

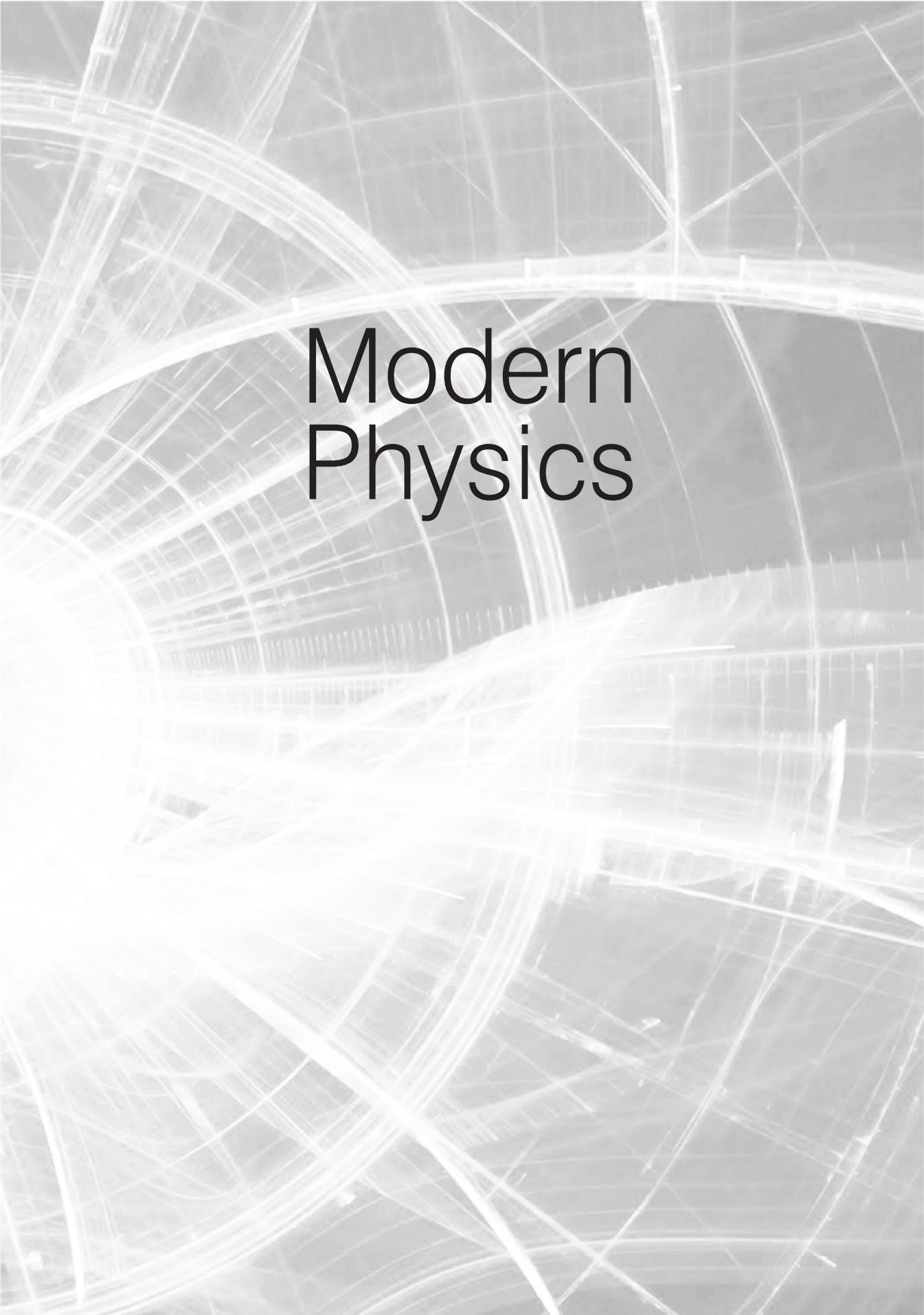
18e

# Modern Physics

As per UGC Choice Based Credit System (CBCS)

R MURUGESHAN  
KIRUTHIGA SIVAPRASATH

S. CHAND

The background of the image consists of a complex network of thin, white, curved lines forming a grid-like pattern. These lines are set against a dark gray, slightly textured background, creating a sense of depth and motion. The lines intersect at various points, forming a web-like structure that covers the entire frame.

# Modern Physics



# Modern Physics

**R MURUGESHAN**

M.Sc., M.Phil.

Head of the Department of Physics (Retired)

Vivekananda College

Tiruvedakam West (Post)

Madurai District-625 217

Tamil Nadu

**KIRUTHIGA SIVAPRASATH**

B.E.



