{-15} \text{ s}$$

5 *Calculate the electrical conductivity of copper. Given atomic weight, density and relaxation time as 63.5, 8.9×10^3 kg/m^3 and 2.48×10^{-14} s respectively. (June 2009, 5 marks)*

Solution Given: $M = 63.5$, $\delta = 8.9 \times 10^3$ kg/m^3 and $\tau = 2.48 \times 10^{-14}$ s

$$n = \frac{N_a \delta}{M}$$

$$= \frac{6.023 \times 10^{26} \times 8.9 \times 10^3}{63.5}$$

$$= 8.44 \times 10^{28} \text{ electrons/m}^3$$

Electrical conductivity σ:

$$\sigma = \frac{ne^2\tau}{m}$$

$$= \frac{8.44 \times 10^{28} \times (1.6 \times 10^{-19})^2 \times 2.48 \times 10^{-14}}{9.1 \times 10^{-31}}$$

$$= 5.88 \times 10^7 \, \Omega^{-1} \, \text{m}^{-1}$$

6 *Calculate the thermal conductivity of copper, if its electrical conductivity at the temperature 300 K is* $5.82 \times 10^7\ (\Omega\,m)^{-1}$*. (Given: Lorentz number is* $2.33 \times 10^{-8}\ W\Omega K^{-2}$*)*

Solution Given: $T = 300$ K, $\sigma = 5.82 \times 10^7\ (\Omega\ \text{m})^{-1}$

$$\frac{K}{\sigma} = LT$$

$$\begin{aligned} K &= \sigma LT \\ &= 5.82 \times 10^7 \times 2.33 \times 10^{-8} \times 300 \\ &= 406.82 \text{ W/mK} \end{aligned}$$

7 *Calculate the thermal conductivity of Ag at 300 K having free electron density 5.85 × 1028 m–3. The thermal velocity of silver at 300 K is 1.4 × 105 m/s and the mean free path of 7.48 nm. [Given: kB = 1.38 ×* 10^{-23} JK^{-1}*]*

Solution Given: $T = 300$ K, $n = 5.85 \times 10^{28}\ \text{m}^{-3}$, $v = 1.4 \times 10^5$ m/s, $\lambda = 7.48$ nm

$$\begin{aligned} K &= \frac{1}{2} n v k_B \lambda \\ &= \frac{1}{2} \times 5.85 \times 10^{28} \times 1.4 \times 10^5 \times 1.38 \times 10^{-23} \times 7.48 \times 10^{-9} \\ &= 422.7 \text{ Wm}^{-1}\text{K}^{-1} \end{aligned}$$

8 *An aluminium disk of electrical conductivity 3.65 × 107 (W m)–1 conducts heat from a heat source to a heat sink at a rate of 10 W. Its radius is 13 × 10–3 m and thickness is 35 × 10–3 m. Calculate the thermal conductivity and thermal resistance at 300 K. [Given Lorentz no. L = 2.06 ×* 10^{-8} $W\Omega K^{-2}$*]*

Solution Given: Electrical conductivity,

$$\sigma = 3.65 \times 10^7\ (\Omega\ \text{m})^{-1}$$

Temperature $T = 300$ K,

Radius is 13×10^{-3} m, so area of cross-section $A = 5.309 \times 10^{-4}\ \text{m}^2$ and thickness is 35×10^{-3} m

Thermal conductivity K:

$$\begin{aligned} K &= \sigma LT \\ &= 3.65 \times 10^7 \times 2.06 \times 10^{-8} \times 300 \\ &= 225.57 \text{ Wm}^{-1}\text{K}^{-1} \end{aligned}$$

Thermal resistance R_T:

$$R_T = \frac{l}{KA} = \frac{35 \times 10^{-3}}{225.57 \times 5.309 \times 10^{-4}} = 0.2922 \text{ KW}^{-1}$$

Multiple-Choice Questions

1. Unit of Lorentz number is
   - (a) watt.ohm/deg^2
   - (b) watt/ohm. deg^2
   - (c) no unit
   - (d) watt. deg^2/ohm

2. The ratio of thermal and electrical conductivity of any conductor is
   - (a) inversely proportional to temperature
   - (b) directly proportional to temperature
   - (c) negative
   - (d) independent of temperature

3. The limitations of classical free-electron theory is due to application of
   - (a) Maxwell–Boltzmann statistics
   - (b) Fermi–Dirac statistics
   - (c) Bose–Einstein statistics
   - (d) none of the above

4. If the material has a conductivity 10^{-12} (Ohm cm)$^{-1}$, it will be
   - (a) a conductor
   - (b) an insulator
   - (c) a semiconductor
   - (d) a superconductor

5. Mean free path of electrons in metals is of the order of
   - (a) Å
   - (b) cm
   - (c) tens of Å
   - (d) hundreds of Å

6. $\frac{\sigma}{k}$ of any conductor is
   - (a) inversely proportional to temperature
   - (b) directly proportional to temperature
   - (c) negative
   - (d) independent of temperature

7. Electrical or thermal conductivity of a conductor like metals and alloys is due to
   - (a) all the electrons present in it
   - (b) only free electrons of the conductor
   - (c) none of the above
   - (d) protons present in metals

ANSWERS TO MCQ

1 (a) 2 (b) 3 (a) 4 (b) 5 (c) 6 (a) 7 (b)

Practice Problems

1. For a given metal, the relaxation time of electrons is 10^{-14}s at 300 K and electron density is 6×10^{28} m^{-3}. Calculate the electrical conductivity of the metal.
[1.7×10^7 Ω^{-1} m^{-1}]

2. Determine the electrical conductivity of copper if the relaxation time at 300 K is 2×10^{-14}s. Electron density in Cu is 8.5×10^{28} m^{-3} **[4.8×10^7 Ω^{-1}m^{-1}]**

3. The electrical resistivity of copper at 300 K is 1.72×10^{-8} Ωm. Find the thermal conductivity of copper. [Given: Lorentz no. 2.26×10^{-8} WΩK^2.
[394W/mK]

4. Find the free electron concentration in silver assuming one free electron per atom.
[Given: At.wt. of Ag: 108, $\delta_{Ag} = 105 \times 10^3$ kg/m^3]

Short-answer Questions

1. Why does wood have smaller heat conductivity than copper?

2. What are the implications for the semiconductor industry that silicon has a relatively good heat conductivity?

3. What are the limitations of the Drude–Lorentz free electron theory?

4. Define and explain giving their units: relaxation time, mean free path, mobility, drift velocity, thermal conductivity, electrical conductivity, current density.

5. What is Lorentz number? What is its unit? Give its significance.

6. State Widemann–Franz law.

7. Explain clearly conduction electrons, free electrons and bound electrons.

8. What are the successes of free electron theory?

9. List out the basic assumptions of classical free-electron theory.

10. Why does resistivity increase as temperature of the conductor increases?

11. Explain the statement 'The average thermal speed is very large but average velocity

is zero'.

12. Explain the drift motion of electrons.

Long-answer Questions

1. Obtain an expression for electrical conductivity based on free-electron theory and hence establish the relationship between thermal and electrical conductivities.

2. Explain free-electron theory of metals and derive an expression for (a) electrical conductivity (b) thermal conductivity

3. State and deduce Widemann–Franz law. Give postulates, success and shortcomings of free–electron theory .

4. State and explain Ohm's law hence derive its point form ($J = \sigma E$). Show that electrical conductivity σ is directly proportional to the relaxation time and also directly proportional to carrier concentration.

Questions From Previous Exams

(i) Discuss the important postulates of free electron theory of metals. (GTU- Jan' 2009)
(Refer Section: 7.3b)

(ii) Explain the classical free electron theory of metal. (GTU- Jan' 2010)
(Refer Section: 7.3)

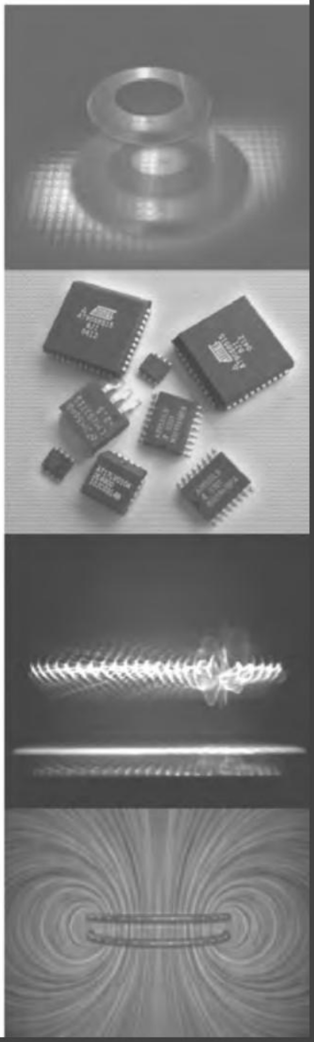
(iii) What are the success and drawback of classical free electron theory? (GTU- Jan' 2010)
(Refer Section: 7.8)

(iv) Derive an expression for electric conductivity. State and deduce Wiedeman–Franz law. (GTU- June' 2010)
(Refer sections: 7.5 and 7.7)

(v) Give the success and drawbacks of classical free electron theory. (GTU- June' 2010)
(Refer Section: 7.8)

Superconductivity and Applications

8



Imagine that you are not getting any electricity bill or the bill is suddenly and drastically reduced to an unbelievable value. Imagine trains are moving with supersonic speed without effectively consuming any electrical energy, or huge iron loads being lifted with minimum effort, a dream for steel industries. What is imagination today can be a reality tomorrow if superconductivity is possible at room temperature and above. To make such dreams possible we should understand this exciting property of materials, i.e., superconductivity.

INTRODUCTION

8.1

One of the most unusual and interesting properties of solids that certain metals and alloys exhibit is almost zero electrical resistivity, i.e., infinite electrical conductivity when they are cooled to sufficiently low temperatures. The phenomenon of reduction in the value of electrical resistivity to zero is known as *superconductivity*.

Superconductivity was first observed in mercury by Professor Heike Kamerlingh Onnes of the University of Leiden in 1911. He was probing the contemporary view that the electrical resistance of a metal arises due to two components. First, the resistance arises as a result of the scattering of the conduction electrons by the thermal vibrations of the metallic atoms and second, due to their scattering by the impurities in the metal. Naturally, the resistance should decrease with temperature as the thermal vibrations decrease with decrease in temperature. Therefore, the scattering of conduction electrons are reduced and hence the resistance of the material should decrease and should vanish at 0 K as thermal vibrations cease. The second component of resistance due to scattering of conduction electrons with impurity was regarded as essentially independent of temperature as there is no change in the impurity level with decrease in temperature.

Mercury was chosen to test the first hypothesis since a very high level of purity can be obtained by repeated distillation of it; and when it was found that the electrical resistivity of mercury below the boiling point of helium (4.2 K) was less than 10^{-11} times of its value above that temperature, it was thought that the hypothesis had been confirmed. Investigations showed that the resistivity, instead of decreasing continuously with decreasing temperature, disappeared sharply at 4.15 K (Fig. 8.1(a)) and that this behaviour was not affected by the presence of impurities. The existence of a new state of mercury characterised by the absence of electrical resistivity came to be recognised and to be known as the **superconducting state.** After two years, in 1913, Kamerlingh Onnes was awarded the Nobel Prize for this discovery.

The temperature at which the metal had passed into the superconducting state is known as the **superconducting transition temperature** (T_c). Onnes in 1914 performed a similar experiment on a coil of lead and established a 'persistent current' in it below 7.2 K. This discovery of superconductivity opened up tremendous curiosity about the scope and nature of the phenomenon.

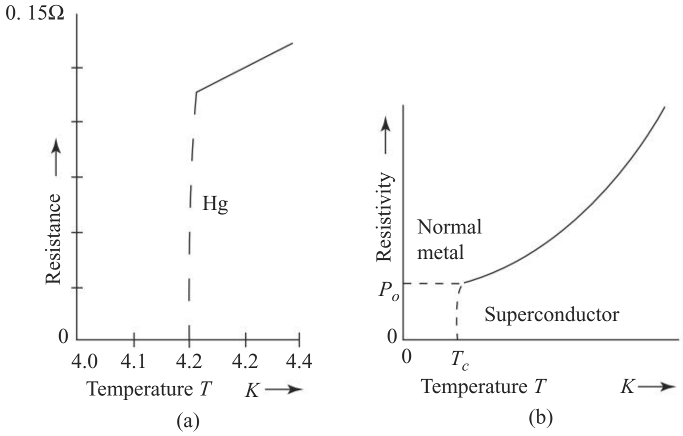


Fig. 8.1 *(a) Variation of resistance with temperature for mercury (b) Resistivity variation with temperature for normal metal and a superconductor*

TRANSITION TEMPERATURE AND A SURVEY OF SUPERCONDUCTIVITY

8.2

Superconductivity has now been observed in many metals, and several hundreds of alloys and compounds. The temperature at which the resistance disappears in zero magnetic field, is called the ***transition temperature*** T_c. The transition from the normal conducting state to the superconducting state can be exceedingly sharp in suitable circumstances (long cylindrical sample, pure metal, low measuring current) , the whole change taking place within a few ten-thousandths of a degree. The resistance becomes zero only when the current is steady or varies slowly. If the current fluctuates or alternates, small absorptions of energy occur which is roughly proportional to the rate of alternation. When the frequency alternation rises above 10 MHz, appreciable resistance arises, and at infrared frequencies (10^{13} Hz), the resistivity is the same in the normal and superconducting states, and is independent of temperature.

Kamerlingh Onnes observed that as the measuring current is increased, transition temperature (T_c) falls, and also that at any temperature below T_c, the superconductivity could be destroyed by the application of a magnetic field exceeding a critical value (H_c). These effects are related, because the lowering of temperature T_c by the current is due to the magnetic field produced by the current, and at any temperature below the transition temperature in a zero magnetic field, the superconductivity can be destroyed by raising the current above a critical value.

Table 8.1 *Superconducting elements*

	Element	T_c(K)	H_c(Am^{-1}) × 10^3	Type of superconductor
1	Al	1.196	7.9	I
2	Ga	1.09	4	I
3	Zn	0.79	3.8	I
4	Sn	3.72	24.5	I
5	Ta	4.48	67	II
6	Nb	9.25	156	II
7	Hg	4.12	33	I
8	V	5.1	105	I

Table 8.1 gives critical temperatures T_c(K) and the critical fields H_c(A/m) at 0 K of many pure metals. These elements show large resistivities in the normal state. It is observed that bismuth, antimony and tellurium become superconducting under high pressure.

A large number of superconducting, intermetallic compounds are known so far. The majority of them contain at least one superconducting component. Certain types of lattice, notably the sodium chloride structure, the nickel arsenide structure and the beta tungsten structure, are particularly common among superconducting compounds, which suggests that the properties of the lattice are significant, as well as those of the conduction electrons. This suggestion is supported by the behaviour of tin, which is superconducting in the a (base centre tetragonal) form, and not in the β (diamond-type) form. Some of the superconducting compounds have a high transition temperature of around 20 K.

Certain metals having low melting points and are mechanically soft and easily obtained in pure, strain-free condition, display similarities in their superconductive behaviour. They are distinguished as ‘ideal’, ‘Type I’ or ‘soft’ superconductors.

The behaviour of many alloys and some of the more refractory superconducting metals is more complex and individual, with respect to the way in which the superconductivity is affected by a magnetic field. These materials are termed as ‘Type II’ superconductors; or, if the superconductivity can be retained in high magnetic fields, as ‘hard’ or ‘high field’ superconductors. Superconducting materials may be either Type I or Type II.

MECHANISM OF SUPERCONDUCTIVITY: BARDEEN, COOPER AND SCHRIEFFER (BCS THEORY)

8.3

The successful theoretical explanation regarding the mechanism of superconductivity has been developed over many years with the first satisfactory explanation given by Bardeen, Cooper and

Schrieffer (BCS theory) in 1957. They explain superconductivity on the basis of formation of pairs of electrons, called **Cooper pairs**, which is a spatial state of affairs existing between the conduction electrons.

According to Coulomb force, two electrons in free space will mutually repel each other but in the solid state of matter, the force between the two electrons will be modified by the interaction of the electrons with the crystal lattice. In certain substances, the lattice interaction is so large that Coulomb repulsive force between electrons becomes modified into an attractive force binding certain electrons together into what are called Cooper pairs. For two electrons to become bound into a Cooper pair, they must be in the equilibrium, which means no current should be flowing between them and they should be having equal but opposite momentum.

On the basis of quantum mechanics in normal metals, plane electron waves, known as de Broglie waves, are assumed to propagate in the direction of motion of the electrons. The free electron gas in a metal can thus be considered as a superposition of several such de Broglie waves. Without any external electric field, these waves are reflected randomly by the surface of the crystal and there is no net electric current. If an external electric field is present, an additional wave component in the direction of the field arises and a finite current starts flowing. For an ideal crystal (no imperfections), there will be zero electrical resistivity (or infinite conductivity). But in real metals, phonons (lattice waves/vibrations) and defects such as impurities and vacancies act as scattering centres for the de Broglie waves, giving rise to a finite resistance. At high temperatures, phonon scattering plays a major role, while at low temperatures, scattering from static defects predominates.

The above fact can also be understood as follows. In an ordinary conductor, the behaviour of an electron is governed by two important principles. Only certain energies are permitted and no two electrons can have the same state. Electrons with the same energy will have opposite momenta. Let us consider a simplified one-dimensional solid. The two electrons with the same energy can have momenta of equal magnitude but of opposite sign, indicating that one electron will be directed to the right and the other to the left. Under normal equilibrium conditions when no current is flowing, the energy levels will be filled from the bottom upwards, with the pair of electrons of opposite momentum in each energy level. Therefore, equal numbers of electrons moving to the left and to the right produce zero net electron drift. Now if an electric field is applied to the solid, the motion of the electrons is modified. The electrons have a net energy gain from the electric field produced by a battery and there will now be more electrons with momentum directed to the right, say, than to the left, and a current is produced. Electrons having momentum to the right will occupy energy levels up to a greater height than electrons with opposite momentum to the left. Any of the electrons may collide with a phonon (a quantum of acoustic or vibrational energy associated with lattice vibrations), provided that the phonon energy and the conditions of the collision are such that the electron after collision still has permitted values of energy and momentum. The phonons in the material are moving at random, therefore, when a current is flowing, the effect of electron collisions with phonons is to destroy the net momentum of the electrons in the direction of the current flow. The effect of electron–phonon collisions is known as resistance to flow of current offered by a material and to remove electrons from high energy states with momentum directed to the right, to lower energy states with momentum to the left. The process continues till the original equilibrium is restored with equal number of electrons moving to the right and left, thereby no current flows.

Bardeen, Cooper and Schrieffer showed that the mechanism of superconductivity is entirely different from that of conductivity in normal, pure metals and is associated with the pairing of conduction electrons into what are known as Cooper pairs. An electron moving through a crystalline lattice draws ions towards it, a situation which attracts another electron. As far as the temperature is kept low enough, the energetically favourable electron pairing is not broken up, and the correlation of the conduction electrons gives rise to an almost infinite conductivity. The only requirement is the presence of a strong electron-lattice (i.e., electron–phonon) interactions.

This indicates that for superconducting materials, the interaction of electrons and phonons is modified by the fact that the electrons may no longer be independent, but may be bound together in Cooper pairs which can drift through the lattice when current flows. If either of the electrons in a Cooper pair is to be scattered by a collision with its momentum in the direction of current flow to be changed, the phonon must provide sufficient energy to break the bond holding the pair of electrons together in addition to the ordinary criteria pertaining to allowed energy and momentum before and after collision. At sufficiently low temperatures, the number of phonons in the material possessing sufficient energy to disrupt a Cooper pair becomes very small, and such pairs may pass through the lattice without scattering, i.e., these electrons experience zero resistance.

In other words, from the quantum mechanical calculation, the free energy is decreased, provided a Cooper pair involves electrons with opposite velocities (or momentum) and the lattice serves as the medium for exchange of momentum. Actually, electrons are being continuously exchanged between the Cooper pairs. Nevertheless, the transition to the superconducting state requires only that a sufficient number of electrons should condense into the paired state. At 0 K, all the conduction electrons are paired, but with increasing temperature, the number of pairs decreases. It was shown by Anderson that nonmagnetic impurities do not usually destroy superconductivity. Therefore, this phenomenon could be observed even for slightly impure samples. Still due to the presence of magnetic impurities, the superconducting transition temperature is lowered and/or it can suppress superconductivity completely.

A Cooper pair may be considered as a new particle having twice the mass and charge of an electron. For the superconducting state, the electrical current is due to the motion of several Cooper pairs in the same direction with the same velocity; only a single de Broglie wave can then be regarded as collectively representing these Cooper pairs. Because the momentum of the Cooper pairs is small, a given current corresponds to only a small velocity of the wave. Consequently, the wavelength of the de Broglie wave becomes very large, scattering from imperfections rarely occurs, and the conductivity tends to infinity:

Properties of superconductors

A few important properties of superconductors are stated below:

1. The current in the superconductors persists for a very long time. It can be demonstrated by laying a ring-type superconductor in a magnetic field and lowering its temperature below a critical temperature T_c and then removing the field. It is observed that the current which is set up persists over a period longer than two years without any attenuation.
2. The magnetic field does not penetrate into the body of the superconductor. This property is known as the **Meissner effect**. It provides the fundamental characterisation of superconductivity. The superconductor becomes a normal conductor when the magnetic field B is greater than a critical value $B_c(T)$. $B_c(T)$ is zero at $T = T_c$ and has the largest value B_0 at $T = 0$ K [Fig. 8.2(a)].
3. The specific heat of the material shows an abrupt change at $T = T_c$, jumping to a large value for $T < T_c$. [Fig. 8.2(b)].
4. If the current through the superconductor is increased beyond a critical value $I_c(T)$, the superconductor again becomes a normal conductor.
5. Certain materials are found to exhibit the superconductivity phenomenon on increasing the pressure over them, e.g., cesium is found to become a superconductor at $T_c = 1.5$ K if the pressure applied is 110 kbar. If stress increases in a superconductor, the critical temperature T_c also increases.
6. The T_c value is lowered if the impurities are added to superconducting elements.

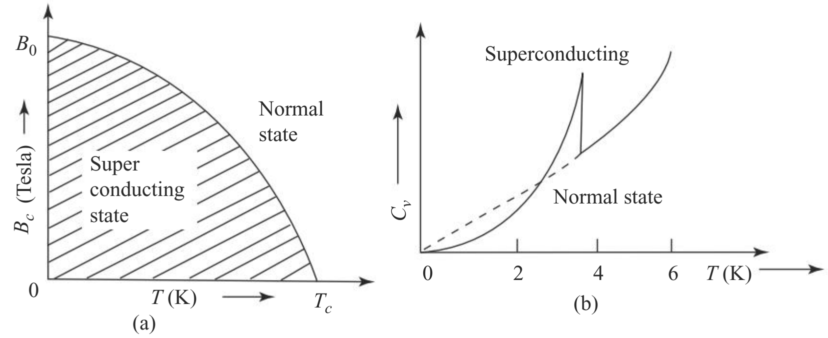


Fig. 8.2 *(a) The critical magnetic field at which superconductivity disappears (b) Specific heat in superconducting and normal state*

The general features of the superconducting materials are as stated below:

(a) Good conductors at room temperature are not superconductors, and superconducting metals are not good conductors at room temperature as the normal metals.

(b) Magnetic materials, like ferromagnetic and antiferromagnetic metals, do not show superconductivity.

(c) Monovalent metals do not exhibit superconductivity.

(d) Amorphous thin films of Be, Bi and Fe exhibit superconductivity below T_c.

(e) Bismuth, antimony and tellurium become superconducting under high pressure.

The superconducting properties of some selected compounds are shown in Table 8.2.

Table 8.2 *For superconducting compounds and alloys, their critical temperature T_c and critical magnetic field H_c (at 0 K)*

	Compounds and alloys	*T_c (K)*	*H_0 (Am^{-1}) × 10^6*	*Type of superconductor*
1	Nb–44%Ti	10.5	9.5	II
2	NbN	16.0	8	II
3	Nb–25%Zr	10.8	5.6	II
4	V_3Si	17.0	13	II
5	Nb_3Sn	18.5	16	II

THE EFFECT OF A MAGNETIC FIELD

8.4

It is observed experimentally that the superconducting state of a metal exists only in a particular range of temperature and magnetic field strength. The most essential condition for superconductivity in a metal is that a given combination of temperature and magnetic field strength should be less than their critical values T_c and H_c. This indicates that if the temperature of the specimen is raised above its T_c, superconductivity will disappear, and the same is true if a sufficiently strong magnetic field is employed above H_c. The critical value of the applied magnetic field necessary to destroy superconductivity and restore the normal resistivity is denoted by $H_c(T)$ and is a function of temperature. At the critical temperature T_c, the critical field $H_c(T)$ is zero.

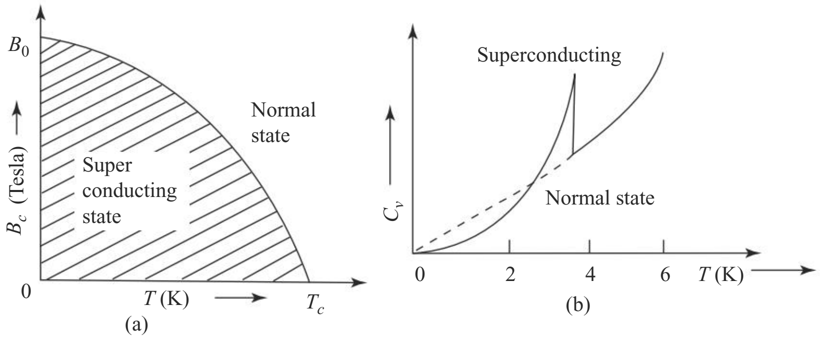


Fig. 8.3 *Variation of critical field H_C as a function of temperature T*

Figure 8.3 depicts the variation of critical field as a function of temperature for some elements. The curves represented in Fig. 8.3 are nearly parabolic and can reasonably be represented by the relation

$$H_c = H_0 \left[1 - \left(\frac{T}{T_c}\right)^2\right] \tag{8.1}$$

where H_c is the critical field strength at the temperature T, H_0 is the maximum critical field strength occurring at absolute zero, and T_c is the critical temperature or the highest temperature to exhibit superconductivity. Thus, the above equation fixes a curve which separates the normal region of the magnetic field– temperature diagram and the superconducting region of the metal.