**S Chand And Company Limited**

(ISO 9001 Certified Company)  
RAM NAGAR, NEW DELHI - 110 055



## S Chand And Company Limited

(ISO 9001 Certified Company)

Head Office: 7361, RAM NAGAR, QUTAB ROAD, NEW DELHI - 110 055

Phone: 23672080-81-82, 66672000 Fax: 91-11-23677446

[www.schandpublishing.com](http://www.schandpublishing.com); e-mail: [info@schandpublishing.com](mailto:info@schandpublishing.com)

### Branches:

Ahmedabad	: Ph: 27541965, 27542369, ahmedabad@schandpublishing.com
Bengaluru	: Ph: 22268048, 22354008, bangalore@schandpublishing.com
Bhopal	: Ph: 4209587, bhopal@schandpublishing.com
Chandigarh	: Ph: 2625356, 2625546, 4025418, chandigarh@schandpublishing.com
Chennai	: Ph: 28410027, 28410058, chennai@schandpublishing.com
Coimbatore	: Ph: 2323620, 4217136, coimbatore@schandpublishing.com (Marketing Office)
Cuttack	: Ph: 2332580, 2332581, cuttack@schandpublishing.com
Dehradun	: Ph: 2711101, 2710861, dehradun@schandpublishing.com
Guwahati	: Ph: 2738811, 2735640, guwahati@schandpublishing.com
Hyderabad	: Ph: 27550194, 27550195, hyderabad@schandpublishing.com
Jaipur	: Ph: 2219175, 2219176, jaipur@schandpublishing.com
Jalandhar	: Ph: 2401630, jalandhar@schandpublishing.com
Kochi	: Ph: 2809208, 2808207, cochin@schandpublishing.com
Kolkata	: Ph: 23353914, 23357458, kolkata@schandpublishing.com
Lucknow	: Ph: 4065646, lucknow@schandpublishing.com
Mumbai	: Ph: 22690881, 22610885, 22610886, mumbai@schandpublishing.com
Nagpur	: Ph: 2720523, 2777666, nagpur@schandpublishing.com
Patna	: Ph: 2300489, 2260011, patna@schandpublishing.com
Pune	: Ph: 64017298, pune@schandpublishing.com
Raipur	: Ph: 2443142, raipur@schandpublishing.com (Marketing Office)
Ranchi	: Ph: 2361178, ranchi@schandpublishing.com
Sahibabad	: Ph: 2771235, 2771238, delhibr-sahibabad@schandpublishing.com

© 1984, R Murugesan and Kiruthiga Sivaprasath

All rights reserved. No part of this publication may be reproduced or copied in any material form (including photocopying or storing it in any medium in form of graphics, electronic or mechanical means and whether or not transient or incidental to some other use of this publication) without written permission of the copyright owner. Any breach of this will entail legal action and prosecution without further notice.

**Jurisdiction:** All disputes with respect to this publication shall be subject to the jurisdiction of the Courts, Tribunals and Forums of New Delhi, India only.

First Edition 1984

Subsequent Editions and Reprints 1994, 95 (Twice), 96, 98, 99, 2001, 2002, 2003, 2005, 2007, 2009, 2010, 2012, 2013, 2014, 2016

Eighteenth Edition 2016

ISBN: 978-93-525-3310-7

# Preface to the Eighteenth Edition

---

This book is an extensive revision of the seventeenth edition with substantial amount of new material being added to reflect changes in technology as well as syllabi of various Indian universities. The primary objective of this book is to serve as the text in Modern Physics courses for B.Sc. students. This book may strike you as forbiddingly mathematical. Don't let the mathematics – which, for us, is only a tool – interfere with the physics.

The new edition has been divided into nine parts. Part I deals with *Theory of Relativity*. *Atomic Physics* is covered extensively in Part II which familiarizes the students with basics of structure of the atom, X-rays, photoelectric effect and Plank's quantum theory. Part III is devoted to *Quantum Mechanics*. Students have to realize that in physical measurements probability takes the place of exactness and as such phenomena which are impossible according to classical ideas may find a small but finite probability of occurrence. As Stephen Hawking states, 'God not only plays dice. He also sometimes throws the dice where they cannot be seen!' Part IV of the book covers *Molecular Physics* which lucidly explains molecular, atomic and x-ray spectra. Part V on *Nuclear Physics* presents a detailed introduction to its fundamental principles and applications. It elucidates the topics such as radiation detectors, particle accelerators, radioactivity, nuclear energy, cosmic rays, and elementary particles. An overall view of *Solid State Physics* is presented in Part VI. It deals with crystal structure, crystal binding, defects in crystals, lattice dynamics, superconductivity, dielectrics and magnetism. An introduction to astrophysics and nanotechnology, which are the two key areas in *Frontiers of Physics*, is presented in Part VII. We are living in the age of *Electronics*, thus Part VIII exposes the reader to *Semiconductor Devices*. Part IX on *Theoretical Physics* explains many of the mathematical techniques necessary for quantum mechanics.

The book is pedagogically enriched. Each chapter begins with the Learning Objectives which outlines the aim of each chapter. A large number of illustrations, diagrams and solved problems are provided throughout the book to help students achieve the practical understanding of the subject. The chapter-end review questions would help the students in comprehending and assessing their understanding of the underlying concepts.

We are thankful to the Management and the Editorial teams at S. Chand for all the help and support in publication of this book.

Suggestions and feedback from the readers will be gratefully acknowledged.