The critical field curves for a number of pure metals are shown in Fig. 8.3. The surprising paradox of superconductivity arises from these curves. Consider lead at absolute zero temperature; the critical magnetic field of lead is 650 A cm^{-1}. This indicates that in the absence of lattice vibrations which has ceased at absolute zero, an increase in the energy content of the metal equivalent to a field of 650 A/cm quenches superconductivity. The energy density is 240 J m^{-3} or 10^{-7} eV/atom. Contrary to this, in zero magnetic-field strength, the critical temperature of lead is 7.175 K, which corresponds to a lattice energy of about 8×10^{-4} eV/atom. This indicates that the magnetic energy required to raise the electrons of lead from the superconducting state to the normal state is only one ten-thousandth of that required to normalise the superconducting state by thermal energy. This is the paradox that any successful theory of superconductivity must explain. We must first briefly review the quantum electronic model of a metal to understand the origin of this phenomenon.

The magnetic field which changes a superconductor to a normal conductor from a superconducting state is not needfully an external applied magnetic field. It may arise as a result of electric current flow in the conductor. For a long superconducting wire of radius r, the superconductivity may be destroyed when the current exceeds the critical value I_c which at the surface of the wire will produce a critical field H_c given by

$$I_c = 2\pi r Hc \tag{8.2}$$

This is called **Silsbee's rule**.

MEISSNER EFFECT

8.5 If we assume that zero resistance is the only basic property of a superconductor and other properties are derived from the perfect electrical conductivity, we would be wrong. This can be understood from the following hypothetical experiment.

Consider a perfect conductor (having zero resistance) kept in an external magnetic field. There will be no change in the distribution of the magnetic field, either inside or outside the material if a solid block of material is first placed in an external magnetic field and is then brought into the perfectly conducting state.

On removal of the external magnetic field, a surface current on the block will be induced to keep the interior magnetic field at its original value. But since the material is a perfect conductor, this surface current will never die out and the magnetic field remains trapped inside the perfect conductor.

The situation will be different if a material is a superconductor instead of a normal conductor. Let the experiment be repeated with a superconductor. A solid block of material in the normal state is kept in an external magnetic field. So the magnetic field will penetrate the block in the ordinary way depending on the permeability of the material. Now, on reducing the temperature below certain value so that the material is brought into the superconducting state, the internal magnetic field is at once expelled out. This phenomenon is known as the **Meissner effect** and is an essential property of the superconductor, but not for the perfect conductor. The superconductor is thus characterised by two essential independent properties (i) electrical resistivity of a superconductor is zero, and (ii) magnetic induction inside a superconductor is zero. This implies that inside the superconductor,

$$\vec{B} = \mu_0(\vec{H} + \vec{M}) = 0 \tag{8.3}$$

Considering the magnitude of the fields now,

$$\Rightarrow \qquad \mu_0 H\left(1 + \frac{M}{H}\right) = B = 0 \tag{8.4}$$

$$\Rightarrow \qquad \mu_0 H(1 + \chi) = B = 0$$

or

$$H + M = \frac{B}{\mu_0} = 0$$

i.e.,

$$M = -H$$

Therefore, a material exhibiting superconducting property is a perfect diamagnetic material whose magnetic susceptibility χ is,

$$\chi = \frac{M}{H} = -1$$

This important result cannot be inferred from the characterisation of a superconductor as a medium of zero resistivity.

Again, from the relation between current density and electric field, known as Ohm's law, we can write

$$\vec{E} = \vec{J}\rho \tag{8.5}$$

Therefore, zero resistivity imposes the value of $\rho = 0$, while current density $\vec{J}$ remains finite; hence the electric field vector $\vec{E}$ must be zero. Now Faraday's law of electromagnetic induction gives a direct relation between an electric field and the rate of change of a magnetic field of induction $\vec{B}$ producing it. This law is written in vector form as

$$\vec{\nabla} \times \vec{E} = \text{Curl } E = -\left[\frac{d\vec{B}}{dt}\right] \tag{8.6}$$

Equation (8.6) describes the relation between emf and rate of change of magnetic flux through an area, i.e.,

$$\varepsilon = \frac{d\phi_B}{dt}$$

For the zero resistance state, since $\vec{E} = 0$, we have

$$\frac{d\vec{B}}{dt} = 0; \quad \text{where } \vec{B} = \text{constant}$$

This implies that the magnetic field cannot change, hence the magnetic flux through the metal cannot change on cooling through the transition. This contradicts the Meissner effect and suggests that perfect diamagnetism and zero resistivity are two independent essential properties of the superconducting state.

Thus, the conditions defining the superconducting state are

$E = 0$ (from the absence of resistivity, i.e., $\rho = 0$), and

$B = 0$ (from the Meissner effect)

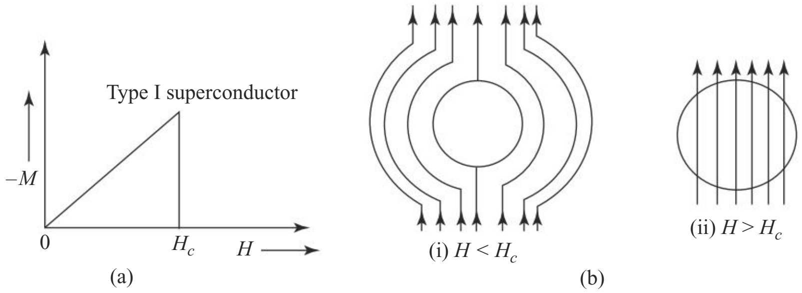


Fig. 8.4 *(a) Magnetism M vs. applied magnetic field H for a perfect superconductor (b) Magnetic field configuration for a spherical superconductor in an applied field with* $H < H_c$ *and* $H > H_c$

In several applications, Type II superconductors are widely used. In these type of superconducting materials, the magnetic flux does not penetrate (i.e., $B = 0$) up to a critical field H_{c1}, called the **lower critical magnetic field.** For fields greater than or equal to H_{c2}, the superconductivity is completely destroyed and the normal state is obtained, where H_{c2} is called the **upper critical magnetic field.** For magnetic fields between the range H_{c1} and H_{c2}, the magnetic flux partially penetrates the superconducting materials although it is still in the superconducting state. The state within the fields H_{c1} and H_{c2} is called the mixed state, i.e, both normal and superconducting states.

The value of critical field for Type II superconductors may be 100 times more or even higher than that for Type I superconducting materials. Ta, V and Nb are examples of Type II superconductors, and Al, Zn, Hg and Sn are some examples of Type I superconductors.

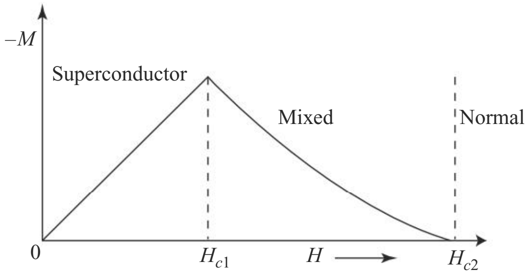


Fig. 8.5 *Magnetisation curve for Type II superconductor*

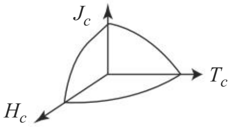


Fig. 8.6 *Critical surface phase diagram*

Three important factors define a superconducting state—critical temperature T_c, critical current density J_c, and critical magnetic field H_c. Each of them depends on the remaining two. To keep a material in the superconducting state, it is required to have all of them to remain below their critical values which depend on the material.

The highest values for H_c and J_c occur at 0 K, while the highest value for T_c occurs when H and J are zero. Thus, the plot of all these three parameters represents a critical surface. The relationship between H_c, J_c and T_c is shown in the phase diagram of Fig. 8.6. Within the surface, the material is superconducting and outside the surface the material is said to be in the normal state.

Comparison of type I and type II superaconductor:

Type I superconductor	*Type II superconductor*
1. They are known as soft superconductors.	1. They are known as hard superconductors.
2. One critical field H_c exists which is very low.	2. Two critical fields H_{c1} and H_{c2} exists which are very high.
3. Perfcet and complete Meissner effect is observed.	3. Meissner effect is not observed perfectly and completely.
4. Due to very low field values, only limited applications are possible.	4. Due to higher field values, large number of applications are possible.
5. Examples are Hg, Zn, Pb	5. Examples are Nb_3, Ge, Y_1 Ba_2 Cu_3O_7

THERMAL CONDUCTIVITY

8.6

The results of thermal conductivity in superconductors were discussed by Prof. Hulm and others. The observations in ideal superconductors show a remarkable decrease in the thermal conductivity which suggests that the electronic contribution to thermal conductivity drops. The implication is that possibly there is no role of electrons in the superconducting state for heat transfer. Results for a specimen of tin are given in Fig. 8.7.

It is observed experimentally that the thermal conductivity of superconductors undergoes a continuous change between the two phases and is usually lower in the superconducting phase. The thermal conductivity of tin is 34 W cm^{-1} K^{-1} at 2 K for the normal phase and 16 W cm^{-1} K^{-1} for the superconducting phase. Now at 4 K there is no superconducting phase for tin as the transition temperature T_c = 3.73 K and at 4 K the thermal conductivity is 55 W cm^{-1} K^{-1}.

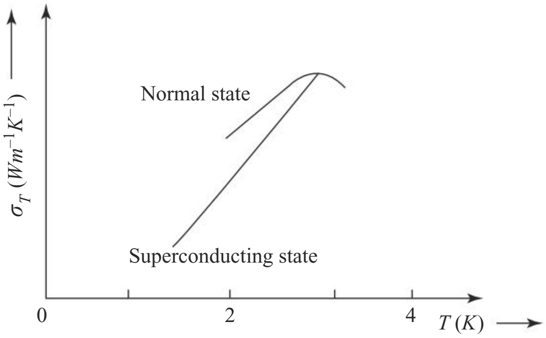


Fig. 8.7 *Thermal conductivity of a specimen of tin in the normal and superconducting states*

PENETRATION DEPTH OF A MAGNETIC FIELD IN A SUPERCONDUCTOR

8.7

Let us have a deeper insight into the surface of a superconductor on which such important phenomena are observed in Fig. 8.8. The figure depicts the boundary between a superconductor having area on the left and a magnetic field shown by the arrows on the right. Circles with points inside show the surface superconducting current flowing outwards out of the page. The field falls smoothly and exponentially with increasing depth from the surface, and the distance at which its strength falls by *e* times its original value is defined as its penetration depth λ_L.

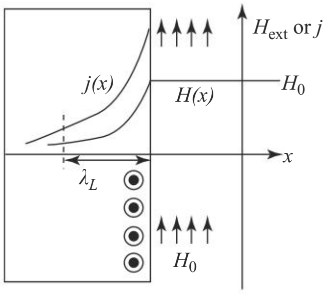


Fig. 8.8 *The penetration depth inside the superconducting body from the surface with the surface current is indicated by circles at the right bottom of the figure.*

Along the boundary, the superconducting current flows which screens the magnetic field and does not let it go inside the material (on the left side of the figure). This current flows in a near-surface layer. If this layer is narrowed, the current density increases, leading to the destruction of the material. But once the screening current is distributed over a certain thickness, the magnetic field penetrates the same distance inside the superconductor and gradually decreases. Figure 8.8 indicates that the behaviour of the external magnetic field strength and the current density depend on the distance inside the superconductor from the flat boundary of the superconductor. Both these quantities gradually decrease with depth into the material, the corresponding depth being customarily denoted by λ_L, called the ***London penetration depth*** (after the London brothers who introduced this quantity).

The penetration depth λ_L appears to be different for different superconductors. It depends on the properties of the material. Its values for several superconductors are listed in Table 8.3.

Table 8.3 *Penetration depth λ_L (Å) at 0 K*

Superconductor	λ_L (Å)
Tin	510
Aluminium	500
Lead	390
Mercury	380 – 450
Niobium	470
Thallium	920
Indium	640

The tabulated λ_L values refer to zero temperature *T*. They reflect the field penetration, as it was, for 'maximum' superconductivity when all the electrons become 'superconducting' and the 'normal' electron liquid disappears. As the temperature is raised from zero to a critical value, λ_L increases. The disappearance of superconductivity upon heating can be imagined as the increase of penetration of the magnetic field until it finally captures the whole of the sample at a critical temperature (see Fig. 8.9).

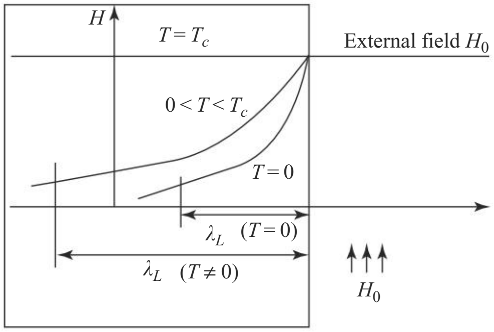


Fig. 8.9 *The penetration depth of a magnetic field into a superconductor as a fuction of temperature*

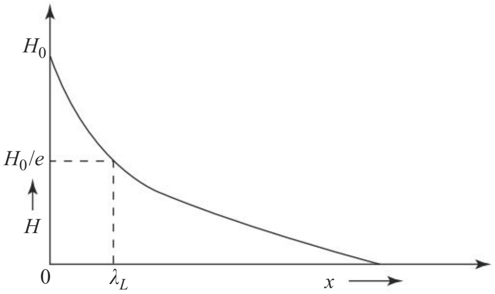


Fig. 8.10 *Decay of magnetic field*

The numerical values of the penetration depth are given in angstroms (Å). Typically, the distances between atoms in crystals make up several angstroms, and a change of these distances even by one hundredth of an angstrom may play an important role for the properties of the crystals. The penetration depth proves to be much larger than these inter-atomic distances. The domain of penetration of the magnetic field, and of the superconducting current, extends to hundreds and thousands of atomic layers. It cannot be 'too' thin, otherwise, the superconducting properties will 'not have enough time' to appear because superconductivity is a property of the whole system of atoms and electrons rather than the individual ones. However, from the point of view of ordinary dimensions, the penetration depth is sufficiently small: $\lambda_L \sim 10^{-6} - 10^{-5}$ cm. Values amounting to several millionths of a centimetre fully justify the use of the words 'forcing out to the surface'. We observe this in experiments with 'thick' samples.

F London and H London (1935) described the Meissner effect and zero resistivity by combining the two conditions (from the absence of resistivity) and (from the Meissner effect) to Maxwell's electromagnetic equations. They predicted that the applied field does not suddenly drop to zero at the surface of the superconductor, but decays exponentially according to the equation

$$H = H_0 e^{-\frac{x}{\lambda_l}} \tag{8.7}$$

where H_0 is the value of magnetic field at the surface and λ_L is the penetration depth, the distance for H to fall from H_0 to (H_0/e) as shown in Fig. 8.10.

Thus, the presence of surface currents and associated magnetic fields will obviously have a profound effect on the properties of thin film superconductors, or indeed of any superconductors whose dimensions are comparable to the penetration depth.

THE ENERGY GAP AND SPECIFIC HEAT

8.8

As shown in Fig. 8.11(a), the density of free electron states continuously changes with energy, and at temperature 0 K every state is occupied up to the Fermi level E_F. In case of superconductors, still those electrons which produce supercurrents are in even lower energy states. When the temperature is below the critical temperature T_c, an energy gap appears to exist in the distribution of density of states. As shown in Fig. 8.11(b), the steep rise in the density of states is a result of the appearance of the superconducting energy gap and has been confirmed experimentally.

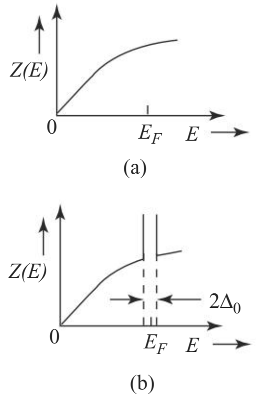
The gap width, $2\Delta_\circ$ is situated at the Fermi level E_F and the density of states is exceptionally large in the neighbourhood of E_F. The specific heat measurements and optical data, particularly infra-red absorption spectra, provide the experimental evidence for this energy gap.

At very low temperatures, the specific heat of a normal conductor, is observed to be of the form

$$C_v = [C_v]_{el} + [C_v]la = \gamma T + \beta T^3 \tag{8.8}$$

The first term of Eq. (8.8) refers to the electronic contribution to specific heat which is linearly proportional to T, while the second term is a lattice contribution which is proportional to T^3 and is explained by the Debye theory. Figure 8.12 shows the electronic contribution to the specific heat, $[C_v]_{el}$, in the superconducting state and it is compared with their contribution in the normal state. The two curves are obtained by measuring the specific heat C_v, with and without a magnetic field whose strength exceeds the critical field $H_c(T)$.

From this, the electronic contribution, $[C_v]_{el}$ is obtained by subtracting the lattice contribution $[C_v]_{la}$. In the presence of a magnetic field, the metal is in the normal state and in its absence it is in the superconducting state. The presence of the magnetic field in the normal state does not significantly modify the contribution to the specific heat by the electrons in the conduction band.



(a)

(b)

Fig. 8.11 *Density of energy states as a function of energy (a) normal metal, and (b) a superconductor*

As shown in Fig. 8.12, at very low temperatures, electronic contribution to specific heat $[C_v]_{el}$ is smaller in the superconducting state as compared to that in the normal state. As the temperature is increased gradually, $[C_v]_{el}$ increases slowly and on a further increase in temperature up to T_c, electronic contribution to specific heat increases rapidly and it has a value almost three times the value in the normal state. This behaviour of the superconducting state for the specific heat is a strong indication of the existence of an energy gap in the excitation spectrum of the electrons in the conduction band. At very low temperatures, very few electrons can get excited to the conduction band because of the energy gap $2\Delta_o$. The variation of the magnitude of the energy gap $\Delta(T)$ with temperature is shown in Fig. 8.13.

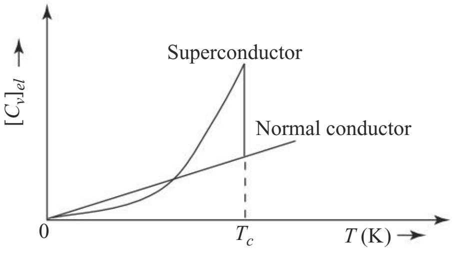


Fig. 8.12 *Temperature variation of the contribution of electrons to the spedific heat of a conductionr in the normal and superconducting states*

The BCS theory can very well explain the above-mentioned dependence of the energy gap on temperature. As the temperature is increased, more electrons get excited to the available states above the energy gap and electronic contribution to specific heat $[C_v]_{el}$ increases exponentially. The energy gap value at 0 K of some superconductors is given in Table 8.4.

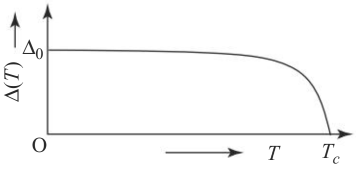


Fig. 8.13 *Energy gap as a fuction of temperture*

According to the BCS theory, the ratio of energy $\frac{2\Delta_o}{k_B T_c}$ gap to the thermal energy at the critical temperature is 3.53, which is same for all superconductors. This ratio is presented in the last column of Table 8.4, showing good agreement with the theory.

Table 9.4 *Energy gap at 0 K for some superconductors*

Element	*$2\Delta_o$ (meV)*	*T_c(K)*	*$2\Delta_o/k_BT_c$*
Niobium	3.05	9.50	3.8
Mercury	1.65	4.16	4.6
Tantalum	1.40	4.48	3.6
Lead	2.90	7.18	4.3
Tin	1.16	3.72	3.5
Aluminium	0.34	1.20	3.3

ISOTOPE EFFECT

8.9 From the observations of Maxwell and others using mercury isotopes, it is experimentally confirmed that the critical temperature T_c of superconductors changes with isotopic mass. For mercury, T_c varies from 4.185 K to 4.146 K as the isotopic mass M varies from 199.5 to 203.4. It was argued that isotopic mass can enter the process of formation of the superconducting phase of the electron states only through the electron–phonon interaction. In the early years of the development of the BCS theory, the simple law (8.9) was thought to be valid for most materials

$$T_c \propto M^{-\beta} \tag{8.9}$$

with $\beta = +0.5$. Thus

$$T_c\, M^{\beta} \tag{8.10}$$

The significant departures from the value of β have been observed afterwards.

HIGH T_C SUPERCONDUCTORS (HTS)

8.10 Extremely low critical temperatures of conventional superconductors (the low T_c type) put the most serious limitation on their use in technological applications. Working with devices that have to be cooled to temperatures in the range of liquid helium temperature (4.2 K) is obviously not viable. This has kept researchers from trying to discover superconductivity near room temperature. A decisive boost to this optimism came in 1986, when Bednorz and Muller synthesised metallic oxygen-deficient copper oxide compounds of an La–Ba(Sr)–Cu-O system with the transition temperature of about 30 K. The research for materials with higher critical temperatures ensued following this Nobel-Prize winning announcement. It has resulted in the development of a variety of materials with the highest critical temperature T_c in the vicinity of 135 K. The T_c values is so high compared to those of conventional superconductors that these materials are called high temperature superconductors or High T_c Superconductors (HTS).

The scope of this book does not permit us to do justice to the explosive development of HTS and their properties. Nevertheless, we give below in Table 8.5, an account of some materials that represent the main classes of HTS.

For the demonstration of superconductivity in suitable materials, they are to be cooled first below certain temperatures. Liquid helium temperature is 4.2 K and liquid nitrogen temperature is 77 K. Therefore, the superconductors that require liquid helium as a coolant are called **low-temperature superconductors** and if the coolant required is liquid nitrogen, the materials are known as **high T_c superconductors.**

A large number of possible practical applications of the phenomenon of superconductivity has not been achieved due to the limit of low values of T_c. For low T_c, the maintenance of superconductivity through refrigeration becomes a costly affair. In recent years, special efforts have been made to discover high T_c superconductors. Some of them are listed in Table 8.5.

Table 9.5 *High T_c Superconductors*

Material	*T_c in K*
Cu-O perovskite type	30 to 100
$La_{1.85}Sr_{0.15}CuO_4$	36
$La_{1.8}Sr_{0.2}CuO_4$	40
Y-Ba-Cu-O ($Y_1Ba_2Cu_3O_7$)	93–95
Hg-Ba-Ca-Cu-O	130–135
Granular Y-Ba-Cu-O	160
Lanthanum compounds	300–500

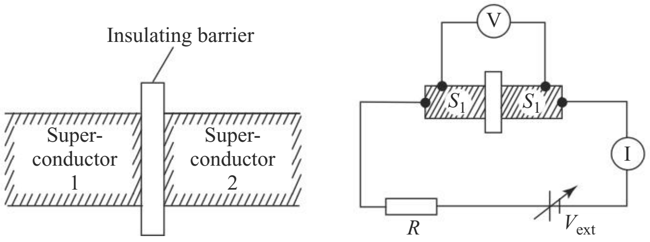
The detailed mechanism for high T_c super conductivity in certain materials is not known precisely. A number of arguments have been advanced to explain high T_c. For example, it has been realised that the role of oxygen is essential for high T_c oxide superconductors. The occurrence of a strong electronic band in the infrared spectrum and its correlation with superconductivity provides evidence that superconductivity of $YBa_2Cu_3O_{7-y}$ material is mediated by an excitonic mechanism.

The high T_c superconductors currently have only limited practical applications. The reason is that most of the high T_c superconductors are ceramics which are brittle. But, in thin-film technology, these materials can find applications in the electronic field such as in high-speed computers.

JOSEPHSON TUNNELLING

8.11

In 18.2, Josephson theoretically showed that an electron pair (the Cooper pair) can tunnel from one superconductor through a fine insulating layer into another superconductor as shown in Fig. 8.14 (a). This is known as ***Josephson tunnelling***. The coupling between the two superconductors provided by their insulating barrier must be very weak so that there is a very low probability of finding a Cooper pair in the insulating region. In practice, this is achieved by restricting the barrier thickness to about 10–20 Å. An arrangement such as that shown in Fig. 8.14(a) is referred to as a ***Josephson junction***. Two superconductors separated by a very thin strip of an insulator form a Josephson junction.



(a) (b)

Fig. 8.14 *(a) Combination of two superconductors so as to show josephson tunnelling (b) Circuit for studying losephson effect*

The significant phenomena associated with Cooper pair tunnelling in a Josephson junction are identified as (i) Josephson effect, and (ii) supercurrent quantum interference.

Josephson Effect We present a basic theoretical description of the Josephson effect for an experimental geometry given by Fig. 8.14 (b). The interpretation of general expressions reveals two types of effects—the dc Josephson effect and the ac Josephson effect.

For simplicity, we assume that the superconductors S_1 and S_2 in our model are of the same material. The temperature being so low, we need to consider the electrons only in their superconducting ground state. The magnetic field near the junction must be screened completely because the highly coherent Cooper pair states are strongly influenced by magnetic fields. Even the magnetic field due to earth may be sufficient to completely obscure the Josephson effect.

The wave nature of moving particles makes the electrons tunnel through the barrier (insulator), i.e., the electrons can tunnel from one superconductor to the other. As a result of the tunnelling of electrons (Cooper pairs) across the insulator, there is a net current across the junction. This phenomenon is called the **dc Josephson effect**. Here, the important observation is that the current flows even in the absence of a potential difference.

The thickness of the insulator, the material nature and the temperature decides the magnitude of the current.

If a potential difference V is applied between the two sides of the junction as in Fig. 8.14 (b), there will be an oscillation of the tunnelling current with angular frequency $\omega = 2e\,V/h$. This is called the **ac Josephson effect**.

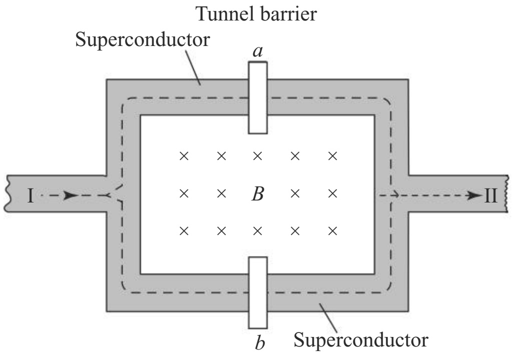


Fig. 8.15 *SQUID—Superconducting Quantum Interference Device*

SQUID Josephson junctions are used in sensitive magnetometers called SQUID—Superconducting Quantum Interference Device.

It is formed by connecting two Josephson junctions in parallel, as shown in Fig. 8.15.

When current is passed from one side as shown, due to two junctions in parallel, the current splits and flows across the two branches. The current through the circuit will have a periodicity which is very sensitive to the magnetic flux passing normally through the closed circuit. Therefore, an extremely small magnetic flux can be detected with this device.

The small voltages of the order of 10^{-15} V can be detected by SQUID. A small change in Magnetic field of the order of 10^{-21} T can be detected. Using SQUID, weak magnetic fields produced by biological currents like that in the brain can also be detected. SQUID detectors are used to measure levels of iron in the liver, so that iron built up can be treated before much harm occurs to the body.

APPLICATIONS OF SUPERCONDUCTORS

8.12 Soon after Kamerlingh Onnes discovered superconductivity, scientists began dreaming up practical applications for this strange new phenomenon. Powerful new superconducting magnets could be made much smaller than a resistive magnet, because the windings could carry large currents with no energy loss. Generators wound with superconductors could generate the same amount of electricity with smaller equipment and less energy. Once the electricity was generated, it could be distributed through superconducting wires. Energy could be stored in superconducting coils for long periods of time without significant loss.

The recent discovery of high-temperature superconductors brings us a giant step closer to the dream of the early scientists. Applications currently being explored are mostly extensions of current technology used with the low-temperature superconductors. Recent applications of high-temperature superconductors include magnetic shielding devices, medical imaging systems, Superconducting Quantum Interference Devices (SQUIDS), infrared sensors, analog signal-processing devices, and microwave devices. As our understanding of the properties of superconducting materials increase, applications such as power transmission, superconducting magnets in generators, energy-storage devices, particle accelerators, levitated vehicle transportation, rotating machinery, and magnetic separators will become more possible.

The ability of superconductors to conduct electricity with zero resistance can be exploited in the use of electrical transmission lines. Currently, a substantial fraction of electricity is lost as heat through resistance associated with traditional conductors such as copper or aluminium. A large-scale shift to superconductivity technology depends on whether wires can be prepared from the brittle ceramics that retain their superconductivity at 77 K while supporting large current densities.

The field of electronics holds great promise for practical applications of superconductors. The miniaturisation and increased speed of computer chips are limited by the generation of heat and the charging time of capacitors due to the resistance of the interconnecting metal films. The use of new superconductive films may result in more densely packed chips which could transmit information more rapidly by several orders of magnitude. Superconducting electronics have achieved impressive accomplishments in the field of digital electronics. Logic delays of 13 picoseconds and switching times of 9 picoseconds have been experimentally demonstrated. Through the use of basic Josephson junctions, scientists have been able to make very sensitive microwave detectors, magnetometers, SQUIDs and very stable voltage sources.

The use of superconductors for transportation has already been established using liquid He as a refrigerant. Prototype levitated trains have been constructed in Japan by using superconducting magnets.

Superconducting magnets are already crucial components of several technologies. Magnetic Resonance Imaging (MRI) is playing an ever-increasing role in diagnostic medicine. The intense magnetic fields that are needed for these instruments are a perfect application of superconductors. Similarly, particle accelerators used in high-energy physics studies are very dependant on high-field superconducting magnets. The recent controversy surrounding the continued funding for the Superconducting Super Collider (SSC) illustrates the political ramifications of the applications of new technologies.

New applications of superconductors will increase with critical temperature. Liquid-nitrogen based superconductors have provided the industry more flexibility to utilise superconductivity as compared to liquid helium superconductors. The possible discovery of room-temperature superconductors has the potential to bring superconducting devices into our everyday lives.

High-temperature superconductors are recent innovations from scientific research laboratories. New commercial innovations begin with the existing technological knowledge generated by research scientists. The work of commercialisation centres on the development of new products and the engineering needed to implement the new technology. Superconductivity has had a long history as a specialised field of physics. Through the collaborative efforts of government funded research, independent research groups and commercial industries, applications of new high-temperature superconductors will become possible in the not-so-distant future.

Electrical Switching Element Due to the infinite ratio between the resistance of a metal in the normal and superconducting state, a superconductor can be used to switch electrical current between different possible paths. Either a magnetic field or temperature can be used to control the change of state of such a switch. A film of lead is used as the switching element for a temperature-controlled switch. The current-carrying film is operated below the transition temperature to make it superconducting. The current can be interrupted by means of a heating coil made of nonsuperconducting metal. As the lead film becomes resistive, the current flowing through it is diverted through a parallel path of low resistance.

A device known as **cryotron,** a magnetically operated current switch is shown in Fig. 8.16. The fact that superconductivity disappears for a magnetic field exceeding the critical value, has been utilised in a cryotron. PQ is a straight wire of superconducting material S_1 on which is wound a coil of superconducting material S_2. The temperature of the entire system is kept below the transition temperatures of both the superconducting materials S_1 and S_2. Now if the magnetic field produced by the coil B exceeds the critical field H_c of the material at the operating temperature then the material S_1 of the core A operates in the normal state. The core A has nonzero resistance and hence the current in the core reduces. Thus, the current in the core A may be controlled by the current in the coil B. This control current I_2 in coil required to make the core A normal depends on the dc current I_1 flowing through the core because I_1 also produces its own magnetic field.

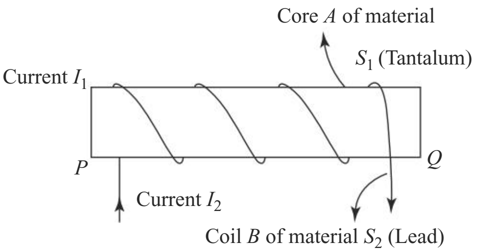


Fig. 8.16 *The wire wound cyclotron*

If the operating temperature is 4.2 K, that of a liquid helium bath, a metal such as tentalum having T_c (= 4.38 K) only slightly greater than 4.2 K may be used for the core A so that the critical magnetic field required to cause the normal state is relatively small. The material chosen for a coil must remain superconducting even with the control current flowing in the coil. Niobium or lead may be used as coil materials. A flip-flop circuit uses a single cryotron shown in Fig. 8.16. Elements used being superconductors, the power consumption of the device is extremely small. A large digital computer using cryotrons may require extremely small power of the order of only half a watt—excluding, of course, the terminal equipment and the helium cryostat. Further, the size of such a computer will be extremely small, occupying a very small space.

Superconductor Transformers and Transmission Lines If superconductors are used for winding of a transformer, the power losses will be very small. By using superconductors, transformers of 2000 to 3000 mW can be manufactured and they will be of the portable type.

Electrical transmission lines with the use of cables with superconductors can be designed. Theoretically, such a line, if necessary cooling is provided, can transmit high power to any desirable distance without losses.

Electric Motors and Generators The manufacture of electrical machines, generators and motors with superconductor windings is one more field of application of superconductors which is very prospective. It is possible to obtain high efficiency and small weight by using superconductors. The possibility of achieving extremely strong magnetic fields and using lossless coils to cut those fields is of great interest to the power engineer. We may note that even in superconductors, there are losses with low-frequency alternating current which may be overcome by using new materials. The most encouraging work on the industrial side is concerned with dc machines. In particular, a superconductor homopolar motor is being widely used for plant operation.

In this simple form, the homopolar machine uses, or generates, electrical power in the form of a very high current at very low voltage, and this is rarely convenient. This difficulty has been overcome by using a segmented structure for the rotating conductor, with a special brush arrangement which allows the 2500 kW motor to accept its input power at the much higher potential difference of about 450 V.

The superconducting homopolar dc generator has, as yet, not been developed very far, but new design studies are proceeding in parallel with the more immediately rewarding work on motors. Applications of particular superconductors for dc superconducting motors appear to lie in fields like ship propulsion and large mills, where very high torques are required at comparatively low speeds, which can be controlled easily. In terms of capital and running costs, they are likely to be competitive with conventional machines for large output powers, and will probably also be smaller, lighter and more efficient. Similar savings in weight, cost, etc., are predicted for alternators.

Superconductor Fuse and Breaker Fuses and circuit breakers can be prepared by using thin- film superconductors. A thin, insulated superconductor film get back to its normal state after the current of more than critical density has passed through it. Such films can be used instead of a fuse. By selecting the necessary appropriate material for a film, it is possible to regulate very precisely the admissible intensity of current.

If a long film is taken and at the moment of its transition into normal state, its resistance is rather high then it can be used as a breaker. Lead is used here.

Magnetic Mirror and Superconducting Bearing Zero magnetic induction, which is responsible for the ‘levitation’ effects, are applied in frictionless bearings. The most famous levitation experiment was that in which a horizontal bar magnet was suspended from a flexible chain, and then lowered over a sheet of lead which had been cooled to the superconducting state. As the magnet approached the superconducting sheet, the supporting chain became limp, and eventually drooped down in a loop below the magnet, which remained floating horizontally above the lead sheet. The magnetic field of the approaching magnet induces a current in the surface of the superconductor. Due to zero resistance, the current persisted, and the field due to this current repelled that due to the bar magnet. The superconducting surface can be regarded as a mirror below which there exists an image of the magnet. Using a different experimental arrangement, the superconducting sheet might have been supported by the magnetic field from the permanent magnet. Instead of using a permanent magnet, the supporting field might have been provided by a superconducting solenoid carrying a current.

The repulsion between the magnetic field of a superconducting solenoid and that due to the current induced in the surface of a nearly superconductor is the basis of most designs for superconducting mechanical support. Unlike simple magnetic repulsion systems, the superconducting support is usually unconditionally stable, and thus needs no associated servo system.

Various arrangements have been proposed to give the two-dimensional support needed for a shaft bearing, and the three-dimensional support needed for a floating sphere. This last arrangement has been applied in a superconducting gyroscope with a spherical rotor spinning for months under nearly frictionless conditions.

Concluding Remark Time lags between new discoveries and practical applications are often great. The discovery of the laser in the early 60’s has only recently been appreciated today through applications such as laser surgery, laser optical communication, and compact disc players. The rapid progress in the field of superconductivity leads one to believe that applications of superconductors are limited only by one’s imagination and time.

Solved Problems

1 *Calculate the critical current which can flow through a long thin superconducting wire of aluminium of 10^{-3} m diameter. The critical magnetic field for aluminium is 7.9×10^3 A/m.*

Solution Givn, $H_c = 7.9 \times 10^3$ A/m, $r = 0.5 \times 10^{-3}$ m

The critical current $I_c = 2\pi r H_c$

$$I_c = 2 \times 3.14 \times 0.5 \times 10^{-3} \times 7.9 \times 10^3$$

$$I_c = 24.806 \text{ A}$$

2 *The critical temperature, T_c, for mercury with an isotopic mass of 199.5 u is 4.185 K. Calculate its T_c when its isotopic mass changes to 203.4 u.*

Solution Given, $M_1 = 199.5$ u, $T_{c1} = 4.185$ K, $M_2 = 203.4$ u, $T_{c2} = ?$

$$T_c M^{+\beta} = \text{const}$$

$$\Rightarrow \quad Tc1\ M_1^{+\beta} = T_{c2} M_2^{+\beta}$$

$$T_{c2} = \frac{M_1^{+\beta}}{M_2^{+\beta}} T_{c1} = \frac{199.5}{203.4} \times 4.185 = 4.104 \text{ K}$$

3 *A metallic, long, thin superconducting wire produces a magnetic field of 1.05×10^5 A/m on its surface due to the current passing through the wire at a temperature T. Calculate the value of T if the critical magnetic field of the metal at 0 K is 1.5×10^5 A/m and its critical temperature is 9.20 K.*

Solution Given, $H_0 = 1.05 \times 10^5$ A/m, $H_c = 1.5 \times 10^5$ A/m, $T_c = 9.20$ K

$$H_c = H_0\left[1 - \left(\frac{T}{T_c}\right)^2\right]$$

$$T = T_c\sqrt{1 - \frac{H_c}{H_0}} = 9.2 \times \sqrt{1 - \frac{1.05 \times 10^5}{1.5 \times 10^5}} = 5.04 \text{ K}$$

4 *A critical magnetic field of lead wire is 6.5×10^3 A/m at 0 K. At what temperature would the critical magnetic field of lead drop to 4.5×10^3 A/m if the critical temperature of lead is 7.18 K? What is the critical current density at that temperature if the diameter of the wire is 2×10^{-3} m?*

Solution $H_0 = 6.5 \times 10^3$ A/m, $H_c = 4.5 \times 10^3$ A/m, $T_c = 7.18$ K, $r = 1 \times 10^{-3}$ m

$$H_c = H_0\left[1 - \left(\frac{T}{T_c}\right)^2\right]$$

$$T = T_c\sqrt{1-\frac{H_c}{H_0}} = 7.18 \times \sqrt{1-\frac{4.5\times10^3}{6.5\times10^3}} = 3.98 \text{ K}$$

Critical current density J_c

$$J_c = \frac{I_c}{A} = \frac{2\pi r H_c}{\pi r^2} = \frac{2H_c}{r} = \frac{2\times4.5\times10^3}{1\times10^{-3}} = 9.0\times10^6 \text{ A/m}^2$$

5 *If the voltage applied across a Josephson junction is 8 mV, what is the frequency of the radiation emitted by the junction?*

Solution Given, $V = 8\ \mu\text{V}$

$$v = \frac{2eV}{h} = \frac{2\times1.6\ 10^{-19}\times8\times10^{-6}}{6.62\times10^{-34}} = 3.86\times10^9 \text{ Hz}$$

6 *If the critical current passing through a long thin superconducting wire of 0.5 mm radius is 22.608 A, calculate the critical magnetic field.*

Solution $I_c = 22.608$ A, $r = 0.5$ mm,

$$I_c = 2\pi r H_c \Rightarrow H_c = \frac{I_c}{2\pi r} = \frac{22.608}{2\times3.14\times0.5\times10^{-3}} = 7.2\times10^3 \text{ A/m}$$

7 *At zero magnetic field a superconducting tin has a critical temperature of 3.7 K. At 0 K, the critical magnetic field is 0.306 T. Calculate the critical magnetic field at 2.0 K.*

Solution Given, $T_c = 3.7$ K, $H_c = 0.306$ T, $T = 2.0$ K, $H_c = ?$

$$H_c = H_0\left[1-\left(\frac{T}{T_c}\right)^2\right]$$

$$H_c = 0.0306\left[1-\left(\frac{2.0}{3.7}\right)^2\right] = 0.022 \text{ T}$$

Multiple-Choice Questions

1. The condition required for superconductivity phenomenon in a material is
   (a) $H = 0, R = 0$ below T_c
   (b) $H = 0$, below T_c
   (c) $R = 0$, below T_c
   (d) $H = 0$, $R = 0$, above T_c

2. Superconductors are __________ materials.
   (a) perfectly ferromagnetic
   (b) perfectly diamagnetic
   (c) perfectly paramagnetic
   (d) dielectric

3. Cooper pairs in the superconductors are nothing but
   (a) pairs of electrons in the conduction band
   (b) pairs of holes in the conduction band
   (c) pairs of electrons in the valence band
   (d) pairs of electrons and holes in the conduction band

4. The magnetic susceptibility of a superconducting material is
- (a) –1
- (b) positive and very large
- (c) negative and very large
- (d) +1

5. High T_c superconductors are the materials having temperature
- (a) very high above the room temperature
- (b) very high above the boiling point of water
- (c) very high above liquid helium but below liquid nitrogen temperature
- (d) above liquid nitrogen temperature

6. For a Type I superconductor,
- (a) two critical magnetic fields exist
- (b) two critical electric fields exist
- (c) only one critical magnetic field exists
- (d) only one critical electric field exists

7. For a Type II superconductor,
- (a) two critical magnetic fields exist
- (b) two critical electric fields exist
- (c) only one critical magnetic field exists
- (d) only one critical electric field exists

8. Good electric conductors at room temperature
- (a) are also good superconductors below T_c
- (b) are bad conductors of electricity below T_c
- (c) are good superconductors above T_c
- (d) may behave as superconductors below T_c

9. High T_c superconductors below T_c are
- (a) normally bad conductors of electricity at room temperature
- (b) behave normally as magnetic materials at room temperature
- (c) metals
- (d) metal alloys

10. The complete expulsion of all the magnetic fields by superconducting materials is called
- (a) Josephson effect (b) Maglev effect
- (c) SQUID (d) Meissner effect

11. The wave nature of a moving particle makes the electrons to tunnel through the insulating barrier from one superconductor to other. This is known as
- (a) Josephson effect (b) Maglev effect
- (c) SQUID (d) Meissner Effect

12. A relay or switch made from superconductors whose size is small is known as
- (a) SQUID (b) cyclotron
- (c) cryotron (d) betatron

13. The effect of impurity on superconductor is that
- (a) T_c value increases
- (b) T_c value decreases
- (c) electrical conductivity increases
- (d) does not change the value of T_c

14. Increase in stress results in... ... . of the T_c value.
- (a) decrease
- (b) increase
- (c) no change
- (d) both increase or decrease

15. Persistent current is the characteristic property of
- (a) superconductivity (b) Meissner Effect
- (c) Josephson effect (d) Hall effect

ANSWERS TO MCQS

1. (a) 2. (b) 3. (a) 4. (a) 5. (d) 6. (c) 7. (a) 8. (b)
9. (a) 10. (d) 11. (a) 12. (b) 13. (b) 14. (b) 15. (a)

Practice Problems

1. A superconducting lead has a critical temperature of 7.26 K at zero magnetic field and a critical field of 8×10^5 A/m at 0 K. Find the critical field at 5 K. **[4.2×10^5 A/m]**

2. The penetration depths for lead are 396 Å and 1730 Å at 3 K and 7.1 K, respectively. Calculate the critical temperature for lead. **[7.193 K]**

3. A voltage of 5.9 μV is applied across a Josephson junction. What is the frequency of the radiation emitted by the junction? **[2851 MHz]**

4. Superconducting Nb has a critical temperature of 9.15 K at zero magnetic field and a critical field of 0.1960 T at 0 K. Find the critical field at 5 K. **[0.137 T]**

Short-answer Questions

1. Define superconductivity.

2. What are superconductors? Give examples.

3. Define critical magnetic field H_c. Give its relation with the critical temperature T_c.

4. Define the critical temperature or transition temperature T_c.

5. What is the effect of impurity and stress on a superconductor?

6. What is an isotopic effect on superconductivity? Give the appropriate relation showing the isotope effect.

7. Show graphically how H_c varies with respect to temperature.

8. Define critical current density J_c.

9. Explain persistent current.

10. What is the Meissner effect?

11. What are the three important factors required to define a superconducting state?

12. State two independent basic conditions defining superconductivity.

13. Show the diagram that depicts Meissner effect.

14. What are the types of superconductors based on their behaviour in an applied magnetic field?

15. What are Type I and Type II superconductors? Give examples.

16. Compare Type I and Type II superconductors.

17. Draw the magnetisation curves for Type I and Type II superconductors.

18. Categorise the superconductors based on coolants. Give the names of coolants used.

19. What are low-temperature and high-temperature superconductors?

20. What are high T_c superconductors? Give examples.

21. What is a Josephson junction?

22. Explain Josephson effects.

23. What is a SQUID?

Long-answer Questions

1. What are superconductors and normal conductors? Explain a few important characteristic properties of superconducting materials. What are Type I and Type II superconductors?

2. Write a note on the terms used in the superconductivity study like critical magnetic field, critical temperature, Meissner effect, isotope effect.

3. Explain the mechanism of superconductivity or BCS theory in detail.

4. Explain the Josephson tunnelling and Josephson effect in detail and discuss the application.

5. Give your comparison between two types of available superconductors.

6. Write a note on high T_c superconductors.

7. Give a brief account of few important applications of superconductors.

8. Discuss the Meissner effect in detail and prove that superconductors are perfectly diamagnetic in nature.

9. Discuss the thermal conductivity of superconducting materials.

10. What do you mean by penetration depth? Explain it giving a suitable diagram. Explain the effect of temperature on penetration depth of a magnetic field into a superconductor.

11. Discuss the energy gap and specific heat for superconductor. Discuss the density of energy states as a function of energy for normal metals and superconductors.

12. What is meant by superconductivity? Describe the effect of

13. (a) magnetic field, (b) frequency, and (c) isotopes on superconductors.

14. Discuss the phenomenon of superconductivity. Explain some of its applications.

15. Write an essay on superconductivity.

16. Explain the chief characteristics of superconductors. Discuss the important applications of superconductors.

17. What is superconductivity? Give an account of the occurrence, properties and uses of superconductors.

18. What are superconductors? Mention the important property changes that occur in materials when they change from normal to superconducting state. Give some examples of practical uses that exploit the above property changes.

Questions From Previous Exams

(i) Define the transition temperature of superconductivity. (GTU-Jan'2009)

The temperature at which a material loses its electrical resistivity to absolute zero is called transition temperature.

(ii) Compare type-I and type-II superconductors. (GTU-Jan' 2009) (GTU-Jan' 2010)

Refer Section: 8.5

(iii) Give the differences between type-I and type-II superconductors. (GTU-June' 2010)'

Refer Sections 8.5 Added file

(iv) Discuss the properties of superconductors. (GTU-Jan' 2009)

Refer Section: 8.3

(v) What is a superconducting material? List the properties of superconducting material and explain in detail. (GTU-Jan' 2010)

Superconductivity is a phenomenon in which certain metals, alloys and ceramics conduct electricity without resistance when they are cooled below a certain temperature.

Properties are electrical resistance, effect of impurities, effect of pressure and stress, isotope effect, magnetic field effect, Meissner effect, persistence current etc.

(vi) Give the application of superconductor. (GTU-Jan' 2010)

Refer Section: 8.12

(vii) $H_c = H_0 (1 - T^2/Tc^2)$ is the formula which gives the relation of critical magnetic field and temperature. (GTU-June' 2010)

(viii) Explain Josephson effect and its applications. (GTU-June' 2010)

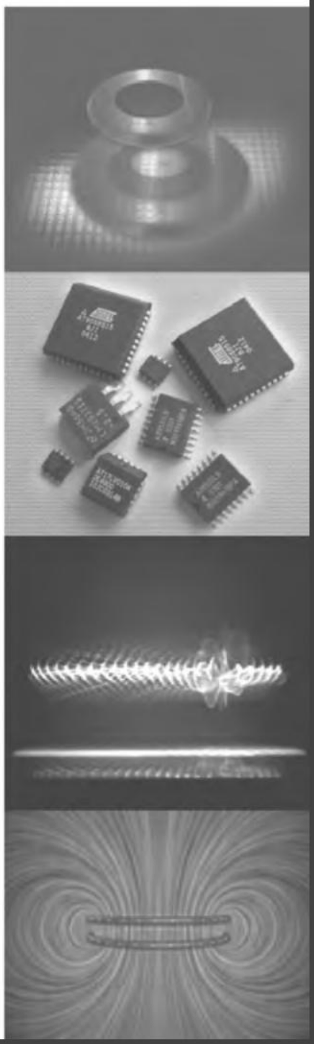
Refer Section: 8.11

The McGraw·Hill Companies

New Engineering Materials

9

Can we obtain glasses (amorphous) with the properties of that of metals (crystals)? Can we obtain materials with high resistivity, high ductility, and higher corrosion resistance and yet Rave them behave as metals? Can we have materials which are smart? Can we have materials which can replace, augment or perform natural functions? Such materials can be used to develop amazing applications. Since they have the properties that scientists can design and manipulate. This chapter describes all of these in a very lucid manner.



INTRODUCTION

9.1 The progress in areas of miniaturization, automation, sophistication, economization, etc., is due to the developments in technology and engineering. These developments have taken place not only due to the new designs and concepts but also due to new engineering materials. There is a continuous search and effort for developing new materials with improved properties such as good mechanical strength, high stability, large electrical conductivity, resistance to wear and tear, better corrosion resistance, and so on. The newly prepared engineering materials find major applications in industries and research.

To develop a new engineering material with the desired properties for a specific application with lower cost and enough safety factor consideration, a thorough knowledge of the properties, composition, structure and preparation is essential. Such materials would have properties that scientists can manipulate as per their need. There are many areas in which such new materials are available, but the following sections would discuss the main areas such as metallic glasses, nanotechnology and materials, Shape Memory Alloys(SMA) and biomaterials. The discussions include types, preparation, properties and applications of these materials.

METALLIC GLASSES

9.2

Metals and alloys are generally orderly materials or crystalline materials but glasses are amorphous materials. Metals are malleable, ductile, ferromagnetic, electrical, thermal conductors and opaque, whereas glasses are brittle, transparent and nonmagnetic. Metallic glasses are new engineering materials which share the properties of both metals and glasses. They are strong, stable, ductile, malleable and opaque like a metal but metallic glasses have amorphous structure, brittleness and high corrosion resistance like glasses. In 1965, it was found that some of the metals could be frozen in the amorphous state if they were rapidly cooled from their molten state to the solid state. Thus, amorphous metallic solids could be produced on cooling molten (liquid) phase to the solid phase at high cooling rates of the order of 106 °C/s. The amorphous metallic solids thus formed were known as **metallic glasses** or **met glasses** or **amorphous metals**. Thus, *metallic glasses are amorphous metals that are produced by solidification of liquid alloys.*

9.2.1 Types of Metallic Glasses

Metallic glasses are mainly of two types:

1. Metal–Metal Glasses: They can be prepared by combining metals of different proportions, e.g., Mg–Zn ,Cu–Zr ,Ni–Nb, etc.
2. Metal–Metalloid Glasses: They can be prepared by combining transition metals such as Fe, Co, Ni and metalloids such as B, Si, P with different proportions, e.g., Fe_{67} Co_{18} B_{14} Si, Fe_{40} Ni_{38} Mo_4 B_{18}

Commercial metallic glasses are generally of this type.

9.2.2 Preparation of Metallic Glasses

Metallic glasses are prepared by cooling molten metal to the glassy state at a cooling rate, high enough to bypass crystallization. The cooling rate depends on the composition and their phase diagram. The temperature at which the transition from liquid to solid takes place is known as glass transition temperature. The glass transition temperature for metallic alloys is in the range of 20°C to 300°C. Various rapid cooling techniques such as spraying, spinning, laser deposition, etc., can be employed to prepare metallic glasses. Melt spinning is one of the most common methods used to prepare metallic glasses at a large-scale level. Therefore, it is discussed in detail.

Melt Spinning Method

It is illustrated in Fig. 9.1.

In order to prepare a metallic glass of a particular type, an appropriate combination of either metal–metal or metal–metalloid alloy in their stoichiometric proportion are taken in a refractory tube, which has a fine nozzle at its bottom.

- The nozzle side of the tube is adjusted exactly over the spinning disc as shown in Fig. 9.1.
- An induction heater attached to the refractory tube melts under the inert gas atmosphere. This melt is kept above its melting point till it gets transformed into a homogeneous mixture.

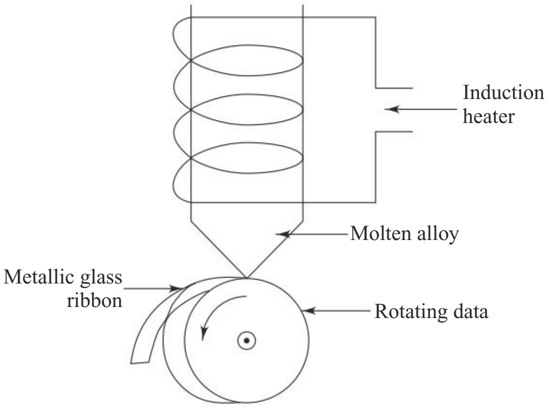


Fig. 9.1 *Melt spinning process*

- A disc made up of copper (which is faces the nozzle) rotates at a very high speed. When the gas pressure is increased, the melt is sprayed through the fine nozzle and it is directed on to a spinning copper disc.
- The molten alloy falling on the cooled copper disc is solidified and forms a thin ribbon, about 0.0025 cm thick and 15cm wide, which is wound on a spool.
- The main advantage of this method is that we can obtain a continuous ribbon of metallic glass of uniform cross section. Its thickness may be varied by varying the speed of the spinning disc.

Metallic glasses can also be prepared by **other techniques**.

- ***Twin-Roller system*** A molten alloy is passed through two rollers rotating in opposite directions.
- ***Melt-extraction system*** A fast-moving roller sweeps off molten droplet into a strip from a solid rod.
- ***Sputtering*** A vacuum deposition process in which atoms are released from a target under the bombardment of positive ions and deposited on a substrate.
- ***BMG*** Recently, a number of alloys with critical cooling rates, low enough, is employed to form an amorphous structure in a thick layer(>1mm). Thus prepared materials are known as **bulk metallic glasses**.

9.2.3 Properties of Metallic Glasses

These materials exhibit a number of superior properties than their crystalline counterparts, which set them apart from crystalline metals.

Structural Properties

- They do not contain grain boundaries and dislocations.
- They have tetrahedral close-packing structure.
- They do not have long-range anisotropy.
- They can be easily drawn into thin ribbons.

Mechanical Properties

- They are stronger than metals and alloys, as they do not have grain boundaries and defects.
- They have better corrosion resistance.
- They are highly ductile.
- Their yield strength and hardness are much higher than crystalline phase, but the elastic constants are smaller.
- They are very susceptible to the brittle fracture, which limits their practical application.

Electrical Properties

- They have high electrical resistivity but do not vary much with temperature; hence eddy current losses are very small.
- They can have both positive and negative Hall coefficients.
- Some metallic glasses based on Mo, Nb, Ru, Rh exhibit superconductivity also.

Magnetic Properties

- Metallic magnetic glasses containing ferromagnetic transition metals such as Fe, Co, Ni with metalloids like B, Si display superior soft magnetic properties.
- They show very narrow hysteresis loops and so and they have very low hysteresis energy losses.
- They do not have long-range anisotropy and grain boundaries. Therefore, domain walls move with extreme ease in them.
- The coercivity of ferromagnetic metallic glasses is lower due to random fluctuations in the amorphous phase.
- They have higher magnetic permeability and hence they can be magnetized and demagnetized easily.
- The Curie temperature is slightly lower than that of their counterparts.

Chemical Properties

- They have excellent corrosion resistance in a variety of chemically hostile environments.
- Their surfaces exhibit a distribution of chemically active sites which lead to catalytic properties.

9.2.4 Applications

- Due to their higher tensile strength they are used widely as reinforcing elements in concrete, plastic and rubber.
- Because of higher ductility and good corrosion resistance, they are used in the production of springs for different applications.
- Because of good corrosion resistance, they can also be used for cutting and making surgical instruments. They can also be used as a prosthetic material and as implants in the human body.
- Magnetic properties of these materials are not affected by irradiation hence they are used in making containers for nuclear waste disposal.
- As they have higher electrical resistivity and about zero temperature coefficient of resistance, they are successfully used in making cryothermometers, magnetoresistance sensors and computer memories.
- The homogeneity, lower coercive force and higher permeability of metallic glasses make them suitable for numerous magnetic applications.

- Thin sheets of metallic glasses are used as the core materials for power distribution transformers, which increases efficiency of power distribution and transmission.
- They are used as inductors in transformers, recording devices, magnetic shielding, motors, etc.
- They are used in the preparation of magnets for fusion reactors and for levitation trains.

NANOMATERIALS

9.3

9.3.1 Nano Science and Technology

Development of nanomaterials and its technology is the major thrust area of research across the globe. This is due to its vital role in the development of science, engineering and technology. At the nanoscale, the properties of materials change significantly. Their physical, chemical and mechanical. properties alter greatly compared to their normal large-sized grain counterparts with the same chemical composition. *Nanoscience* is the study dealing with the fundamental principles of molecules and structures with at least one dimension roughly less than 100 nm. (**1nm = 10^{-9} m = 10^{-3} μm = 10 Å**) Nanotechnology refers to the design, characterization, production and application of structures, devices and systems by controlling shape and size at the nanometre scale.

These objects are so small that they can only be viewed or imaged either with a very powerful electron microscope (**SEM**) or with an atomic force microscope (**AFM**). They can be rods or wirelike, platelike, or complex-geometry-like. Materials that have one dimension in the nanoscale and are extended in the other directions are nanolayers such as thin films or surface coatings. Materials with nanoscals in two dimensions are referred as nanowires and nanotubes. However, materials having nanoscales in three dimensions are nanoparticles, e.g., precipitates, colloids.

Nanomaterials exhibit far superior properties compared to their bulk counterparts. Intrinsic physical *properties such as density, conductivity, chemical reactivity, etc., of bulk materials (e.g., Cu wire) are independent of their size*. On the other hand, at the ***nanoscale, almost all physical, chemical, electrical, etc., properties of materials become size dependent***. The properties of nanomaterials alter significantly due to the following reasons.

- As particle size decreases, a greater number of atoms are found at the surface than inside and hence they have a larger surface area compared to the same mass of material produced in a larger form. Due to this fact, they become chemically more reactive and attain better strength and electrical properties.
- Quantum effects also contribute a lot which affects the optical, electrical, and magnetic behaviour of such materials.

9.3.2 Some Important Nanomaterials

Fullerene

- Elemental carbon has allotropes such as diamond, graphite, and fullerene.
- They were isolated in macroscopic amounts in 1990.
- The name fullerene was given after **R Bukminster Fuller,** the inventor of the geodesic domelike molecule which conforms to the same underlying structural formula.

- It is also known as bucky ball as its structure resembles a football. In this structure, the atoms lie at the vertices of a polyhedron with 12 pentagonal faces and any number (other than one) of hexagonal faces. It is a hollow pure carbon molecule as shown in F**ig. 9.2**.
- Each carbon atom is bound to three other carbon atoms in a pseudo-spherical arrangement consisting of alternating pentagonal and hexagonal rings just like a football.
- Every carbon atom in this structure is equivalent, so NMR spectra of this molecule shows a single line.
- They can be recovered from the inside of the scrapes of chimneys.

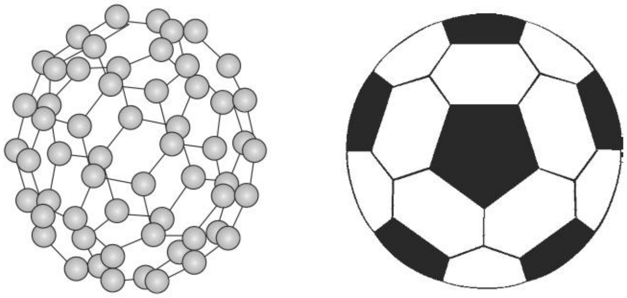


Fig. 9.2 *Chemical structure of Buckminsterfullerene C60*

Quantum Dots

- A quantum dot is a nanoparticle with some special properties. It is a too small crystal such that its properties are subject to quantum effects. They are also referred as zero-dimensional objects.
- A quantum dot is not as small as an atom but is not much larger also. The smallest such quantum dot may contain only 3 atoms.
- Same quantum dots are made up of virtual matter consisting of confined electrons, with which their insignificant masses manage to display a crystal structure.

Carbon Nanotubes (CNT)

- Carbon nanotubes are tubular form of carbon atoms, and look like rolled graphite sheets into cylindrical form. Their diameter is of the nanoscale but their lengths can be extended up to several micrometres.
- They are made up of a hexagonal network of covalently bonded carbon atoms. They are of two types, namely, single walled (SWNT) and multi walled (MWNT). SWNT consists of a single grapheme cylinder and MWNT consists of several concentric grapheme cylinders.
- The structure of CNTs influences their many properties such as electrical and thermal conductivity, density and the lattice structure. These structures give rise to excellent electronic, thermal, mechanical, optical and chemical properties which pave the way for the future applications.
- The diameter of the structure is very important. The wider nanotubes behave like graphite.

9.3.3 Preparation

Nanomaterials can be prepared by a variety of methods such as physical, chemical, biological, self-assembly or hybrid methods. Commonly used methods by the workers in this field are discussed briefly.

1. High-Energy Ball Milling

Basically, it is solid-state material processing and a physical method. In this method, the grain size in powder samples is reduced to nanometre size by the ball-milling process. A container is filled with stainless steel balls with diameters of a few millimeters, and then material to be crushed is added in the form of powder. A rotating shaft grinds the material in the presence of liquid nitrogen. Though this method has the problem of contamination, highly poly disperse size distribution and low surface; it is widely used for preparing magnetic and catalytic nanoparticles efficiently.

2. Chemical Vapour Deposition Method (CVD)

It is a well-known process in which vapour of the material gets deposited on a heated surface through a chemical reaction route. In thermal CVD, the reaction is activated by a high temperature. The deposition chamber is an evacuated chamber, which is well connected with a gas-supply system. In the presence of gas at a medium pressure, a dissociation or reaction between two species takes place, and the newly formed molecule deposits on the heated wafer surface as a nanolayer. Silane (SiH_4) nanolayer can be formed by dissociating hydrogen gas into elementary silicon.

3. Sol–gel Process

A sol (aerosol) is a colloidal dispersion in liquid, and gel is a jellylike substance. Nitrates or carbonates are dissolved in d-ionised water. This solution is maintained at a suitable temperature, and then some amount of gelling agent is added to it. Sol−gel forms through stages of hydrolysis, condensation, growth of particles, and agglomeration of particles. Therefore, physical conditions such as viscosity, pH, temperature, etc., in this process are very crucial. The nanomaterials in the form of thin-film coatings can be prepared by this method. To get a thin-film coating, substrates like copper, nickel, or glass may be dipped in the solution before gel formation, and then annealing has to be done.

This method is most widely preferred due to the following reasons.

- Very high purity (99.9999%) can be obtained.
- Large quantities of nanomaterials at a cheaper cost can be produced.
- Co-synthesis of two or more materials simultaneously is possible.
- Microstructure, physical, chemical, and mechanical properties of the final product can be controlled in this technique.
- It can produce extremely homogeneous alloys, ceramics and composites.

Nanomaterials can also be prepared by arc method, flame synthesis, laser method, inert gas condensation and electrodeposition.

9.3.4 Properties

- These materials are more brittle.
- They are very strong, hard, and tough. They are made up of small crystalline grains and the boundaries between the grains arrest the propagation of defects. The interface area within the material greatly increases at a nanoscale and hence its hardness, toughness and strength increases.

- The atoms in the clusters have lower coordination numbers than the interior atoms. Decreasing the size of the clusters up to a nanoscale, therefore, increases the magnetic moments, e.g., rhenium clusters show ferromagnetism at the nanoscale. They also exhibit higher coercivity and magnetization.
- They show super plasticity even at lower temperatures. They can be deformed largely without any fracture.
- The melting point reduces significantly on reducing cluster/particle size.
- The reactivity of these materials is excellent as compared to that of bulk on account of higher surface-to-volume ratio of materials at the nanoscale.
- Due to change in size, their electrical and thermal conductivity can be altered drastically.
- Clusters of different sizes have different absorption spectra, e.g., nanoscale gold particles look orange, purple, red or greenish depending on the size of the cluster.
- CNTs can be either metallic or semiconductors, depending on their structure. Some nanotubes have conductivities higher than copper, but others behave more like silicon.

9.3.5 Applications

Due to some unique and important physical, chemical and mechanical properties, they are used in a number of applications. Some of them are enumerated below.

(i) Nanomaterials such as tungsten carbide, tantalum carbide, titanium carbide, etc., are much harder, rugged, wear resistant, and corrosion resistant, have better shelf life and are widely used ***in cutting-tool applications***. Their use not only reduces production cost but also increases productivity.

(ii) They can be used ***to make soft and permanent magnets with ease***.

(iii) Nanomaterials by sol–gel route results in a foamlike structure called an 'aero gel'. They are extremely light weight, porous and yet tougher. Because they are porous they can trap air, and are lavishly used ***as thermal insulators instead of PUF (Poly Urethane Foam).***

(iv) They are also used ***as 'smart' window materials, which adjust themselves to the brightness or darkness of daylight.***

(v) Due to their smaller impedance, they are used ***in impedance matching***.

(vi) They are used ***for the fabrication of signal processing elements*** such as filters, switches, wave guides, etc.

(vii) They can be used ***in optical devices*** such as semiconductor lasers, ***electronic devices*** like nanotransistors, memory devices and ***magnetic devices*** like magnetic storage devices, recording heads, etc.

(viii) Nanomaterials such as tungsic oxide gel (WO_3 xH_2O) are used ***in low cost, very large flat panel electrochromic displays.***

(ix) Due to their enhanced chemical activity, they can be used ***as catalysts*** to react with toxic and noxious gases in automobile catalytic converters and power generators ***to prevent environmental pollution***.

(x) Nanocrystalline materials such as silicon carbide and silicon nitride are used ***in automotive applications such as high-strength springs, ball bearings, valve lifters, etc.,*** due to their excellent physical, chemical and mechanical properties.

(xi) They are used as components ***in high-temperature furnaces***.

(xii) Nano crystalline sensor materials are extremely sensitive materials and so, they are used as smoke detectors, ice detectors on aircraft wings, automobile engine performance sensors, etc.

(xiii) CNTs are used in several applications such as ***membranes for molecular transport, filling materials in polymer composites, nanoscale reactors and channels, fuel cells, electrodes in electrochemistry,*** etc.

(xiv) Nanowires are extremely thin wires with a diameter of the order of a few nanometres or less. They have about 10-atom thickness, but length can be longer even up to thousands of nanometres**!** They are prepared from a growth at molecule-to-molecule organization. Due to their excellent semiconductor properties, they are used ***in Field Effect Transistors (FET), Light Emitting Diode (LED), nanolasers, solar cells, etc***. A tiny nanowire is an extremely sensitive biological sensor, hence, they are used ***in DNA tests and other diagnostic tests with great accuracy and faster results.***

This field is fast developing; new applications and new materials with improved properties are developed at a much faster rate.

SHAPE MEMORY ALLOYS

9.4

9.4.1 Introduction

A majority of the new inventions come from the observations of nature and its analysis. The human body and biological systems are also wonderful objects of study. Biological functions such as sensing, actuation and control by the organs also inspire material scientists and engineers. In order to develop better and improved functional materials, such smart or intelligent materials were explored. Shape memory alloys are such smart materials which can also perform functions like sensing, actuation and control. These materials can respond to external stimuli (physical conditions) such as temperature, pressure, moisture, light, electric field, magnetic field, etc.

The development of such materials opens new avenues of applications such as aerospace, medical engineering, automobile engineering, civil engineering, electronics engineering etc. A **Shape Memory Alloy (SMA)** is an alloy that 'remembers' its shape, and can be returned to that shape after being deformed on application of heat to the alloy. When an SMA is cold (or below its transformation temperature), it has a low yield strength and hence can be deformed easily into any shape. But, it regains its original shape, when the material is heated above its transformation. This effect is known as **Shape Memory Effect (SME)**, and the alloys are termed as **Shape Memory Alloy (SMA).** The shape-recovery process occurs due to solid–solid phase change on account of molecular rearrangement even though they are closely packed. Some piezoelectric and fibre optic sensors with this quality are widely used in several applications.

9.4.2 Two Phases

There are two stable solid phases which occur in SMA. They are

1. the high-temperature phase, called ***austenite***(small platelet structure), and,
2. the low-temperature phase, called ***martensite*** (needle like structure)**.**

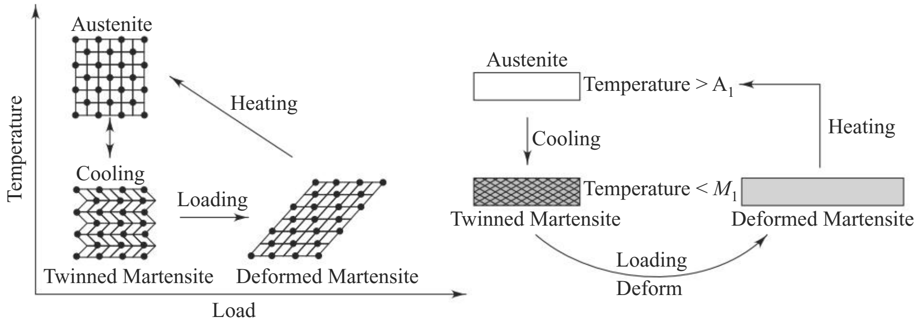
SMAs are metals which have two very unique properties namely,

Shape Memory Effect (SME) and Pseudo Elastic Effect or, Super Elastic Effects (SE).

Shape Memory Effect

The SME was first pioneered in 1932 with a silver–cadmium alloy. It is the property of the alloy by which a specimen deformed at lower temperature returns to its original shape when heated to a higher temperature.

At the phase transformation temperature, the alloy undergoes a crystalline reversible solid-state phase change from martensite to austenite. Both these phases have different properties. It is possible to have a transformation temperature below ambient temperature and such an alloy behaves like a spring. **Figures 9.3 and 9.4** explain clearly crystallographic formation of martensite and its recovery to austenite upon heating (with or without loading).



(a) (b)

Fig. 9.3 *Crystalline arrangement during SME*

It is clear that the high-temperature austenite structure undergoes twinning upon cooling. This twinned structure is microscopically a needlelike structure known as martensite. This phase is relatively soft and easily deformed phase which occurs at lower temperature. On applying external load (stress), this phase wears a particular shape due to detwinning and is known as martensite. In this process, the material undergoes a large elastic strain. However, upon heating, the deformed martensite returns to the stable austenite structure by recovering from the elastic strain. Thus, SME can be explained.

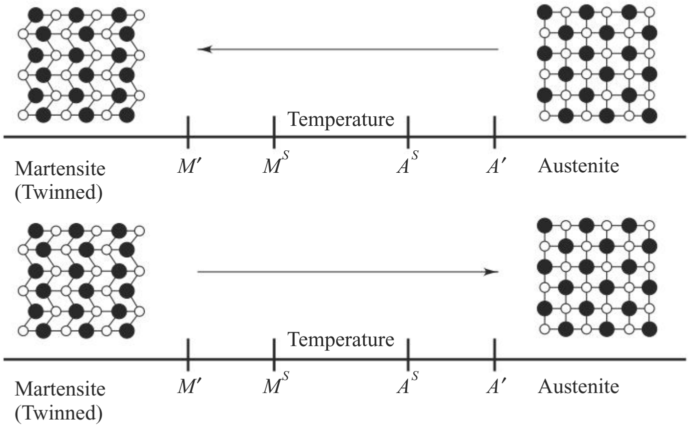


Fig. 9.4 *Temperature induced phase change of SMA without loading*

Pseudo Elastic Effect

The ability of SMA to regain its original shape upon unloading even after a substantial deformation is known as Super Elasticity (SE) or pseudo elasticity. This phenomenon occurs at a constant temperature, hence it is purely stress induced.

On increasing the stress (load) on the SMA at a constant temperature, the austenite phase transforms into a martensite phase. However, as soon as the loading is decreased, the martensite phase begins to transform back to austenite phase. The temperature during this effect is well above T_f, i.e., the temperature at which the reverse transformation from martensite to austenite starts. In this effect, the material is found to be extremely elastic, nonlinear but temperature and strain dependent.

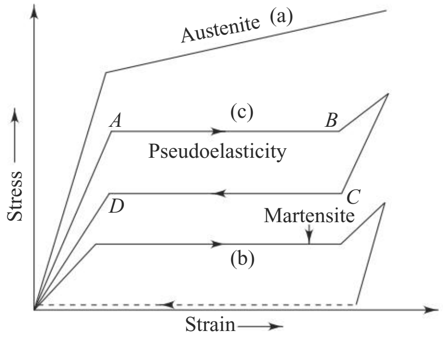


Fig. 9.5 *Stress–strain curves of different shapes of SMA at constant temperature*

The stress–strain relation of different phases at constant temperature of SE made of nickel–titanium is displayed in **Fig. 9.5.**

Figure 9.5 shows that the austenite phase has much higher yield and flow stresses, at a temperature above T_f. Martensite is easily deformed (strained) at a lower stress. This martensite reverts to the austenite phase after removing the stress as shown by the dashed line. No such shape recovery is found in the austenite phase upon straining or heating, as no phase change occurs therein. The above discussed two behaviours are termed as thermomechanical behaviours. The figure also shows an interesting stress–strain property of SE material as curves *AB* and *CD*. Here, the material is completely of austenite phase ($t > T_f$). At this constant temperature, the martensite phase can have stress induced strain (curve *AB*). However, upon unloading the material returns to its original shape (curve *CD*) of austenite phase without heating. This is a pure mechanical behaviour and is referred as pseudo elasticity or super elasticity.

9.4.3 Commercially Important SMAs

A wide variety of alloys display SME but only a few of them are used in commercial applications. Nickel–Titanium (Ni–Ti) alloy is a most commonly used SMA. Its generic name is **NITINOL** (Nickel Titanium Naval Ordnance Laboratory). Other commercially used SMAs are Cu–Al–Ni, Cu–Zn–Al, Co–Ni–Al and ferromagnetic shape memory alloys (**FSMA**) such as Ni–Mn–Ga.

9.4.4 Merits and Demerits

Merits and Applications

- They are **bio-compatible** and hence can be used in **medical applications** such as stents in the human body, vena cava filter used to trap blood clots, dental and orthodontic arch wires, nitinol needle wires to locate and mark breast tumor sites, to make tweezers to remove foreign objects through small incisions, as guide wires for catheters through blood vessels, to design microsurgical instruments and micro grippers.
- It has memory effects to the physical conditions and hence they can be used in number of applications such as making of eye-glass frames, cellphone antennas, fluid fittings, fine tuned helicopter blades and SMA couplings in **aerospace and marine industry**, spring actuators, clutch systems, thermostats, oil pressure control unit in **automobile industries**.

- They have good mechanical properties (strong and corrosion resistance) and hence, they are used as couplers and fasteners, as industrial valves.

Demerits and Limitations

- Relatively expensive.
- Poor fatigue properties
- Technology still in development stage.

BIOMATERIALS

9.5

9.5.1 General Information

In order to replace a damaged or injured part of a human body, novel materials or biomaterials came into light. A biomaterial can be defined by many ways but explicitly it can be defined as the following. ***A biomaterial is any material, natural or artificial, which comprises whole or part of a living structure or biomedical device that can perform, augment or replace a natural function***. Due to many advantages and their successful uses in medical applications, this field is rapidly developing. Since they are used to make devices that would be in close or direct contact with the body to augment or replace a damaged or injured part, these materials have made a major impact in medicine science and patient care to improve the quality of life of humans. They are used in numerous applications such as artificial skin, artificial blood vessels, artificial heart and its parts, dental fillings, pacemakers, wires, pins and plates for bone repair, total artificial joint replacements, contemporary medical devices, smart drug delivery systems, diagnostic products, etc.

9.5.2 Types

They can be classifies as ***metals and alloys, polymers, ceramics and glasses, carbon and its composites***. Such materials are used as molded or machined parts, coatings, fibres, films, foams and fabrics. They are prepared and developed to have a combination of properties. They must meet following strict standards, criteria and stringent controls before they are used as prosthetic material or devices.

- ***The Biofunctionality (Functional Performance)*** In order to use them as prosthetic devices to deliver their desired function most effectively, the physiochemical properties such as electrical and thermal conductivity, corrosion resistance durability, the adverse side effects in case of failure of the device, mechanical properties like strength, toughness, fatigue, hardness, etc., are to be evaluated and analysed.
- ***Biocompatibility*** It is the ability of a material to perform with an appropriate host response in a specific application. This material should not undergo degradation in its properties within the biological environment of the body and should not cause any adverse reaction. The various processes that may degrade the biomaterials are corrosion, chemical modification, swelling, leaching and wear, etc. The properties which may get adversely affected are wear resistance, fracture, toughness, strength, surface roughness, etc.

To meet such requirements the following biomaterials are used.

1. ***Metals and Alloys(Biometals)***

   Currently three types of biometals are in vogue, namely, **pure Ti–Ti alloy, stainless steel 316L, Co-Cr–Mo alloy.**

Due to their excellent mechanical properties, they are widely used for load-bearing implants such as simple wires, plates, pins, screws used in orthopedic or orthodontic applications. They are also used in cardiovascular surgery and maxillofacial surgery.

2. ***Polymers***

   There are polymers such, **poly amides(nondegradable), Poly propylene(PP), Polyethylene (PE), Poly Tetra Fluoro Ethylene(PTFE), Poly Ethylene Terephthalate(PET), Polyesters, Poly Methyl Methacrylate(PMMA), etc.**

   The polymers may undergo either **unintentional degradation or controlled degradation.** Chemical reactions such as oxidation, hydrolysis, and reduction may cause unintentional degradation. This may release chemicals and can induce adverse local systematic host reaction which may cause clinical complications. However, some polymers such as polylactic acid, polyglycolic acid, polyorthoesters, etc., may be synthesized to undergo controlled degradation. These materials degrade into smaller fragments and get eliminated by routine metabolic processes of the body.

3. ***Ceramics and Glasses(Bio Ceramics)***

   The ceramic materials are nonmetallic inorganic compounds. These materials are stiff, corrosion resistant, excellent wear resistant and lighter, with constraints such as brittle, difficult to process, poor tensile strength etc. Bioceramics such as **alumina, zirconia, porcelain, bio active glasses, carbons, apatite ceramics** are widely used in many medical applications. Alumina has excellent corrosion resistance, high compressive strength, good biocompatibility and excellent wear resistance properties. Therefore, it is widely used in dentistry since many decades**.**

4. ***Composites***

   Composites of biomaterials such as PMMA-glass filters and BIS-GMA-quartz filter are used in the field of dentistry as restorative material. They are lighter and still stronger; so they can be used for prosthetic limbs.

Besides these types of materials, **natural biomaterials** derived from animals and plants such as **collagen, keratin (obtained from hair), cellulose (obtained from plants), and chitin (obtained from insects or crustaceans)** can also be used. The main advantage of these materials is that they are nontoxic as they are materials familiar to the body. However, they have limitations such as immunogenic and tendency to decompose at temperatures below melting point with natural polymers of the body.

9.5.3 Applications

Applications of each type of materials have been enumerated. However, some of the popular applications are as under:

- They are commonly used as implants like hip joints, artificial lenses.
- Peptides and DNA molecules are used as building blocks to build new nanostructures.
- They are used for improved dressings for chronic wounds.
- Regeneration of damaged or diseased tissues, e.g., nerve regeneration for spinal chord injuries can be engineered using the tissues.
- ‘Smart’ surfaces for the culture of embryonic stem cells, artificial muscles with the help of electroactive polymers can be developed.
- ‘Bio nanotech’ engineering approaches can be used to manipulate cell function.

Multiple-Choice Questions

1. Shape memory alloys can repeat deformation and shape recovery
(a) once only (b) twice only
(c) 100 times only (d) for unlimited times

2. The ability of SMA to regain its original shape upon unloading even after a substantial deformation is
(a) elastic effect (b) pseudo elastic effect
(c) hysterisis effect (d) none of the above

3. Thermal conductivity of CNT is
(a) very high (b) very low
(c) neutral (d) negative

4. Bulk metallic glasses have their strength-to-weight ratio
(a) high (b) low
(c) positive but less than one
(d) negative

5. CNT can be used to form
(a) inductor (b) resistor
(c) transistor (d) transmission lines

6. Temperature coefficient of resistance of metallic glasses is
(a) high (b) very high
(c) positive but less than one
(d) negative

7. Fe_{67} Co_{18} B_{14} Si and Fe_{40} Ni_{38} Mo_4 B_{18} are the examples of
(a) metal–metal (b) metal–metalloid
(c) metal–nonmetal (d) none of the above

8. Twin roller system is used to prepare
(a) nano material (b) biomaterial
(c) SMA (d) metallic glasses

9. Nanomaterial can be prepared using
(a) ball milling (b) melt spinning
(c) twin roller system (d) none of the above

10. Containers for nuclear waste disposal can be prepared using metallic glasses due to their
(a) magnetic (b) chemical
(c) mechanical (d) electrical properties.

11. NMR spectra of fullerenes show
(a) single line (b) single band
(c) multiple lines (d) none of the above

12. On decreasing the size of the clusters up to a nano scale, the magnetic moments
(a) remain constant (b) increase
(c) decrease (d) none of the above

13. Nano wires can be used to make
(a) LED (b) transformers
(c) antenna (d) transmission lines

14. The interface area within the material at a nano scale
(a) remains constant (b) increases
(c) decreases (d) none of the above

15. NITINOL is a commercially used SMA of
(a) nickel–tantalum (b) nickel–titanium
(c) nichrome–titanium
(d) nichrome–Tantalum

16. The high-temperature phase of SMA is called
(a) austenite (b) martensite
(c) austmatensite (d) none of the above

17. The natural biomaterial *chitin* can be obtained from
(a) hair (b) insects
(c) plants (d) flowers

ANSWERS TO MCQ

1. (d) 2. (b) 3. (a) 4. (a) 5. (c) 6. (c) 7. (b) 8. (d)
9. (a) 10. (a) 11. (a) 12. (b) 13. (a) 14. (b) 15. (b) 16. (a)
17. (b)

Short-answer Questions

1. What are the differences between glasses and metallic glasses?

2. Enumerate 'new engineering materials' and explain their significance.

3. Explain briefly the process of making of metallic glasses.

4. Why is it easier to form metallic glasses from alloys than from pure metals?

5. Describe unique properties of metallic glasses.

6. List out common techniques used in preparing metallic glasses.

7. What are the types of metallic glasses? Give their examples.

8. List out mechanical/electrical/ magnetic properties of metallic glasses.

9. Give applications of metallic glasses.

10. What are advantages of use of metallic glasses in transformers?

11. What are nanoscience and nanotechnology?

12. What are advantages of nanotechnology?

13. How are nanomaterials refereed in one, two, and three dimensions?

14. What are quantum dots, nanowires, nanotubes, nanocolloidal?

15. Why do properties of nanomaterials alter and become size dependent?

16. Enumerate some important nanomaterials.

17. Explain the structures of bucky ball, CNT and quantum dots.

18. How can one see or image nanomaterials?

19. What are the types of CNT? Give their differences.

20. List out the methods of preparing nanomaterials.

21. Why is sol–gel method widely preferred?

22. Enumerate properties of nanomaterials.

23. What are applications of CNT?

24. What are the applications of nanomaterials?

25. Which nanomaterials are used in cutting-tool applications? Why?

26. What are Shape Memory Alloys (SMA)?

27. What is pseudo elasticity? Explain shape memory effect.

28. State and explain types of SMA.

29. What are smart materials? Where they are used?

30. State unique properties of SMA.

31. Name the crystal structures or phases exhibited by SMA.

32. What causes the SMA to remember its shape?

33. Explain the types of phase transformations in SMA.

34. What are Ni-Ti alloys? Give its commercial name.

35. List out the various fields of applications of SMA.

36. What are merits and demerits of SMA?

37. What is ferromagnetic SMA? Where they are used?

38. What are biomaterials?

39. What are the qualities that bio materials should have for using them in devices?

40. Enumerate the types of biomaterials.

41. Explain biofunctionality.

42. Explain biocompatibility.

43. List out the processes by which biomaterials can degrade.

44. List out the types of biomaterials.

45. What do you mean by controlled degradation and biodegradable materials?

46. Give names of commercial biomaterials.

47. Mention the applications of biodegradable materials.

48. What are natural biomaterials?

49. What are merits and demerits of natural biomaterials?

50. How are natural biomaterials obtained?

Long-answer Questions

1. Discuss the properties, types and applications of metallic glasses.

2. Write a short note on melt spinning method to prepare the metallic glasses.

3. What are the various properties of metallic glasses. Explain how they can be used in various applications.

4. Discuss the properties, types and applications of nanomaterials.

5. What is nanotechnology? Explain various applications of nanomaterials.

6. Write a short note on carbon nanotubes.

7. Discuss the methods of preparing nanomaterials.

8. Discuss structures and applications of some important nanomaterials.

9. Write a short note on SME and SMA.

10. Discuss temperature induced crystallographic changes in SMA.

11. Explain stress–strain behaviour of SMA at a constant temperature.

12. Write a short note on pseudoelastic effect of SMA.

13. What are commercial SMAs? Give their merits and demerits.

14. Discuss the properties, types and applications of biomaterials.

15. Write a short note on biomedical materials.

16. Write a short note on natural biomaterials.

17. Discuss the applications of biomedical materials.

Questions From Previous Exams

(i) What do you mean by metallic glasses? **(GTU- Jan′ 2009)**

Metallic glasses are metal alloys that are amorphous. They do not have a long-range atomic order. They share the properties of both metals and glasses.

(ii) Discuss the properties, types and applications of metallic glasses. **(GTU- Jan′ 2009)**

(Refer Section: 9.2)

(iii) Write the properties of metallic glasses. **(GTU- Jan′ 2010)**

(Refer Section: 9.2)

(iv) What is nanotechnology? Write the application of nanotechnology. **(GTU- Jan′ 2010)**

(Refer Section: 9.3)

(v) What is a biomaterial? What are the types of biomaterials? **(GTU- Jan′ 2010)**

(Refer Section: 9.5)

(vi) Give any two applications of biomaterials. **(GTU- June′ 2010)**

Artificial body parts, ceramic material.

(vii) Mention any five properties and application of nanomaterials. **(GTU- June′ 2010)**

(Refer Section: 9.3)

(viii) Explain the types, properties of metallic glasses and melt spinning technique to prepare the metallic glasses. **(GTU- June′ 2010)**

(Refer Section: 9.2)

(ix) Short notes (1) Shape Memory Effect **(GTU- Jan′ 2009) (GTU- Jan′ 2010)**

(Refer Section: 9.4)

(x) Draw only the material phase transformation diagrams for shape memory effect (temperature Vs load and pseudo-elasticity (stress Vs strain) **(GTU- June′ 2010)**

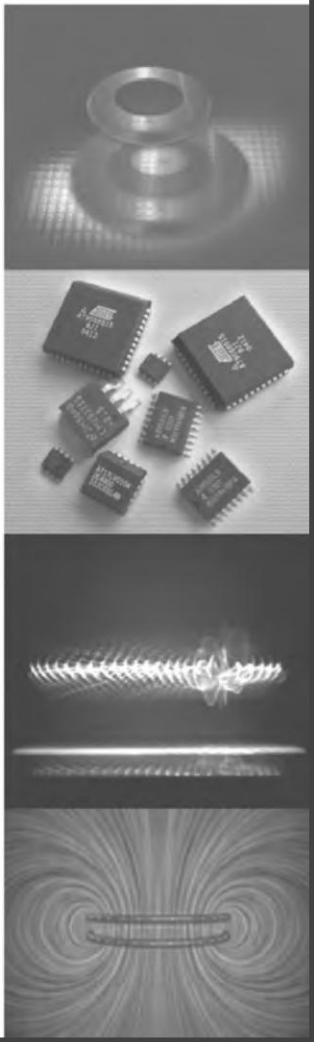
(Refer Section: 9.4)

The McGraw-Hill Companies

Non-Destructive Testing

10

Suppose there is a crack in the wall or gate of any dam which is full of water. Think of a nuclear reactor or chemical reactor with some flaw. How dangerous it can be! They have to be inspected and tested to avoid major damage and accidents. How can we test and measure such things? Every engineer must know the basic methods of non-destructive testing. A list of such techniques and their methods are described in this chapter.



INTRODUCTION

10.1 An industrial product is designed to perform a specific function most reliably and most accurately. The reliability of a product depends upon the individual reliability factor of all the parts or components used in that product. Thus, to produce high-quality products, it is important that each individual component is reliable and performs its function accurately for an assigned period of time. (Just like each player plays well in a winning team! One weak or bad player can gift you a defeat!)

The quality of products, components or parts depends much upon many factors like the design, raw material, and fabrication techniques. Quality is also related to the presence of the defects and imperfections present in the components or finished product which impair the performance level. Knowledge of these defects with a view to detect and evaluate them and then minimizing in the product is essential to achieve improved or acceptable level of quality. Therefore, the techniques based on physical principles such as Non-Destructive Testing (NDT), Non-Destructive Evaluation (NDE), Non-Destructive Inspection (NDI) have been developed. Non-Destructive Testing (**NDT**) refers to the entire test methods which detect the harmful defects and imperfections in the finished products or their part without affecting its future usefulness, i.e., without damage or reduction of the service life of the product or component. After the test process, the product can be used again for which it was designed.

OBJECTIVES OF NON-DESTRUCTIVE TESTING

10.2

Defects, faults, flaws, or imperfections can occur during manufacturing, or during assembly, or installation, or during pre-service life. Such irregularities are inevitable so, with the help of NDT, one can detect the defects and give information about their distribution. The main objectives of NDT are enumerated as under:

- ***To improve productivity and hence to increase the profitability***

  With the help of NDT, useless wear and tear on the plant can be avoided. Testing at an early stage of basic workpieces like casting, welds and components increases the quality of product and hence the profit.
- ***To increase service life of the product***

  Faults, defects and flaws are known to decrease the service life of the products or components used therein. By detection of these irregularities and then, taking corrective measures accordingly, can increase service life of the product greatly.
- ***To prevent accidents and increase the safety***

  Defects may develop during the service life of the product. Such defects may be hazardous, e.g., high pressure reactors and vessels to be used in automobile and chemical industries, aircraft structures, etc. Cracks and flaws can be dangerous and can lead to accidents. So, with the help of NDT we can detect them and can replace if the component/product has defects greater than the allowed limit. Therefore, periodical testing of the component and product is a routine exercise in industry to ensure the safety.
- ***To maintain strict quality control standards***
- ***To improve reliability and precision of the product***

COMPARISON OF DESTRUCTIVE AND NON-DESTRUCTIVE TESTING

10.3

The objectivity of NDT can be well resolved by understanding bellow given table which describes the merits and demerits of NDT over destructive testing.

Table 10.1 *Merits and demerits of NDT over destructive Testing*

Destructive Tests	*Non-destructive tests*
Merits	**Demerits**
Direct and reliable measurement.	Being indirect measurements, their reliability must be verified.
Generally quantitative measurement	Generally, qualitative measurement, so judgmental rather than analytical. Measurements can also be done quantitatively.
Direct relation between material properties and test measurements.	Skilled judgement and experienced professionals can interpret the measurements.

Demerits	Merits
Due to the direct tests on the objects correlation between the objects and the specimen used needs to be established.	Testing on actual component is possible.
A single test may measure only one or a few of the properties.	Many NDT methods can be applied on the same many properties of interest can be measured.
In service, testing is not possible.	In service, testing is possible.
Preparation of the test specimen is costly.	Very little preparation is enough hence relatively cheaper.
Repeated measurements and checks over a period of time are not possible.	Repeated measurements and checks over a period of time are possible.
They are time-consuming methods.	They are very speedy methods.

TYPES OF DEFECTS

10.4

Defects are unavoidably and inherently present in many of the metallurgical specimens. They can be classified broadly in two categories.

- ***Defects, disorders, irregularities present at the surface***

  Cracks developing due to unsatisfactory conditions either during cooling or during heat treatment, *surface scales* and *defects* present at the surface due to material processing or faulty technique are included in this category.

- ***Flaws, cracks, porosity, etc., present below the surface***

  During or after solidification of metals discontinuity such as ***cracks*** due to fracture takes place at subsurface level. During casting (pouring or injecting molten metal into a cavity of desired shape) entrapped gas develops ***porosity, Blow holes, nonmetal inclusions*** (sand, slag, oxides, etc.) in casting. ***Seams*** or surface irregularities such as cracks develop on the slabs or metallic vessels during rolling or extrusion process. ***Lamination*** is the flattened discontinuities. They occur during forging process wherein materials are shaped by pressing or hammering, etc.

VARIOUS METHODS OF NON-DESTRUCTIVE TESTING

10.5

NDT tests are specific for a given application. Selection of the method and equipments necessary for these tests can be based on the specifications as per the suggested codes or standards. It is a matter of fact that no single method is capable of revealing all surface and subsurface discontinuities in all types of materials. Therefore, in order to get complete evaluation of the quality of the product, part or component combinations of tests are carried out.

Most popularly used NDT methods by industrial quality control engineers and scientists are listed as per following:

- Visual inspection
- Liquid penetrant testing
- Magnetic particle (magnaflux) inspection
- Eddy current testing

- Radio graphic methods
  - (a) X-ray radiography
  - (b) Gamma ray radiography
- Ultrasonic inspection methods
- Thermography and pulse echo methods
- The details of some important methods are described in detail in following sections.

10.5.1 Visual Inspection Method

It is the most widely used method as it is simple, easy, quick, and cheaper. It can reveal gross surface defects. The visual inspection method has been developed to a very high degree of precision due to advancement of microprocessors, computers, visual pattern recognition, image processing, etc. Generally, this method is employed on the specimen even though it has to be tested by other NDT methods as well, because this method reveals the following first-hand information.

(i) Surface porosity
(ii) General information of the component
(iii) Presence or absence of corrosive products on the surface of the specimen
(iv) Presence or absence of cracks
(v) Orientation and position of cracks
(vi) Potential sources of mechanical weakness such as sharp notches or misalignment, etc.

The basic procedure used involves the adequate cleaning of the specimen before visual inspection followed by illumination of the specimen with light and then, examined with naked eye or by optical aids such as photodetector, microscope, bore scope, endoscope, flexi scope, telescope, holography, etc. For visual inspection, adequate lighting, i.e., 800–1000 lux is essential. The human eye has excellent visual perception and hence is the most valuable NDT tool. It has limitations for distinguishing the differences in brightness and closer wavelengths. A **microscope** is used by visual inspectors to magnify the minute or finer defects which cannot be detected by the unaided eye. **Bore scope** helps to inspect the inside of a narrow tube, bore, or chamber. The **endoscope** has a better optical system and a high intensity light source than the bore scope. A unique feature of the endoscope is that objects are constantly in focus from about 4 mm to infinity. Hence, it has ‘no focusing’ feature. The **flexiscope** is a flexible fibre optic bore scope which permits the visual inspection of inaccessible regions such as corners and passages with directional changes. **Holography** is a very important technique to obtain an accurate three-dimensional image of a given object. It is used for highly complicated and precision components.

10.5.2 Liquid/dye Penetrant Method

This method is familiar among the professionals either as **LPI** (LIQUID Penetrant Inspection) or as **DPI** (Dye Penetrant Inspection). This method is best suitable for detecting very fine cracks and surface openings as narrow as 150 nm. It is simple to use and relatively cheaper than any other NDT methods. Penetrant inspection uses capillary action to attract low surface tension fluid to the discontinuity as compared to surroundings, which creates a recognizable indication of a crack or other surface opening. It is most widely used to detect casting and forging defects, cracks, and leaks in new products, and fatigue cracks on in-service components. It is not suitable for dirty and rough surfaces.

Basic Principle

When a liquid penetrant is applied over a clean surface to be inspected, the penetrant seeps/penetrates into cavities that are open to the surface by the combined action of surface tension and capillary action. After removal of excess penetrant, development by blotting action of the developer powder and then, inspection either by visible light or by UV light, one can recognize indication of the defect. Fluorescence can also be exploited to improve the reliability and resolution of the method.

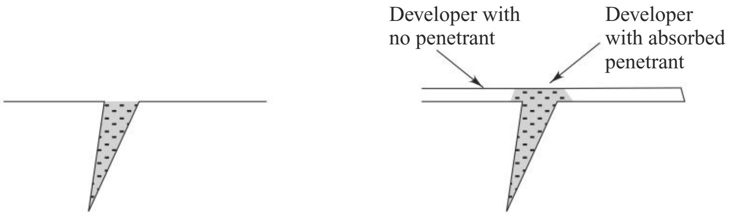
Basic steps involved in Penetrant Inspection

Basic steps involved in penetrant inspection are enumerated as under and illustrated in Fig. 10.1

1. Cleaning It is the foremost step in the penetrant inspection method. Initial cleaning of the surface area to be inspected is very important. If the surface to be inspected prevents the penetrant to enter into the discontinuity due to contaminants like scale, paint, flakes, dirt, grease, etc., it can lead to either masking of real indications or faulty creation of defect indication. A satisfactory combination of solvents, brushes, rags, etchants, etc., must be chosen carefully. It is essential that the cleaned surface be adequately dried before the application of the penetrant because the presence of solvents may dilute the penetrant and decrease the brilliance of the indication of the defects.

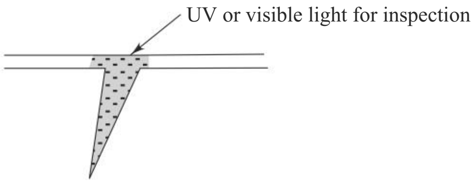
(a) Open defect

(b) Applied penetrant in the crack



(c) Removal of excess penetrant from the surface but leaving it in the cavity

(d) Enlarged width of penetrant due to developer



(e) Detection of defect by visible inspection

Fig. 10.1 *Sequence of steps involved in Liquid/ dye penetrant method*

2. Penetrant Application All liquids can flow into various types of very fine surface openings by capillary action however very viscous (sticky) liquids are unsuitable as penetrants; because they cannot flow rapidly in the cavity hence the *dwel time* (the amount of time required to enter into the cavity) would be more for such liquids. Besides this, the penetrant must also be able to carry a dye that can mark a metal. The dye floated within the liquid penetrant can either be a fluorescent or visible dye. Application of penetrant may be performed on the cleaned surface by dipping, immersing, spraying, brushing, etc. A minimum dwel time of about 20 to 30 minutes may be allowed after penetrant application.

3. Removal of excess penetrant This step though simple involves a lot of skill. In this step, the surface must be clear of penetrant but the defect/cavity/crack must retain all of the penetrant which moved into it. Excess cleaning may remove the penetrant from the upper region of the defect hence developer does not reach the penetrant and, therefore, no defect is indicated. On the other hand, insufficient cleaning will retain a background penetrant on the surface which ultimately results into the poor indication of the defect. To remove excess penetrant, cleaning of a surface using a solvent and then rinsing with water or first treating with emulsifier and then rinsing with water can be undertaken.

4. Application of developer A thin coat of developer is applied over the surface to draw the penetrant out of the crack to enhance visibility of the indication. It acts as a blotter. It also covers the surface with a colour to enable good contrast.

The developer material such as dry powder, aqueous/wet powder-suspension, solvent suspension, plastic film, etc., may be used for the purpose. Wet developers, either aqueous/wet powder-suspension or solvent suspension can be applied on the surface by several techniques like spray, immersion, fluidized bed, electrostatic means, etc. They are preferred when the surface is still damp to touch. Dry powder can be applied either by sprinkling on the surface or by passing the part through a developer dust cloud chamber so that the powder absorbs the liquid from the flaws which are immediately revealed. The dry developer should only be applied to the thoroughly dry surface.

5. Inspection and evaluation The last step in the process is the scanning of the surface for indications. It can be carried out under visible, UV or laser light conditions and the defect recognition may be made with the human eye or with automated optical scanners. Two types of indications are possible, true or false. *True-type indications are caused by penetrant bleeding out from actual defects in the sample.* True indications reveal surface discontinuities such as large cracks, pits, porosities, etc. *False indications are not caused by flaws at the surface of the sample.* They are caused by following incorrect liquid penetrant applications or rough and irregular surfaces on the test material.

Merits and Demerits of Liquid Penetrant Method

Merits

- It is a simple, cheap and well-established method.
- It is used to detect the flaws irrespective of size, shape, internal structure, chemical composition and flow orientation.
- The defect indication can have a high visual contrast.
- The developer draws the penetrant out of the crack over a wide area of the discontinuity and hence it resolves better.

Demerits

- It cannot be used to detect subsurface defects or discontinuities.
- Proper cleaning and removal of excess penetrant is necessary for reliable results.
- It is necessary to check compatibility on the test material especially while using plastic materials.

Applications

- This method is used for detecting surface defects such as fatigue cracks, porosity, lack of joints in welds, etc.

- This method is applicable for a wide variety of materials like ferrous, nonferrous, nonmagnetic metals (plastic, glass or ceramic).
- It is suitable for detecting cracks in fabrication and regrinding of the casting, steam and gas turbine blades, electronic applications, etc.

10.5.3 Radiographic Method

Radiography is an NDT method to indicate the presence and nature of defects or other structural discontinuities in the interior of the materials with the help of ionizing radiations like X rays, gamma rays, electron beam, neutron beam, etc. Radiography is essentially based on the principle of shadow projection. In this process, a permanent image on a film is produced on shinning the material with radio (short wave length) waves. This image may be termed as radiograph, image film or xero-radiograph. It can produce a real-time image which can be viewed on a fluorescent screen or image intensifier plates.

It is most widely used for the detection of internal defects such as porosity and voids. With proper orientation, planner defects can also be detected. It can also be used to know the changes in material composition, thickness measurements, and locating unwanted or faulty components hidden from view in an assembled part.

Basic Principle The basic principle of this technique is illustrated in Fig.10.2. Shorter wavelength radiations have high penetration power in the material. The amount of absorption of this radiation is a function of density and thickness of the material. If there is any cavity or discontinuity in the interior of the material, the beam of radiation will penetrate less in such a defective area than in solid material. This will result in the variation in the absorption of the rays by the material in the defective area. This variation in absorption is measured or recorded as shadow on a film sensitive to X rays or gamma rays, which produces an image and indicates the presence of the defect. Variation in the darkness on the obtained xerograph (using X rays) or gamma graph (using gamma rays) can provide information about the internal structure of the material. The basic set-up includes a source of radiation, the object to be radiographed and a detector (usually, a sheet of photographic film or image intensifier plate).

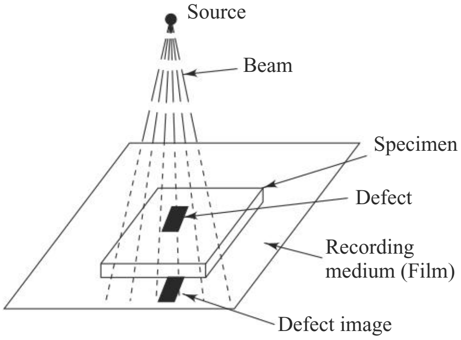


Fig. 10.2 *Basic Principle involved in Radiography*

Absorption of Radiation by the Object under Test

As radiation passes through the material, the intensity of the radiation decreases. This decrease in intensity dI is proportional to the intensity I of the incident radiation and the thickness dx of the absorbing material. Therefore,

$$dI = -\mu I dx \tag{10.1}$$

where μ is a constant of proportionality and the–ve sign is the decrease in intensity.

On integrating this equation, we get,

$$I = I_0 e^{-\mu x} \tag{10.2}$$

where I_0 and I are the intensities of the incident and emergent rays respectively, x is the thickness and μ is the linear coefficient of absorption of radiation for the medium through which the rays are passing. μ is lower for shorter-wavelength radiations.

Two popular radiographic methods are *X-ray radiography* and *gamma ray radiography*.

X ray radiography

The experimental-set up shown in Figure 10.3 illustrates the principle of *X*-ray radiography.

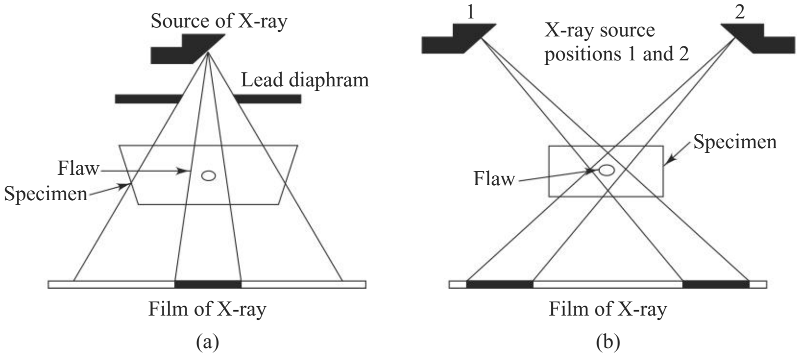


Fig. 10.3 *(a) Principle of X ray Radiography; (b) Loction of the flaw within the specimen volume*

X-rays are short-wave electromagnetic waves of wavelength of about 0.1Å. They can be produced by a Coolidge tube and is known as source. These rays have very good penetration power and so are made to pass through the specimen. The specimen is a plate of uniform thickness containing some discontinuities which have absorption characteristics different from those of the surrounding material. The amount of radiation which impinges on the film in the area below the defect is different from the amount that impinges on adjacent areas. This can produce a latent image of the discontinuity on the film or detector. On developing, this image of the defective area emerges as a shadow of different photographic density than its surrounding.

To get the exact location of the flaw or defect, at least two radiographs with the X-ray source at two different positions, as shown in Fig.10.3b, in a line parallel to film must be taken. It can be obtained using the relation

$$d = \frac{FS}{S + P} \tag{10.3}$$

where F is the distance between the film and the position of the X-ray source, S is the distance between the images of the flaw, and P is the distance between two positions of the X-ray source.

Merits

- No prior preparation of the specimen surface or extra skill is necessary.
- The basic advantage is due to the use of ionizing radiation, which enables examining objects of different shapes and size ranging from microminiature electronic parts to giant missile or power-plant structures.
- This method can be applicable on a large variety of materials.

Demerits

- Hazards due to longer exposure to radiation can cause adverse biological effects on the body.
- It is very expensive.
- Defects cannot be indicated unless they are oriented properly to the radiation.
- Very fine discontinuities like inclusions in wrought materials, flakes, microporosity, etc., cannot be detected easily.

Applications

- It can be used on castings, forgings and welding.
- It can be used on forging and mechanical assemblies.
- This method is widely used to inspect most types of solid materials such as ferrous and nonferrous alloys, nonmetallic materials composites, etc.
- It is well suited to the inspection of semiconductor devices for detection of cracks, broken wires, unsoldered connections, etc.
- This method is very good to inspect the condition and proper placement of components, for liquid-level measurement in sealed components, etc.

10.5.4 X-ray Fluoroscopy

Another method similar to radiography is *X*-ray fluoroscopy. The difference in this method is that the X-ray image is projected on a fluorescent screen instead of being photographed on a photographic film. It is widely used for the inspection before getting approval of the final product, e.g., inspection of fruits before packing, canned foods, etc.

In this technique, a fluorescent screen is held behind the test object instead of a photographic film. The X-rays are converted into visible light and an image of the object can be seen on this screen. The quality of the image obtained on the fluorescent screen is usually very faint and it is difficult to ascertain the details of defects. But by using a suitably sensitive closedcircuit television (CCTV) camera focused on the fluorescent screen, a brighter image can be obtained on the television monitor screen. The radiation hazards can be eliminated because the image is produced on a television monitor, which can be kept at a large distance from the *x*-ray equipment. The image on the television can be converted into digital data and using digital image processing technique, image enhancement can be achieved. This method is also known as real-time radiography. The term *fluoroscopy* is used for real-time radiography. A basic system is shown in Fig. 10.4.

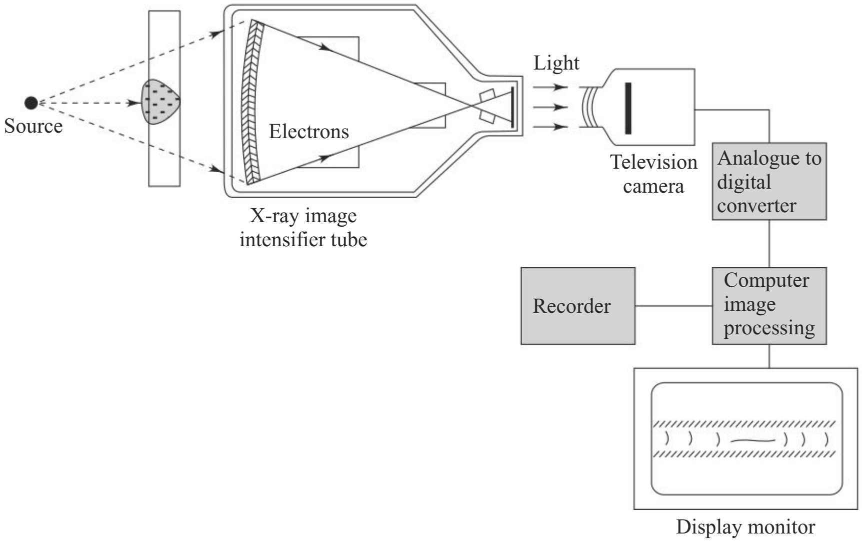


Fig. 10.4 *Schematic diagram of real-time radiography system/ X-ray fluoroscopy*

As shown in Fig.10.4, a source of radiation, test object, X-ray image intensifier, television camera, analogue to digital converter, computer with image processing unit, recorder and display monitor is arranged. The X-rays from the source pass through the object and strike the double layer screen in the image intensifier. The X-rays incident on the fluorescent layer of the screen convert it into light. An electron image of the visible image produced by the fluorescent layer is generated by adjoining photocathode layer. From the double layer input screen, the electrons are accelerated and focused onto a smaller output screen. The output screen now reconverts the electron image into a visible image such that the dim image from the input screen is made hundred to ten thousand times more intense. The television camera views the image and converts it into digital data by the analogue-to-digital converter. The computer now processes the digital output and the real-time image of the object is presented on the display monitor.

10.5.5 Ultrasonic Inspection Method

Ultrasonic testing is a versatile NDT method which has a wider applicability. It is most successfully used to detect surface as well as internal discontinuity such as seams, voids, cracks, laps, blow holes, inclusions, etc. It is already learnt that ultrasonic waves are the sound waves with frequency 20 kHz to 25 MHz. We also know their properties and the production methods.

Basic Principle This method is based on one of the properties of these waves that they reflect at the interfaces and at the change in the medium. These reflections are used to detect the flaws, defects or discontinuity in the form of cracks, blowholes, porosity, air bubbles, etc., in the internal structure of a material. Thus, by transmitting ultrasonic waves to a test specimen and by recording its reflections at the receiving detector, the discontinuities can be analyzed to study the location, shape and size of the defects. Better results can be obtained with higher frequencies (smaller wavelengths) and therefore ultrasonic waves of the frequencies between 0.5 MHz to 20 MHz are used.

Basic Equipments used in Ultrasonic Methods The following equipments are used in ultrasonic test method.

- An electronic signal generator to generate ac voltage
- A transmitting transducer such as piezoelectric device, which transmits ultrasonic waves when ac voltage from the generator is applied.
- A receiving transducer to record the reflected ultrasonic waves from the test specimen and to convert them in the corresponding electrical signal.
- An amplifier to modify the signal from the receiving transducer to project it on a display (CRO).
- An electronic timer to measure the time delay (difference) between transmitted and received signals.

Mostly, a single piezoelectric transducer acts alternatively as a transmitter as well as a receiver. The following self-explanatory block diagram of Fig. 10.5 illustrates the ultrasonic pulse echo method.

Types of Commercial Ultrasonic Test System

Three types of ultrasonic test systems are commercially widely used.

They are

1. Pulse echo system
2. Resonance system
3. Through transmission system

The pulse echo system is most popular among NDT professionals.

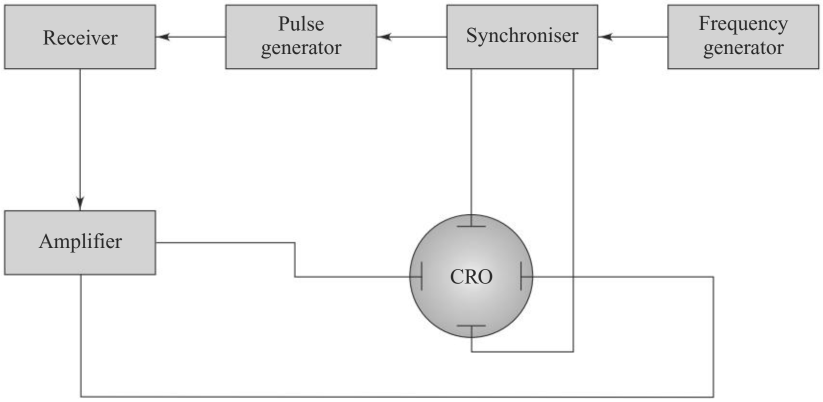


Fig. 10.5 *Block diagram of pulse echo method with basic equipments*

Pulse Echo Method In this system, a single transducer is used for the purpose of transmission as well as reception of the ultrasonic waves as shown in Fig. 10.6. To improve the contact between the test material and the transducer couplets, materials such as glycerin, jelly, thin oil, etc., may be used. From the transmitter, the electric pulse is fed to the transducer probe. The transducer is excited by the electric signals of ultrasonic range of frequencies and vibrates at its resonant frequency. This signal is propagated into the test object through the couplet layer. The pulse also triggers the time base generator of CRO, so that the ultrasonic signal starts to move into the test object. At the same time, the luminous spot moves across the CRO screen. The transducer is applied to the Y plates of CRO through an amplifier.

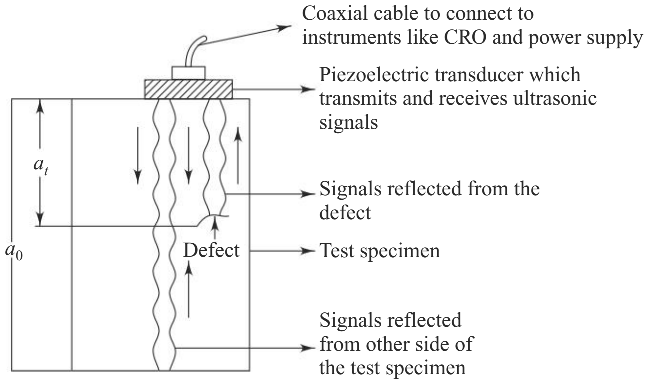


Fig. 10.6 *Pulse echo ultrasonic system*

As the signal passes through the top surface of the test object, there will be a pattern on the CRO screen as shown in Fig. 10.7 and labeled as T therein, but if there is no defect in the test specimen, the reflected pattern will appear at R in the same figure.

However, if there is a flaw within it, the wave will be reflected (echo) and a change in pattern will appear on the screen as shown in Fig. 10.8 and labeled as E therein. The specimen is also tested through its entire thickness, marked as R in Fig. 10.8.

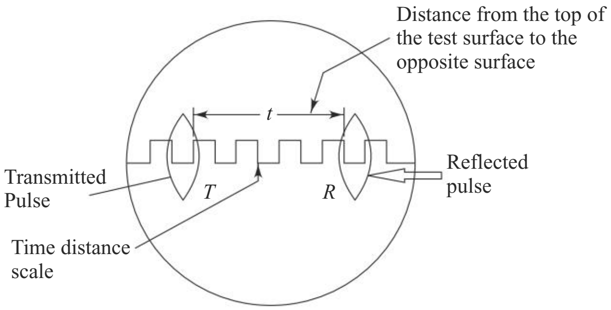


Fig. 10.7 *No defect shown on CRO Display*

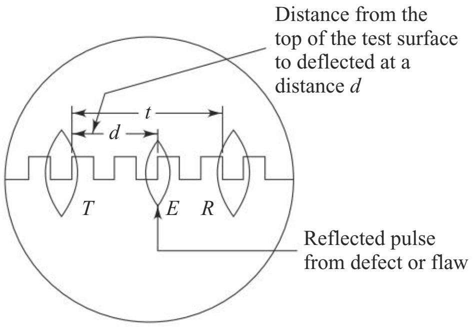


Fig. 10.8 *Defect indicated at a distance*

The reflected pulse from the flaw will appear as a smaller 'pip' between the other two, i.e., one from the top and the other from bottom. The distance to a flaw can be accurately measured as the distance between the pips on the oscilloscope screen represent elapsed/delay time of the reflected pulse.

Solved Problems

1 *Find the echo time of ultrasonic pulse traveling with a velocity of 6 km/sec in mild steel whose correct thickness is 18 mm.*

Solution Given, thickness $t = 18$ mm $= 18 \times 10^{-3}$ m, velocity $v = 6 \times 10^3$ m/sec

$$\therefore \quad \text{thickness} = \frac{vt}{2}$$

$$\therefore \quad \text{Echotime} = \frac{2d}{v} = \frac{2 \times 18 \times 10^{-3}}{6 \times 10^3} = 6 \times 10^{-6} \text{ s}$$

2 *Calculate the frequency to which a transducer oscillator circuit should be tuned so that a piezoelectric transducer of 0.1 cm thickness oscillates in its fundamental mode to produce ultrasonic waves. (Young's modulus of the transducer crystal = 80 Gpa, Density of material of crystal = 3200 kg/m^3)*

Solution Given, $t = 0.1 \times 10^{-2}$ m, $E = 80$ Gpa $= 80 \times 10^9$ pa

$$\rho = 3200 \text{ kg/m}^3\ f = ?$$

$$f = \frac{1}{2t}\sqrt{E/\rho}$$

$$= \frac{1}{2 \times 0.1 \times 10^{-2}}\sqrt{80 \times 10^9/3200}$$

$$= \frac{1}{0.002}\sqrt{10^9/40} = 2.88 \times 10^6 \text{ Hz}$$

3. *Velocity of ultrasonic waves in mild steel is 6 km/sec. The velocity of ultrasonic waves in brass measured by an ultrasonic gauge meter which was calibrated for mild steel was found to be 4 km/sec. If the thickness of the brass plate measured by the same meter is 15 cm what is its real thickness?*

Solution Given, $V_{ms} = 6 \times 10^3$ m/sec, Vbrass $= 4 \times 10^3$ m/sec, $t = 15 \times 10^{-2}$ m

We Know that $f\lambda = v$ and $f\,\mu$ 1/thickness

$$\therefore \quad v_1 t_1 = v_2 t_2$$

$$t_1 = \frac{v_2 t_2}{v_1} = \frac{4 \times 10^3 \times 15}{6 \times 10^3}$$

$$t_1 = 10 \text{ m}$$

Multiple-Choice Questions

1. Dye penetrant method is best suitable for
   (a) nonferrous metal (b) ferrous metal
   (c) For *a* and *b* both (d) None of the above

2. While removing excess penetrant
   (a) all the penetrant on the surface is removed.
   (b) Penetrant trapped in defect is removed.
   (c) *a* and *b* both are true
   (d) None of the above is true.

3. The objectives of NDT are
   (a) to prevent accidents and increase the safety
   (b) to improve productivity and hence to increase the profitability
   (c) a and b both are true
   (d) none of the above is true

4. Bore scope helps to inspect
   (a) an accurate three-dimensional image of a given object.
   (b) the inside of a narrow tube, bore, or chamber
   (c) For a and b both
   (d) none of the above

5. Flexiscope is used in
   (a) ultrasonic testing method
   (b) liquid penetrant testing method

(c) for a and b both

(d) visual Inspection system

6. Ionizing radiation of short wavelength is used in

(a) ultrasonic testing method

(b) liquid penetrant testing method

(c) xerography testing method

(d) visual inspection system

7. The following instrument is needed for ultrasonic testing

(a) voltmeter (b) thermometer

(c) CRO (d) none of the above

8. Couplets such as glycerin, jelly, thin oil, etc., between the test material and the ultrasonic transducer may be used

(a) to make surface smooth

(b) to ensure proper contact

(c) to make test cheaper

(d) none of the above

ANSWERS TO THE MCQ

1 (a) 2 (a) 3 (c) 4 (b) 5 (d) 6 (c) 7 (c) 8 (b)

Practice Problems

1. The frequency of the piezoelectric oscillator circuit is 2.7 MHz. Calculate the thickness of the crystal needed to tuned with the oscillator in its fundamental mode to generate ultrasonic waves. Given, density of the crystal is 2654 kg/m^3 and Young's modulus of the material of the crystal is 80 GPa. **[Ans: 0.1 cm]**

Short-answer Questions

1. What do you mean by NDT?

2. Enumerate methods of NDT.

3. What are the objectives of NDT?

4. What are the advantages of NDT?

5. Distinguish clearly between destructive testing and nondestructive testing.

6. Which type of defects can be detected /tested using nondestructive testing?

7. What are the demerits or limitations of NDT?

8. Explain the basic principle of Liquid Penetrant Testing (LPT) method.

9. Give the characteristics of liquid penetrant to be used in LPT method.

10. List out sequential steps of LPT method.

11. What care must be taken while following LPT? Why?

12. What is radiography?

13. Explain the principle of radiography.

14. What are the advantages of xeroradiograph?

15. Distinguish between radiography and fluoroscopy.

16. Explain the principle of ultrasonic testing method.

17. What are the basic components used in ultrasonic inspection system?

18. Which type of defects can be detected using visual inspection method?

19. State and explain the optical aids used in visual inspection method.

20. What are the limitations of human eye for visual inspection method?

Long-answer Questions

1. Discuss the objectives of NDT methods.

2. Discuss the general principles behind the various types of NDT methods.

3. Draw a schematic diagram of ultrasonic flaw detector and hence, explain its operation.

4. Write a short note on visual inspection method.

5. Giving sequence of steps involved explain LPT. What are its merits and demerits?

6. Explain the principle of X-ray radiography and describe the technique to detect the location of the flaws by X-rays.

7. Which NDT method should be employed to test the steel components? Explain it in detail.

8. Compare different NDT methods. Mention where they are best suitable.

Questions From Previous Exams

(i) Define NDT. **(GTU- Jan' 2009), (GTU- June' 2010)**

It is a testing of specimen to detect the presence of defects such as cracks, holes and flaws without causing any damage to the material

(ii) Discuss in detail the ultrasonic flaw detection. **(GTU- Jan' 2009)**

(Refer Section: 10.5.5)

(iii) Discuss the liquid penetration method of NDT in detail. **(GTU- Jan' 2009)**

(Refer Section: 10.5.2)

(iv) What are the objectives of NDT? Discuss about the ultrasonic inspection method-pulse echo system. **(GTU- Jan' 2010)**

(Refer Section: 10.5.5)

(v) Explain X-ray Radiography. **(GTU- Jan' 2010)**

(Refer Section: 10.5.3)

(vi) What is radiography? **(GTU- June' 2010)**

(Refer Section: 10.5.3)

(vii) Discuss the dye penetrant method of NDT in detail. **(GTU- June' 2010)**

(Refer Section: 10.5.2)

(viii) Draw a block diagram of an ultrasonic flaw detector and explain its actions. **(GTU- June' 2010)**

(Refer Section: 10.5.5)

Gujarat Technological University

B.E. Sem-I Remedial Examination December/January 2009

Q.1

(i) Define the unit cell

It is a smallest geometric figure, the repetition of which gives the actual crystal structure.

(ii) State the properties of LASER.

It has four properties namely, Coherence, High Intensity, High Directionality and High Monochromatic.

(iii) What is the life time of charge carrier in metastable state?

The life time of charge carrier in metastable state is 10^{-3} to 10^{-2} sec.

(iv) What is the standard Intensity? Give its value.

The minimum sound intensity which a human ear can sense is called the standard intensity. It value is 10^{-12}Wm^{-2}.

(v) Define the reverberation time.

The time taken by a sound intensity to fall below the minimum audibility level after the source stopped sounding.

(vi) Define piezoelectric effect.

When pressure is applied to one pair of opposite faces of crystals like quartz, tourmaline, Rochelle salt etc, cut with their faces perpendicular to its optic axis equal and opposite charges appear across its other faces.

(vii) Define total internal reflection.

The light approaches the wall of the fiber with an angle of incident Φ that is greater than the critical angle; the incident light does not refracted but reflects inside the denser medium.

(viii) State the main components of optical fiber communication system.

The important components of a fiber optic communication system are

Transmitter, Fiber optic cable and Receiver.

(ix) Define the transition temperature of super conductivity.

Temperature at which a material looses their electrical resistivity to absolute zero is called transition temperature.

(x) Define NDT.

It is a testing of specimen to detect the presence of defects such as cracks, holes and flaws without causing any damage to the material

(xi) What do you mean by metallic glasses?

Metallic glasses are metal alloys that are amorphous. They do not have a long range atomic order. They share the properties of both metals and glasses.

(xii) Give two examples of pentavalent impurities.

Phosphorus (P), Arsenic (As)

(xiii) Define lattice and basis.

A three dimensional collection of points in space is called crystalline lattice.

A unit assembly of atoms identical in composition, arrangement and orientation is called basis.

(xiv) State the full form of LASER

Light **A**mplification by **S**timulated **E**mission of **R**adiation.

Q.2

(a)

(i) Discuss the advantage of optical fibre communication system over conventional coaxial communication system. **(Refer Section 6.6)**

(ii) Compare type-I and type-II superconductor. **(Refer Section 8.5)**

(b)

(i) Derive the expression for acceptance angle and numerical aperture of an optical fibre. **(Refer Section 6.4)**

(ii) Calculate the NA, the acceptance angle of the fiber having $n_1 = 1.48$ and $n_2 = 1.43$.

Given, $n_1 = 1.48$ and $n_2 = 1.43$

$$NA = (n_1^2 - n_2^2)^{1/2}$$

$$NA = 0.3872$$

Acceptance Angle $\varphi_{\max} = \sin^{-1}$ NA,

$$\varphi_{\max} = \sin^{-1}(0.3872) = 22^\circ\ 46'$$

OR

(b)

(i) Discuss the properties, types and applications of metallic glasses. **(Refer Section 9.2)**

(ii) The volume of the room is 600 m^3. The wall area of the room is 220 m^2, the floor area is 120 m^2 and ceiling area is 120 m^2. The average sound absorption coefficient for wall is 0.03, for ceiling is 0.8 and for floor it is 0.06. Calculate reverberation time.

Given

$$V = 600\ \text{m}^3$$

Area of the room $S_1 = 220\ \text{m}^2$, $a_1 = 0.03$

Area of the floor $S_2 = 120\ \text{m}^2$, $a_2 = 0.06$

Area of the ceiling $S_3 = 120\ \text{m}^2$, $a_3 = 0.8$

Average absorption coefficient

$$\bar{a} = a_1S_1 + a_2S_2 + a_3S_3/S_1 + S_2 + S_3$$

6.6 + 7.2 + 96/460 = 0.2374 O.W.U.

$$\bar{a}\,\Sigma S = 0.2374 \times 460 = 109.8\ \text{O.W.U.-m}^2$$

$$T = 0.167\ V/a\,\Sigma S = 0.91\text{sec}$$

Q.3

(i) Describe the construction and working of Nd-YAG laser (Refer Section 5.9)

(ii) Derive the relation between Einstein's A and B coefficients. (Refer Section 5.2)

(iii) Calculate the frequency to which piezoelectric oscillator circuit should be tuned so that a piezoelectric crystal of thickness 0.1 cm vibrates in its fundamental mode to generate ultrasonic waves. (Young modulus and the density of the material of crystal are 80 GPa and 2654 Kgm^{-3})

Given $E = 80$ GPa $= 80 \times 10^9$ Pa, $\rho = 2654$ Kgm^{-3}, $t = 0.1$ cm $= 0.1 \times 10^{-2}$ m

The frequency of vibration is given by

$$f = 1/2t\,(E/\rho)^{1/2}$$

$$1/2 \times 0.1 \times 10^{-2}\,(80 \times 10^{9}/2654)$$

$$f = 2.7451 \text{ Hz}$$

OR

Q.3

(i) Explain the terms magnetostriction and piezoelectric effect. Discuss any one method of production of ultrasonic waves. (Refer Section 2.7)

(ii) What is meant by time of reverberation? Discuss the Sabine's formula. (Refer Section 1.8)

(iii) The Hall coefficient (R_H) of a semiconductor is 3.22×10^{-4} m^{3}C^{-1}. Its resitivity is 9×10^{-3} ohm-m. Calculate the mobility and carrier concentration of the carrier.

$$R_H = 3.22 \times 10^{-4} \text{ m}^3\text{C}^{-1}$$

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

$$\mu = ?$$

$$n = ?$$

$$R_H = 1/ne$$

$$n = 1/R_H e = 1/(3.22 \times 10^{-4})(1.6 \times 10^{-19}) = 1.94099 \times 10^{22}/\text{m}$$

$$\mu = \sigma R_H$$

$$\mu = R_H/\rho = 3.22 \times 10^{-4} \text{ m}^3\text{C}^{-1}/9 \times 10^{-3} \text{ ohm-m} = 0.35777\text{m}^2\text{V}^{-1}\text{s}^{-1}$$

Q.4

(i) Discuss in detail the ultrasonic flaw detection. (Refer Section 2.6)

(ii) What are Miller indices? Explain with proper example how to determine Miller Indices. (Refer Section 3.6)

(iii) Calculated the interplanar spacing for a (3, 1, 1,) plane in a simple cubic lattice whose lattice constant is 2.109×10^{-10} m.

Given Miller Indices $(h, k, l,) = (3, 1, 1,)$

Lattice Constant $a = 2.109 \times 10^{-10}$ m

Inter planner Distance $d = a/(h^2 + k^2 + l^2)^{1/2}$

$$d = 2.109 \times 10^{-10} \text{ m}/(11)^{1/2}$$

$$d = 2.109 \times 10^{-10} \text{ m}/3.31$$

$$d = 6.358 \times 10^{-11} \text{ m}$$

OR

(i) Explain how the materials are classified into conductors, semiconductors and insulators with the help of energy band diagram. (Refer Section 4.2)

(ii) State any five factors affecting the acoustics of the building and give at least two remedies. (Refer Section 1.9)

(iii) What is the resultant sound level when a 70 dB sound is added to a 80 dB sound?

Given $I_{L1} = 70 \text{ dB } I_{L1} = 80 \text{ dB}$

$$I_{L1} = 10 \log_{10} (I_1/I_o) = 70 \text{ dB}$$

$$I_{L2} = 10 \log_{10} (I_2/I_o) = 80 \text{ dB}$$

$$I_1/I_o = 10^7$$

$$I_2/I_o = 10^8$$

$$I = I_1 + I_2 = 10^7\, I_o + 10 \times 10^7\, I_o = 11 \times 10^7\, I_o$$

$$I_L = 10 \log_{10} (I/I_o)$$

$$I_L = 80.41 \text{ dB}$$

Q.5

(i) Explain the term Hall Effect. Derive the relation between Hall voltage and Hall coefficient. (Refer Section 4.7)

(ii) Discuss the important postulates of free electron theory of metals. (Refer Section 7.3)

(iii) Short notes (1) LED (2) Solar Cell (Refer Section 4.3)

OR

(i) Discuss the liquid penetration method of NDT in detail. (Refer Section 10.5)

(ii) Discuss the properties of superconductors. (Refer Section 9.3)

(iii) Short notes (1) crystal system (2) shape memory effect (Refer Section 3.4, 9.4)

Gujarat Technological University

B.E. Sem-I Remedial Examination January 2010

Q.1

(i) Explain the various factors affecting the acoustics of building and give their remedies.

Reverberation Time, Loudness, Focusing, Echelon Effect, Resonance

(ii) Briefly explain stimulated emission.

It is process in which there is emission of a photon whenever an atom transits from a higher energy state to lower energy state under the influence of an external agency i.e. inducing photon

(iii) What is the absorption coefficient? Explain how to determine the absorption coefficient.

The sound absorbing coefficient of a material is define as the ratio of sound energy absorbed by it to the total sound energy incident on it.

Suppose T_1 is reverberation time for empty hall, so $T_1 = 0.167\text{V}/\Sigma aS$

Suppose T_2 is reverberation time of hall with absorbing material,

so $T_2 = 0.167\ \text{V}//\Sigma aS + a_1S_1$

Therefore, $1/T_2 - 1/T_1 = a_1S_1/0.167\ \text{V}$ and hence $a_1 = 0.167\text{V}/S_1\ [1/T_2 - 1/T_1]$Sabine.

(iv) A cinema hall has a volume of 7500 m^3. What should be the total absorption in the hall if the reverberation time of 1.5 sec is to be maintained?

$T = 0.167\text{V}/\Sigma as$

$T = 1.5$ sec

$V = 7500$ m^3

$\Sigma as = ?$

$\Sigma as = 835$ Sabin-m^2

(v) Give the properties of an ultrasonic wave.

Frequency is greater than 20 KHz

It can be travel long distance

Speed of propagation depends upon frequency.

(vi) What is Kevlar?

It is a yarn type of material having high tensile strength. It gives an additional strength to cable.

Q.2

(a)

(i) Describe the principle and the method of production of ultrasonic waves by magnetostriction method. **(Refer Section 2.2)**

(ii) What is the resultant sound level when a 70 dB sound is added to a 85 dB sound?

Given $I_{L1} = 70$ dB $I_{L1} = 85$ dB

$I_{L1} = 10 \log_{10} (I_1/I_o) = 70$ dB

$$I_{L2} = 10 \log_{10} (I_2/I_o) = 85 \text{ dB}$$

$$I_1/I_o = 10^7$$

$$I_2/I_o = 10^{0.5}\, 10^8 = 3.16\, (10^8)$$

$$I = I_1 + I_2 = 10^7\, I_o + 3.16 \times 10^8\, I_o = [1 + (3.16)\, 10] \times 10^7\, I_o$$

$$I = (32.6)\, 10^7\, I_o$$

$$I_L = 10 \log_{10} (I/I_o)$$

$$I_L = 85.13 \text{ dB}$$

(iii) What are intrinsic semiconductors?

It is pure form without any impurities

Q.2

(b)

(i) Calculate the thickness of a quartz plate needed to produce ultrasonic wave of frequencies (i) 2 MHz (ii) 30 KHz. (Given ρ = 2650 Kg/m^3 and Young Modulus = 8 × 10^{10} N/m^2)

Given $f = 2$ MHz $= 2 \times 10^6$ Hz, $\rho = 2654$ Kgm^{-3}, Young Modulus $= 8 \times 10^{10}$ N/m^2, $t_1 = ?$

The frequency of vibration is given by

$$f = 1/2t\, (E/\rho)^{1/2}$$

Therefore,

$$t_1 = 1/2f\, (E/\rho)^{1/2}$$

$$t_1 = 1.37 \times 10^{-3} \text{ met}$$

Given $f = 30$ KHz $= 30 \times 10^3$ Hz, $\rho = 2654$ Kgm^{-3}, Young Modulus $= 8 \times 10^{10}$ N/m^2, $t_2 = ?$

The frequency of vibration is given by

$$f = 1/2t\, (E/\rho)^{1/2}$$

Therefore,

$$t_2 = 1/2f\, (E/\rho)^{1/2}$$

$$t_2 = 0.09157 \text{ met}$$

(ii) What are Miller Indices? Draw the plane from given Miller Indices; (i) (1 1 0), (-1 0 0) (iii) (1 1 2) **(Refer Section 3.6)**

OR

(b)

(i) What are the extrinsic semiconductors? Explain the term Hall Effect.

(Refer Section 4.2, 4.7)

(ii) Explain LED (principle and application) **(Refer Section 4.3)**

Q.3

(a)

(i) Explain (i) population inversion (ii) pumping (iii) optical resonator.

Population Inversion It is the state of achieving more number of atoms in excited state compared to the ground state i.e. $N_2 > N_1$.

Pumping It is a mechanism of exciting atoms from the lower energy state to higher energy state by supplying energy from an external source.

Optical resonator It is pairs of reflecting mirrors; of which one is being perfect reflector and other being partial reflector.. It is used for amplification of photons thereby producing an intense and highly coherent output.

(ii) Describe the construction and working of Nd: YAG Laser with a suitable energy level diagram. (Refer Section 5.9)

Q.3

(b) State the characteristics of LASER. Explain the method of construction and reconstruction of a hologram. (Refer Section 5.12, 5.14)

OR

(a) Describe the construction of fiber optic cable and compare the advantage of fiber optic cable over metallic cable. (Refer Section 6.2, 6.6)

(b)

(i) Give the conditions to be satisfied for total internal reflection for the optical fiber.

The refractive index n_1 of the core must always be greater than the refractive index n_2 of the cladding

The light must be approach the wall with an angle of incidence φ that is greater than the critical angle φ_C

(ii) The hall coefficient (R_H) of a semiconductor is 3.22 × 10^{-4} m^3C^{-1}. Its resitivity is 9 × 10^{-3} ohm-m. Calculate the mobility and carrier concentration of the carrier. (Given $e = 1.6 \times 10^{-19}$C)

$R_H = 3.22 \times 10^{-4}\ m^3C^{-1}.$

$\rho = 9 \times 10^{-3}$ ohm-m

$\mu = ?$

$n = ?$

$R_H = 1/ne$

$n = 1/R_H e = 1/(3.22 \times 10^{-4})(1.6 \times 10^{-19}) = 1.94099 \times 10^{22}/m$

$\mu = \rho R_H$

$\mu = R_H/\rho = 3.22 \times 10^{-4}\ m^3C^{-}1/9 \times 10^{-3}$ ohm-m $= 0.35777 m^2V^{-1}s^{-1}$

Q.4

(a)

(i) Explain the classical free electron theory of metal. (Refer Section 7.3)

(ii) An optical fiber fore and its cladding have refractive indexes of 1.545 and 1.495 respectively. Calculate the critical angle $\varphi_{C,}$ acceptance angle $\varphi_{\text{in(max)}}$ and Numerical aperture.

$n_1 = 1.545$

$n_2 = 1.495$

Critical angle $\varphi_{C,} = \sin^{-1}\ \varphi_C = n_2/n_1$

$\varphi_C = \sin^{-1}(0.967)$

$\varphi_C = 75°14'$

Acceptance angle $\varphi_{\text{in(max)}} = \sin^{-1} \text{NA} = \sin^{-1}(0.3898) = 22°56'$

Numerical aperture = $\text{NA} = (n_1^2 - n_2^2)1/2 = 0.3898$

(b)

(i) What are the success and drawback of classical free electron theory? (Refer Section 7.3)

(ii) Compare Type-I and Type-II superconductors. (Refer Section 8.5)

OR

(a) What is superconducting material? List the properties of superconducting materials and explain in detail.

Superconductivity is a phenomenon in which certain metal, alloys and ceramics conduct electricity without resistance when it is cooled below a certain temperature.

Properties are electrical resistance, Effect of impurities, Effect of pressure and Stress, Isotope Effect, Magnetic Field Effect, Meissner Effect , Persistence Current etc.

(b)

(i) Give the application of superconductor. (Refer Section 8.12)

(ii) Write the properties of metallic glasses. (Refer Section 9.2)

Q.5

(a) What is nanotechnology? Write the application of nanotechnology. (Refer Section 9.3)

(b)

(i) What is biomaterial? Write the type of biomaterials. (Refer Section 9.5)

(ii) Mention the name of the various NDT methods. (Refer Section 10.5)

OR

(a) What are the objectives of NDT? Discuss about the ultrasonic inspection method-pulse echo system. (Refer Section 10.2)

(b)

(i) Explain X-ray Radiography. (Refer Section 10.5)

(ii) Draw only the material phase transformation diagrams for shape memory effect (Temp Vs Load) and pseudo-elasticity (Stress Vs Strain) (Refer Section 9.4)

Gujarat Technological University

B.E. Sem-I Remedial Examination March/April 2010

Q.1

(a) Explain the method to determine the absorption coefficient of material.

Reverberation Time of the empty Hall is $T_1 = 0.167V/\sum aS$

Reverberation Time of the Hall with an absorbing material $T_2 = 0.167V/\sum aS + a_1S_1$

where the area of an absorbing material is S_1 and absorption coefficient a_1.

Hence, $\quad 1/T_1 - 1/T_2 = a_1S_1/0.167\ V$

(b) A liquid column subjected to ultrasonic waves constitutes an acoustical grating, explain.

The stationary wave acts as a diffraction grating for an ultrasonic wave in liquid column.

(c) Why Miller indices are reciprocals of the intercepts of the plane along the three axes?

The reciprocal of the intercept value become less that one which can be insert in cube to locate the position of the plane in cube.

(d) Show that the superconductors are perfectly diamagnetic in nature.

For the super conductor the susceptibility is negative and maximum

(e) Three levels laser system is rare whereas four level systems are common, explain.

Increases the number of metstable states which are helpful to increase population inversion.

(f) Why nano materials exhibit different physical properties of materials from those at a large scale?

Because nano means 10^{-9}meter = 1 nm and hence particle of an active element cab be fit into the structure of host material as a result the material can be developed as per requirement an application.

(g) State the requirements for infinite bandwidth optical fiber link.

As per the properties of an optical fiber, out put signal or signal carrying capacity is proportional to the bandwidth i.e large bandwidth for large carrying signal.

Q.2

(a) Distinguish between

(i) Spontaneous Emission and Stimulated Emission **(see Ref Sec:5.2)**

(ii) Monomode and Multimode optical fiber **(see Ref Sec:6.3)**

(b)

(i) Show that for a cubic lattice the relation between inter planar distance and Miller indices of a family of planes (*hkl*) is given by **(see Ref Sec:3.7)**

$$d = a/[h^2 + k^2 + l^2]^{1/2}$$

(ii) Show for simple cubic lattice that the ratio of separation between successive (100), (110) and (111) lattice planes is $d_{100} : d_{110} : d_{111} = a : a/2^{1/2} : a/3^{1/2}$

$$d = a/[h^2 + k^2 + l^2]^{1/2}$$

Hence,

$$d(100) = a/[1^2 + 0^2 + 0^2]^{1/2} = \mathrm{a}$$
$$d(110) = a/[1^2 + 1^2 + 0^2]^{1/2} = a/2^{1/2}$$
$$d(111) = a/[1^2 + 1^2 + 1^2]^{1/2} = a/3^{1/2}$$

OR

(b)

(i) Draw circuit diagram of oscillator (Pierce's) and explain magnetostriction method to produce ultrasonic waves. (See Ref Sec:2.2)

(ii) Calculate the length of an iron rod which can be used to produce ultrasonic waves of 20 KHz. Density of iron is 7.23 × 10^{3}kg/m^3 and Young's modulus is 11.6 × 10^{10} N/m^2.

Given $f = 20$ MHz $= 2 \times 10^7$ Hz, $\rho = 7.23 \times 10^3$ kg/m^3, Young Modulus $= 11.6 \times 10^{10}$ N/m^2.

$\mathrm{L} =?$

The frequency of vibration is given by

$$l = 1/2f\,(E/\rho)^{1/2}$$

Hence

$$l = 1/22 \times 10^7 \text{ Hz } (11.6 \times 10^{10} \text{ N/m}^2 / 7.23 \times 10^3 \text{ kg/m}^3)^{1/2}$$
$$l = 1.820 \times 10^{-5}\text{met}$$

Q.3

(a) Draw the diagram of variation of electron energy as a function of interatomic spacing; explain the formation of energy bands in a solid. Also explain how it helps to classify the materials into conductors, semiconductors and insulators? (See Ref Sec:4.1& 4.2)

(b) The Hall coefficient of a semiconductor is 3.22 × 10^{-4} m^{3}C^{-1} and its resistivity is 9 × 10^{-3} ohm-m. Calculate concentration and the mobility of the charge carrier.

Given:

$$e = 1.6 \times 10^{-19}\text{C.}$$
$$R_H = 3.22 \times 10^{-4} \text{ m}^3\text{C}^{-1}$$
$$\rho = 9\times10^{-3} \text{ ohm-m}$$
$$\mu =?$$
$$n =?$$
$$R_H = 1/ne$$
$$n = 1/R_H \quad e = 1/(3.22 \times 10^{-4})(1.6 \times 10^{-19}) = 1.94099 \times 10^{22}/\text{m}$$
$$\mu = \sigma R_H$$
$$\mu = R_{H/\rho} = 3.22 \times 10^{-4} \text{ m}^3\text{C } 1/9 \times 10^{-3} \text{ ohm-m} = 0.35777\text{m}^2\text{V}^{-1}\text{s}^{-1}$$

OR

Q.3

(a) Silicon and Germanium are not suitable for LED's, explain. (see Ref Sec:4.3.4)

(b) Discuss the dependence of critical magnetic field and London penetration depth on temperature. (See Ref Sec:8.7)

(c) Calculate the mean free path between collisions of the free electrons in copper at 20 C°. The resistivity of copper at 20 C° is 1.27 × 10^{-3} ohm-m and density of free electrons is 8.48 × 10^{28} m^{-3}.

Given:

$$e = 1.6 \times 10^{-19}\text{C}$$
$$m = 9.11 \times 10^{-31} \text{ kg and}$$
$$kB = 1.38 \times 10^{-23} \text{ JK}^{-1}$$
$$\rho = 1.27 \times 10^{-3} \text{ ohm-m}$$
$$T = 20 \text{ C}° = 293\text{K}$$
$$n = 8.48 \times 10^{28} \text{ m}^{-3}$$
$$\lambda = ?$$
$$\sigma = ne^2\, \tau/\text{m Hence}$$
$$\tau = 2.439 \times 10^{-14} \text{ sec}$$
$$\tau = \lambda/v$$
$$v = [3\, K_B\, \text{T/m}]^{1/2}$$
$$v = 115368.97 \text{ ms}^{-1}$$
$$\lambda = 2.8138 \times 10^{-9} \text{ m}$$

Q.4

(a) Show that the ratio of Einstein A coefficient for spontaneous emission to that of Einstein B coefficient for stimulated emission is given by $8\pi h v^3/C^3$. (See Ref Sec:5.2.1)

(b) Explain self focusing property of Graded Index optical fiber. (See Ref Sec:6.3.2)

(c) A refractive index of core for step index fiber is 1.52, diameter is 2.9 μm and a fractional difference of refractive index is 0.0007. It is operated at a wavelength of 1.3 μm. Find the number of modes the fiber will support.

$$V = 2\pi a/\lambda\, [n_1 - n_2]^{1/2}$$

where

Diameter or core $a = 2.9\ \mu\text{m}$

$$n_1 = 1.52$$
$$\Delta = 0.0007$$
$$n_2 = ?$$
$$\Delta = n_1 - n_2/n_1$$
$$n_1 =$$
$$\lambda = 1.3\ \mu\text{m}$$

OR

Q.4

(a) Draw the schematic diagram of hologram construction and explain the formation hologram. (see Ref Sec:6.8.1)

(b) Derive the expression of the numerical aperture of step index optical fiber. Show that it does not depend on the physical dimensions of the fiber. (see Ref Sec:6.4)

(c) Find the core radius, of step index fiber, necessary for single mode operation at 850 nm. The refractive indices of core and cladding are 1.48 and 1.47 respectively.

$$n_1 = 1.48$$
$$n_2 = 1.47$$

$\lambda = 850$ nm

$V = 2\pi a/\lambda\,[n_1 - n_2]^{1/2}$

For SI fiber, Single mode,

$1 = V^2/2$

V = less than 2.405

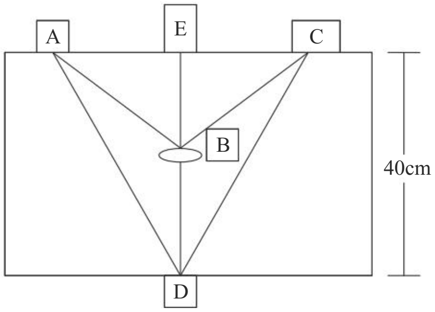
Q.5

(a) Discuss the melt spinning process technique for preparing metallic glasses. **(See Ref Sec:9.2)**

(b) Explain working of ultrasonic flow detector. **(See Ref Sec:2.6)**

(c) The ultrasonic pulse-echo is used to locate the position of defect in a steel bar of 40 cm thick. If pulse arrival times form defective and non defective portion are 30 μs and 80 μs respectively. Calculate the distance of defect from the top surface.

As per following diagram



Thickness of the bar = 40 cm

A distance $AD + DC$ travel by the pulse in time = 80 μs. (for non-defective surface)

A distance $AB + BC$ travel by the pulse in time = 30 μs. (for defective surface)

Distance of the defect from the top surface EB =?

Hence

For non-defective surface,

The velocity of pulse $V = [AD + DC]/t$

But, $V = 2ED/t$

Here, $ED = 40$ cm

$V = 2 \times 40 \text{ cm}/80\ \mu\text{s}$

$V = 1 \text{ cm}/\mu\text{s}$

Now

For defective surface,

The velocity of pulse $V = [AB + BC]/t$

But, $V = 2EB/30\ \mu\text{s}$

$EB = 15$ cm

OR

Q.5

(a) Explain Shape Memory effect and Pseudo Elastic effect. (See Ref Sec:9.4c, d)

(b) Why clean surfaces are required for surface and subsurface defect detection incase of liquid penetration NDT method? (See Ref Sec:10.5b)

(c) State and explain the factors to select a biomaterial for prosthetic device. (See Ref Sec:9.5)

Gujarat Technological University

B.E. Sem-II Remedial Examination June 2010

Q.1

(a) What is radiography?

Radiography is the use of X-rays to view a cross sectional area of a non uniformly composed material such as the human body.

By utilizing the physical properties of the ray an image can be developed displaying clearly, areas of different density and composition.

(b) Give four factors which affect acoustics of building.

Reverberation Time, Loudness, Focusing, Echelon Effect, Resonance

(c) Frequency range of audible sound wave is 20Hz < f <20 KHz

What is the absorption coefficient? Explain how to determine the absorption coefficient.

(d) What is piezoelectric method?

When pressure is applied to one pair of opposite faces of crystals like quartz, tourmaline, Rochelle salt etc, cut with their faces perpendicular to its optic axis equal and opposite charges appear across its other faces.

(e) The total number of atoms per unit cell in SC structure is 1.

(f) The atoms or molecules in a solid are arranged in some regular fashion known as Crystalline solid

(g) What is population inversion?

Population Inversion It is the state of achieving more number of atoms in excited state compared to the ground state i.e. $N_2 > N_1$.

(h) The active medium in Nd: YAG laser is Nd^{+3}

(i) Define fiber optic system

The important components of fiber optic communication systems are Transmitter, Fiber optic cable and Receiver.

(j) Define NDT.

It is a testing of specimen to detect the presence of defects such as cracks, holes and flaws without causing any damage to the material

(k) Give any two applications of biomaterials.

Artificial body parts, ceramic material.

(l) The relation between electrical conductivity of material and mobility of charge carries is given by $\sigma = n e \mu$

(m) A semiconductor behaves as a perfect insulator at 0K.

(n) $H_C = H_O(1 - T^2/Tc^2)$ is the formula which gives the relation of critical magnetic field and electrical temperature.

Q.2

(a)

(i) Explain the Hall Effect and derive an expression of Hall coefficient. (Refer Section:4.5)

(ii) Show that $d = a/(h^2 + k^2 + l^2)^{1/2}$. (Refer Section:3.7)

(b)

(i) An optical fiber has refractive index of core and classing is 1.514 and 1.48 respectively. Calculate the acceptance angle and the fractional index change.

Given, $n_1 = 1.514$, $n_2 = 1.48$

$\Delta = (n_1 - n_2)/n_1 = 0.022$

Acceptance angle $\varphi_{\text{in(max)}} = \sin^{-1} \text{NA} = \sin^{-1} (0.316) = 18° 25'$

Numerical aperture = $\text{NA} = (n_1^2 - n_2^2)^{1/2} = 0.316$

(ii) In carbon dioxide laser, the energy difference between two levels is 0.121eV. Calculate the wavelength of radiation.

Given 0.121 eV and hence 1 eV = 1.602×10^{-19}J

$$E_2 - E_1 = 0.121 \times 1.602 \times 10^{-19}\text{J}$$

$$h = 6.626 \times 10^{-34} \text{ Js}$$

$$c = 3 \times 10^{8}\text{m/s}$$

$$\lambda = hc/(E_2 - E_1)$$

$$\lambda = 1.0225 \times 10^{-5} \text{ met}$$

OR

(b) (i) Find the miller indices of a plane which intercepts at $a/2$, $b/2$ along X-axis and Y-axis respectively and parallel to Z-axis in a simple cubic unit cell. Draw a (0 1 1) plane in a cubic system. (Refer Section:3.6)

(ii) Calculate the frequency at which piezoelectric oscillator circuit should be tuned so that a piezoelectric crystal of thickness 0.1 cm vibrates in its fundamental mode to generate ultrasonic waves.

Given, Young Modulus $Y = 80\text{GPa}$, and density of crystal material = 2654 Kgm^{-3},

The frequency of vibration is given by

$$f = 1/2t\,(Y/\rho)^{1/2}$$

$$f = 2.74 \times 10^{6} \text{ Hz}$$

Q.3

(a) Explain the construction and working of CO_2 laser with a suitable energy level diagram (Refer Section:5.9.2)

(b) Established the relation between Einstein's coefficients. (Refer Section:5.2.1)

(c) Give the differences between type-I and type-II superconductor. (Refer Section:9.5)

OR

Q.3

(a) Derive an expression for electric conductivity. State and deduce Wiedeman-Franz law. (Refer Section: 7.5 & 7.7)

(b) Describe the construction of fiber optic cable. (Refer Section:6.3)

(c) Give the difference between step index fiber and graded index fiber. (Refer Section:6.3)

Q.4

(a) Mansion any five properties and application of nanomaterials. (Refer Section:9.3d&e)

(b) Define the term atomic radius and packing fraction. Calculate the above for SC, FCC and BCC structures. (Refer Section:3.9)

(c) Give the success and drawbacks of classical free electron theory. (Refer Section:7.8)

OR

Q.4

(a) Explain the types, properties of metallic glasses and melt spinning technique to prepare the metallic glasses. (Refer Section:9.2a,b,c)

(b) Discuss the dye penetrant method of NDT in detail. (Refer Section:10.5b)

(c) Draw a block diagram of an ultrasonic flaw detector and explain its actions. (Refer Section:2.6)

Q.5

(a) Explain Josephson Effect and its applications. (Refer Section:8.11)

(b) Derive the Sabine's formula for reverberation time. (Refer Section:1.8)

(c) What are the characteristics of musical sound? Explain them in details. (Refer Section:1.4)

OR

Q.5

(a) Write a short note on (i) LED and (ii) Photo diode. (Refer Section:4.3.4, 4.3.6)

(b) What is acoustic grating? Explain the acoustic grating method to determine the velocity of ultrasonic waves in liquid. (Refer Section: 2.5)

(c) Describe the various properties of ultrasonic. (Refer Section:10.4)

Frequency is greater than 20 KHz

It travels long distance

Speed of propagation depends upon frequency.

Questions Paper

Chapter-1

(i) What is the standard Intensity? Give its value. (GTU- Jan' 2009)

The minimum sound intensity which a human ear can sense is called the standard intensity. It value is 10^{-12}Wm^{-2}.

(ii) Define the reverberation time. (GTU- Jan' 2009)

The time taken by a sound intensity to fall below the minimum audibility level after the source stopped sounding.

(iii) The volume of the room is 600 m^3. The wall area of the room is 220 m^2, the floor area is 120 m^2 and ceiling area is 120 m^2. The average sound absorption coefficient for wall is 0.03, for ceiling is 0.8 and for floor it is 0.06. Calculate reverberation time. (GTU- Jan' 2009)

Given

$V = 600 \text{ m}^3$

Area of the room $S_1 = 220 \text{ m}^2, a_1 = 0.03$

Area of the floor $S_2 = 120 \text{ m}^2, a_2 = 0.06$

Area of the ceiling $S_3 = 120 \text{ m}^2, a_3 = 0.8$

Average absorption coefficient

$$\bar{a} = a_1S_1 + a_2S_2 + a_3S_3/S_1 + S_2 + S_3$$

$$6.6 + 7.2 + 96/460 = 0.2374 \text{ O.W.U.}$$

$$\bar{a}\,\Sigma S = 0.2374 \times 460 = 109.8 \text{ O.W.U.-m}^2$$

$$T = 0.167\text{ V}/\bar{a}\,\Sigma S = 0.91 \text{ sec}$$

(iv) What is meant by time of reverberation? Discuss the Sabine's formula. (GTU- Jan' 2009)

(v) State any five factors affecting the acoustics of the building and give at least two remedies. (Refer Section) (GTU- Jan' 2009)

(vi) What is the resultant sound level when a 70 dB sound is added to a 80 dB sound? (GTU- Jan' 2009)

Given $I_{L1} = 70 \text{ dB}\ I_{L1} = 80 \text{ dB}$

$$I_{L1} = 10 \log_{10} (I_1/I_o) = 70 \text{ dB}$$

$$I_{L2} = 10 \log_{10} (I_2/I_o) = 80 \text{ dB}$$

$$I_1/I_o = 10^7$$

$$I_2/I_o = 10^8$$

$$I = I_1 + I_2 = 10^7\, I_o + 10 \times 10^7\, I_o = 11 \times 10^7\, I_o$$

$$I_L = 10 \log_{10} (I/I_o)$$

$$I_L = 80.41 \text{ dB}$$

(vii) Explain the various factors affecting the acoustics of building and give their remedies. **(GTU- Jan' 2010)**

Reverberation Time, Loudness, Focusing, Echelon Effect, Resonance

(viii) What is the absorption coefficient? Explain how to determine the absorption coefficient. **(GTU- Jan' 2010)**

The sound absorbing coefficient of a material is define as the ratio of sound energy absorbed by it to the total sound energy incident on it.

Suppose T_1 is reverberation time for empty hall, so $T_1 = 0.167\ \text{V}/\Sigma aS$

Suppose T_2 is reverberation time of hall with absorbing material, so $T_2 = 0.167\ \text{V}//\Sigma aS + a_1S_1$

Therefore, $1/T_2 - 1/T_1 = a_1S_1/0.167\ \text{V}$ and hence $a_1 = 0.167\ \text{V}/S_1\ [1/T_2 - 1/T_1]$Sabine.

(ix) A cinema hall has a volume of 7500 m^3. What should be the total absorption in the hall if the reverberation time of 1.5 sec is to be maintained? **(GTU- Jan' 2010)**

$T = 0.167\ \text{V}/\Sigma as$

$T = 1.5$ sec

$V = 7500\ \text{m}^3$

$\Sigma as = ?$

$\Sigma as = 835$ Sabin-m^2

(x) What is the resultant sound level when a 70 dB sound is added to a 85 dB sound?

Given $I_{L1} = 70$ dB $I_{L1} = 85$ dB

$I_{L1} = 10 \log_{10} (I_1/I_o) = 70$ dB

$I_{L2} = 10 \log_{10} (I_2/I_o) = 85$ dB

$I_1/I_o = 10^7$

$I_2/I_o = 10^{0.5}10^8 = 3.16\ (10^8)$

$I = I_1 + I_2 = 10^7\ I_o + 3.16 \times 10^8$

$I_o = [1 + (3.16)10] \times 10^7\ I_o$

$I = (32.6)\ 10^7\ I_o$

$I_L = 10 \log_{10} (I/I_o)$

$I_L = 85.13$ dB

(x) Give four factors which affect acoustics of building.

Reverberation Time, Loudness, Focusing, Echelon Effect, Resonance

(xi) Derive the Sabine's formula for reverberation time. (Refer Section) **(GTU- June' 2010)**

(xii) What are the characteristics of musical sound? Explain them in details. (Refer Section) **(GTU- June' 2010)**

Chapter-2

(i) Define piezoelectric effect. **(GTU- Jan' 2009)**

When pressure is applied to one pair of opposite faces of crystals like quartz, tourmaline, Rochelle salt etc, cut with their faces perpendicular to its optic axis equal and opposite charges appear across its other faces.

(ii) Explain the terms magnetostriction and piezoelectric effect. Discuss any one method of production of ultrasonic waves. (Refer Section) (GTU- Jan' 2009)

(iii) Give the properties of an ultrasonic wave. (GTU- Jan' 2010)

Frequency is greater than 20 KHz

It can be travel long distance

Speed of propagation depends upon frequency.

(iv) Describe the principle and the method of production of ultrasonic waves by magnetostriction method. (Refer Section) (GTU- Jan' 2010)

(v) Calculate the thickness of a quartz plate needed to produce ultrasonic wave of frequencies (i) 2 MHz (ii) 30 KHz. (Given $\rho = 2650$ Kg/m^3 and Young Modulus = 8×10^{10} N/m^2) (GTU- Jan' 2010)

Given $f = 2$ MHz $= 2 \times 10^6$ Hz, $\rho = 2654$ Kgm^{-3}, Young Modulus = 8×10^{10} N/m^2

$$t_1 = ?$$

The frequency of vibration is given by

$$f = 1/2\,t\,(E/\rho)^{1/2}$$

Therefore,

$$t_1 = 1/2\,f\,(E/\rho)^{1/2}$$

$$t_1 = 1.37 \times 10^{-3} \text{ met}$$

Given $f = 30$ KHz $= 30 \times 10^3$ Hz, $\rho = 2654$ Kgm^{-3}, Young Modulus = 8×10^{10} N/m^2

$$t_2 = ?$$

The frequency of vibration is given by

$$f = 1/2\,t\,(E/\rho)^{1/2}$$

Therefore,

$$t_2 = 1/2\,f\,(E/\rho)^{1/2}$$

$$t_2 = 0.09157 \text{ met}$$

(vi) What is the absorption coefficient? Explain how to determine the absorption coefficient. (GTU- Jan' 2010)

(vii) Frequency range of audible sound wave is 20 Hz $< f <$20 KHz. (GTU- June' 2010)

(viii) What is piezoelectric method? (GTU- June' 2010)

When pressure is applied to one pair of opposite faces of crystals like quartz, tourmaline, Rochelle salt etc, cut with their faces perpendicular to its optic axis equal and opposite charges appear across its other faces.

(ix) Calculate the frequency at which piezoelectric oscillator circuit should be tuned so that a piezoelectric crystal of thickness 0.1 cm vibrates in its fundamental mode to generate ultrasonic waves. (GTU- June' 2010)

Given, Young Modulus $Y = 80$ GPa, and density of crystal material = 2654 Kgm^{-3},

The frequency of vibration is given by

$$f = 1/2\,t\,(Y/\rho)^{1/2}$$

$$f = 2.74 \times 10^6 \text{ Hz}$$

(x) Describe the principle and the method of production of ultrasonic waves by magnetostriction method. (Refer Section) (GTU- June' 2010)

(xi) What is acoustic grating? Explain the acoustic grating method to determine the velocity of ultrasonic waves in liquid. (Refer Section) (GTU- June' 2010)

(xii) Describe the various properties of ultrasonic. (GTU- June' 2010)

Frequency is greater than 20 KHz

It travels long distance

Speed of propagation depends upon frequency.

Chapter-3

(i) Define the Unit Cell (GTU- Jan' 2009)

It is a smallest geometric figure, the repetition of which gives the actual crystal structure.

(ii) Define lattice and basis. (GTU- Jan' 2009)

A three dimensional collection of points in space is called crystalline lattice.

A unit assembly of atoms identical in composition, arrangement and orientation is called basis.

(iii) Calculate the frequency to which piezoelectric oscillator circuit should be tuned so that a piezo-electric crystal of thickness 0.1cm vibrates in its fundamental mode to generate ultrasonic waves. (Young modulus and the density of the material of crystal are 80 GPa and 2654 Kgm^{-3}) (GTU- Jan' 2009)

Given $E = 80$ GPa $= 80 \times 10^9$ Pa, $\rho = 2654$ Kgm^{-3}, $t = 0.1$ cm $= 0.1 \times 10^{-2}$ m

The frequency of vibration is given by

$$f = 1/2\,t\,(E/\rho)^{1/2}$$

$$1/2 \times 0.1 \times 10^{-2}\,(80 \times 10^9/2654)$$

$$f = 2.7451 \text{ Hz}$$

(iv) What are Miller indices? Explain with proper example how to determine Miller Indices. . (Refer Section) (GTU- Jan' 2009)

(v) Calculated the inter planner spacing for a (3, 1, 1,) plane in a simple cubic lattice whose lattice constant is 2.109 × 10^{-10} m. (GTU- Jan' 2009)

Given Miller Indices $(h, k, l,) = (3, 1, 1,)$

Lattice Constant $a = 2.109 \times 10^{-10}$ m.

Inter planner Distance $d = a/(h^2 + k^2 + l^2)^{1/2}$

$$d = 2.109 \times 10^{-10} \text{ m}/(11)^{1/2}$$

$$d = 2.109 \times 10^{-10} \text{ m}/3.31$$

$$d = 6.358 \times 10^{-11} \text{ m}$$

(vi) Short notes (1) crystal system (GTU- Jan' 2009)

(vii) What are Miller Indices? Draw the plane from given Miller Indices; (GTU- Jan' 2010)

(viii) (1 1 0), (-1 0 0) (iii) (1 1 2) (Refer Section) (GTU- Jan' 2010)

(ix) The total number of atoms per unit cell in SC structure is 1

(x) The atoms or molecules in a solid are arranged in some regular fashion known as Crystalline solid (GTU- June' 2010)

(xi) Explain the Hall Effect and derive an expression of Hall coefficient.(Refer Section) **(GTU- June' 2010)**

(xii) Show that $d = a/(h^2 + k^2 + l^2)^{1/2.}$ (Refer Section) **(GTU- June' 2010)**

(xiii) (i) Find the miller indices of a plane which intercepts at $a/2$, $b/2$ along X-axis and Y-axis respectively and parallel to Z-axis in a simple cubic unit cell. Draw a (0 1 1) plane in a cubic system. (Refer Section) **(GTU- June' 2010)**

(xiv) Define the term atomic radius and packing fraction. Calculate the above for SC, FCC and BCC structures. (Refer Section) **(GTU- June' 2010)**

Chapter-4

(i) Give two examples of pentavalent impurities. **(GTU- Jan' 2009)**

Phosphorus (P), Arsenic (As)

(ii) The hall coefficient (R_H) of a semiconductor is 3.22 × 10^{-4} m^3C^{-1}. Its resitivity is 9 × 10^{-3} ohm-m. Calculate the mobility and carrier concentration of the carrier. **(GTU- Jan' 2009)**

$R_H = 3.22 \times 10^{-4}\ m^3C^{-1}.$

$\rho = 9 \times 10^{-3}$ ohm-m

$\mu = ?$

$n = ?$

$R_H = 1/ne$

$n = 1/R_H \quad e = 1/(3.22 \times 10^{-4})(1.6 \times 10^{-19}) = 1.94099 \times 10^{22}/m$

$\mu = \sigma R_H$

$\mu = R_H/\rho = 3.22 \times 10^{-4}\ m^3C^{-}1/9 \times 10^{-3}$ ohm-m $= 0.35777 m^2V^{-1}s^{-1}$

(iii) Explain how the materials are classified into conductors, semiconductors and insulators with the help of energy band diagram. (Refer Section) **(GTU- Jan' 2009)**

(iv) Explain the term Hall Effect. Derive the relation between Hall voltage and Hall coefficient. (Refer Section) **(GTU- Jan' 2009)**

(v) Short notes (1) LED (2) Solar Cell (Refer Section) **(GTU- Jan' 2009)**

(vi) What are intrinsic semiconductors? **(GTU- Jan' 2010)**

It is pure form without any impurities **(GTU- Jan' 2010)**

(vii) What are the extrinsic semiconductors? Explain the term Hall Effect. (Refer Section) **(GTU- Jan' 2010)**

(vii) Explain LED (principle and application) (Refer Section) **(GTU- Jan' 2010)**

(viii) The hall coefficient (R_H) of a semiconductor is 3.22 × 10^{-4} m^3C^{-1}. Its resitivity is 9 × 10^{-3} ohm-m. Calculate the mobility and carrier concentration of the carrier. (Given $e = 1.6 \times 10^{-19}C$) **(GTU- Jan' 2010)**

$R_H = 3.22 \times 10^{-4}\ m^3C^{-1}.$

$\rho = 9 \times 10^{-3}$ ohm-m

$\mu = ?$

$n = ?$

$R_H = 1/ne$

$n = 1/R_H \quad e = 1/(3.22 \times 10^{-4})(1.6 \times 10^{-19}) = 1.94099 \times 10^{22}$/m

$\mu = \rho\, R_H$

$\mu = R_H/\rho = 3.22 \times 10^{-4}\ m^3C^{-1}/9 \times 10^{-3}$ ohm-m $= 0.35777 m^2V^{-1}s^{-1}$

(ix) The relation between electrical conductivity of material and mobility of charge carries is give by $\sigma = n\, e\, \mu$ (GTU- June’ 2010)

(x) A semiconductor behaves as a perfect insulator at 0 K (GTU- June’ 2010)

(xi) Write a short note on (i) LED and (ii) Photo diode. (Refer Section) (GTU- June’ 2010)

Chapter-5

(i) State the properties of LASER. (GTU- Jan’ 2009)

It has four properties namely, Coherence, High Intensity, High Directionality and High Monochromatic.

(ii) What is the life time of charge carrier in metastable state? (GTU- Jan’ 2009)

The life time of charge carrier in metastable state is 10^{-3} to 10^{-2} sec.

(iii) State the full form of LASER (GTU- Jan’ 2009)

Light **A**mplification by **S**timulated **E**mission of **R**adiation.

(iv) Describe the construction and working of Nd-YAG laser (Refer Section) (GTU- Jan’ 2009)

(v) Derive the relation between Einstein’s A and B coefficients. (Refer Section) (GTU- Jan’ 2009)

(vi) Briefly explain stimulated emission. (GTU- Jan’ 2010)

It is process in which there is emission of a photon whenever an atom transits from a higher energy state to lower energy state under the influence of an external agency i.e. inducing photon

(vii) Explain (i) population inversion (ii) pumping (iii) optical resonator. (GTU- Jan’ 2010)

Population Inversion It is the state of achieving more number of atoms in excited state compared to the ground state i.e. $N_2 > N_{1.}$

Pumping It is a mechanism of exciting atoms from the lower energy state to higher energy state by supplying energy from an external source.

Optical resonator It is pairs of reflecting mirrors; of which one is being perfect reflector and other being partial reflector.. It is used for amplification of photons thereby producing an intense and highly coherent output.

(viii) Describe the construction and working of Nd: YAG Laser with a suitable energy level diagram. (Refer Section) (GTU- Jan’ 2010)

(viii) State the characteristics of LASER. Explain the method of construction and reconstruction of a hologram. (Refer Section) (GTU- Jan’ 2010)

(ix) What is population inversion? (GTU- June’ 2010)

Population Inversion It is the state of achieving more number of atoms in excited state compared to the ground state i.e. $N_2 > N_{1.}$

(x) The active medium in Nd: YAG laser is Nd^{+3} (GTU- June’ 2010)

(xi) In carbon dioxide laser, the energy difference between two levels is 0.121 eV. Calculate the wavelength of radiation. (GTU- June’ 2010)

Given 0.121 eV and hence 1 eV = 1.602×10^{-19} J

$$E_2 - E_1 = 0.121 \times 1.602 \times 10^{-19} \text{ J}$$
$$h = 6.626 \times 10^{-34} \text{ Js}$$
$$c = 3 \times 10^8 \text{ m/s}$$
$$\lambda = hc/(E_2 - E_1)$$
$$\lambda = 1.0225 \times 10^{-5} \text{ met}$$

(xii) Explain the construction and working of CO_2 laser with a suitable energy level diagram (Refer Section) (GTU- June' 2010)

(xiii) Established the relation between Einstein's coefficients. (Refer Section) (GTU- June' 2010)

Chapter-6

(i) Define total internal reflection. (GTU- Jan' 2009)

The light approaches the wall of the fiber with an angle of incident Φ that is greater than the critical angle; the incident light does not refracted but reflects inside the denser medium.

(ii) State the main components of optical fiber communication system. (GTU- Jan' 2009)

The important components of a fiber optic communication system are

Transmitter, Fiber optic cable and Receiver.

(iii) Discuss the advantage of optical fiber communication system over the conventional coaxial communication system. (Refer Section) (GTU- Jan' 2009)

(iv) Derive the expression for acceptance angle and Numerical aperture of an optical fiber.(Refer Section) (GTU- Jan' 2009)

(v) Calculate the NA, the acceptance angle of the fiber having $n_1 = 1.48$ and $n_2 = 1.43$. (GTU- Jan' 2009)

Given, $n_1 = 1.48$ and $n_2 = 1.43$

$$NA = (n_1^2 - n_2^2)^{1/2}$$
$$NA = 0.3872$$

Acceptance Angle $\varphi_{\max} = \sin^{-1} NA,$

$$\varphi_{\max} = \sin^{-1}(0.3872) = 22°46'$$

(vi) What is Kevlar? (GTU- Jan' 2010)

It is a yarn type of material having high tensile strength. It gives an additional strength to cable.

(vii) Describe the construction of fiber optic cable and compare the advantage of fiber optic cable over metallic cable. (Refer Section) (GTU- Jan' 2010)

(viii) Give the conditions to be satisfied for total internal reflection for the optical fiber. (GTU- Jan' 2010)

The refractive index n_1 of the core must always be greater than the refractive index n_2 of the cladding

The light must be approach the wall with an angle of incidence φ that is greater than the critical angle φ_C.

(ix) An optical fiber fore and its cladding have refractive indexes of 1.545 and 1.495 respectively. Calculate the critical angle φ_C, acceptance angle $\varphi_{in(max)}$ and Numerical aperture. **(GTU- Jan' 2010)**

Given

$$n_1 = 1.545$$
$$n_2 = 1.495$$

Critical angle $\varphi_C = \sin^{-1} \varphi_C = n_2/n_1$

$$\varphi_C = \sin^{-1}(0.967)$$
$$\varphi_C = 75^\circ 14'$$

Acceptance angle $\varphi_{in(max)} = \sin^{-1} \text{NA} = \sin^{-1}(0.3898) = 22^\circ 56'$

Numerical aperture = $NA = (n_1^2 - n_2^2)1/2 = 0.3898$

(x) Define fiber optic system **(GTU- June' 2010)**

The important components of fiber optic communication systems are Transmitter, Fiber optic cable and Receiver.

(xi) An optical fiber has refractive index of core and classing is 1.514 and 1.48 respectively. Calculate the acceptance angle and the fractional index change. **(GTU- June' 2010)**

Given, $n_1 = 1.514$, $n_2 = 1.48$

$$\Delta = (n_1 - n_2)/n_1 = 0.022$$

Acceptance angle $\varphi_{in(max)} = \sin^{-1} \text{NA} = \sin^{-1}(0.316) = 18^\circ 25'$

Numerical aperture = $NA = (n_1^2 - n_2^2)^{1/2} = 0.316$

(xii) Describe the construction of fiber optic cable. (Refer Section) **(GTU- June' 2010)**

(xiv) Give the difference between step index fiber and graded index fiber. (Refer Section) **(GTU- June' 2010)**

Chapter-7

(i) Define the transition temperature of super conductivity. **(GTU- Jan' 2009)**

Temperature at which a material looses their electrical resistivity to absolute zero is called transition temperature.

(ii) Compare type-I and type-II superconductor. (Refer Section) **(GTU- Jan' 2009)**

(iii) Discuss the properties of superconductors. (Refer Section) **(GTU- Jan' 2009)**

(iv) Compare Type-I and Type-II superconductors. (Refer Section) **(GTU- Jan' 2010)**

(v) What is superconducting material? List the properties of superconducting materials and explain in detail. **(GTU- Jan' 2010)**

Superconductivity is a phenomenon in which certain metal, alloys and ceramics conduct electricity without resistance when it is cooled below a certain temperature.

Properties are electrical resistance, Effect of impurities, Effect of pressure and Stress, Isotope Effect, Magnetic Field Effect, Meissner Effect, Persistence Current etc.

(vi) Give the application of superconductor. (Refer Section) **(GTU- Jan' 2010)**

(vii) $H_C = H_O (1 - T^2/Tc^2)$ **is the formula which gives the relation of critical magnetic field and electrical temperature.** **(GTU- June' 2010)**

(viii) Give the differences between type-I and type-II superconductor (Refer Section) **(GTU- June' 2010)**

(ix) Explain Josephson Effect and its applications. (Refer Section) **(GTU- June' 2010)**

Chapter-8

(i) Discuss the important postulates of free electron theory of metals (Refer Section) **(GTU- Jan' 2009)**

(ii) Explain the classical free electron theory of metal. (Refer Section) **(GTU- Jan' 2010)**

(iii) What are the success and drawback of classical free electron theory? (Refer Section) **(GTU- Jan' 2010)**

(ix) Derive an expression for electric conductivity. State and deduce Wiedeman-Franz law. **(GTU- June' 2010)**

(x) Give the success and drawbacks of classical free electron theory. (Refer Section) **(GTU- June' 2010)**

Chapter-9

(i) What do you mean by metallic glasses? **(GTU- Jan' 2009)**

Metallic glasses are metal alloys that are amorphous. They do not have a long range atomic order. They share the properties of both metals and glasses.

(ii) Discuss the properties, types and applications of metallic glasses. (Refer Section) **(GTU- Jan' 2009)**

(iii) Write the properties of metallic glasses. (Refer Section) **(GTU- Jan' 2010)**

(iv) What is nano technology? Write the application of nano technology. (Refer Section) **(GTU- Jan' 2010)**

(v) What is biomaterial? Write the type of biomaterials. (Refer Section) **(GTU- Jan' 2010)**

(vi) Give any two applications of biomaterials. **(GTU- June' 2010)**

Artificial body parts, ceramic material.

(vii) Mansion any five properties and application of nanomaterials. (Refer Section) **(GTU- June' 2010)**

(viii) Explain the types, properties of metallic glasses and melt spinning technique to prepare the metallic glasses. (Refer Section) **(GTU- June' 2010)**

Chapter-10

(i) Define NDT. **(GTU- Jan' 2009)**

It is a testing of specimen to detect the presence of defects such as cracks, holes and flaws without causing any damage to the material

Q.10 *S.Q.P.*

(ii) Discuss in detail the ultrasonic flaw detection. (Refer Section) **(GTU- Jan' 2009)**

(iii) Discuss the liquid penetration method of NDT in detail. (Refer Section) **(GTU- Jan' 2009)**

(iv) Short notes (1) Shape Memory Effect (Refer Section) **(GTU- Jan' 2009)**

(v) What are the objectives of NDT? Discuss about the ultrasonic inspection method-pulse echo system. .(Refer Section) **(GTU- Jan' 2010)**

(vi) Explain X-ray Radiography.(Refer Section) **(GTU- Jan' 2010)**

(vii) Draw only the material phase transformation diagrams for shape memory effect (Temp Vs Load0 and pseudo-elasticity (Stress Vs Strain) (Refer Section) **(GTU- Jan' 2010)**

(viii) What is radiography? **(GTU- June' 2010)**

(jx) Define NDT. **(GTU- June' 2010)**

It is a testing of specimen to detect the presence of defects such as cracks, holes and flaws without causing any damage to the material

(x) Discuss the dye penetrant method of NDT in detail. (Refer Section) **(GTU- June' 2010)**

(xi) Draw a block diagram of an ultrasonic flaw detector and explain its actions. (Refer Section) **(GTU- June' 2010)**