## AUTHORS

**Disclaimer :** While the authors of this book have made every effort to avoid any mistakes or omissions and have used their skill, expertise and knowledge to the best of their capacity to provide accurate and updated information, the authors and S. Chand do not give any representation or warranty with respect to the accuracy or completeness of the contents of this publication and are selling this publication on the condition and understanding that they shall not be made liable in any manner whatsoever. S.Chand and the authors expressly disclaim all and any liability/responsibility to any person, whether a purchaser or reader of this publication or not, in respect of anything and everything forming part of the contents of this publication. S. Chand shall not be responsible for any errors, omissions or damages arising out of the use of the information contained in this publication. Further, the appearance of the personal name, location, place and incidence, if any; in the illustrations used herein is purely coincidental and work of imagination. Thus the same should in no manner be termed as defamatory to any individual.



## Preface

---

The purpose of this book is to give an introductory account of the basic principles of atomic, nuclear and solid state physics. It is meant to serve as a textbook for the B.Sc. Physics students of Indian Universities. SI system of units has been used throughout the text. Great care has been taken in dealing with the subject with modern outlook. A large number of questions and problems have been given at the end of each chapter.

All available standard books on the subject have been freely consulted during the preparation of this book. I acknowledge gratefully my indebtedness to the authors and publishers of different books on the subject. I am grateful to my revered teachers Dr. S. Hariharan M.Sc., Ph.D. (Leeds), Prof. S.N. Neelakantan Nambisan B.Sc. (Hons.), Prof. P. Radhakrishna Menon M.A., Prof. P. Balakrishnan M.A., Prof. K. Padmanabha Sharma M.Sc., and others of Govt. Victoria College, Palghat, Kerala, who contributed to my education in Modern Physics.

I am extremely thankful to Prof. K. Ramamurthi M. Com., Secretary and Prof. K. Subrahmanyam M.A. Principal, of our College for their encouragement and help.

I wish to dedicate this book at the holy feet of Srimad Swami Chidbavanandaji, President of our college. To Swamiji, I owe more than words can describe for his lifelong devotion and guidance in unfolding my hidden potentialities and developing my personality. I can only hope that this book reflects something of his insistence on lucid and concise writing.

Lastly, I would like to thank my Publishers, M/s. S. Chand & Company Ltd.

I especially want to thank the Proprietor, Mr. S.L. Gupta and the Madras Branch Manager Mr. S. Venkatachalam.

Tiruvedakam

**R Murugesan**



# CONTENTS

## PART – I : RELATIVITY

### **1. Theory of Relativity** **2—26**

1.1 Introduction; 1.2 Frame of Reference; 1.3 Newtonian Principle of Relativity; 1.4 Galilean Transformation Equations; 1.5 The Ether Hypothesis; 1.6 The Michelson-Morley Experiment; 1.7 Special Theory of Relativity; 1.8 The Lorentz Transformation Equations; 1.9 Length Contraction; 1.10 Time Dilation; 1.11 Relativity of Simultaneity; 1.12 Addition of Velocities; 1.13 Variation of Mass with Velocity; 1.14 Mass Energy Equivalence; 1.15 Minkowski's Four Dimensional Space-Time Continuum; 1.16 The General Theory of Relativity; 1.17 Predictions of General Relativity; 1.18 Photons and Gravity; 1.19 Gravitational Red Shift.

## PART – II : ATOMIC PHYSICS

### **2. The Electron** **28—31**

2.1 Determination of the Electronic Charge : Millikan's oil-drop Method; 2.2 Dunnington's Method for Determining e/m.

### **3. Positive Rays** **32—42**

3.1 Discovery; 3.2 Properties of Positive Rays; 3.3 Positive Ray Analysis — Thomson's Parabola Method; 3.4 Aston's Mass Spectrograph; 3.5 Bainbridge's Mass Spectrograph; 3.6 Dempster's Mass Spectrograph; 3.7 Mass Defect and Packing Fraction; 3.8 Separation of Isotopes.

### **4. Structure of the Atom** **43—89**

4.1 Rutherford's Experiments on Scattering of  $\alpha$ -particles; 4.2 Theory of  $\alpha$ -particle Scattering; 4.3 Bohr Atom Model; 4.4 Effect of Nuclear Motion on Atomic Spectra; 4.5 Evidences in Favour of Bohr's Theory; 4.6 Bohr Correspondence Principle; 4.7 Critical Potentials; 4.8 Atomic Excitation; 4.9 Experimental Determination of Critical Potentials; 4.10 Drawbacks of Bohr Atom Model; 4.11 Sommerfeld's Relativistic Atom Model; 4.12 The Vector Atom Model; 4.13 Quantum Numbers Associated with the Vector Atom Model; 4.14 Coupling Schemes; 4.15 The Pauli Exclusion Principle; 4.16 The Periodic Classification of Elements; 4.17 Some Examples of Electron Configurations with their Modern Symbolic Representations; 4.18 Magnetic Dipole Moment Due to Orbital Motion of the Electron; 4.19 Magnetic Dipole Moment Due to Spin; 4.20 The Stern and Gerlach Experiment; 4.21 Optical Spectra; 4.22 Fine Structure of  $H_{\alpha}$  Line; 4.23 Zeeman Effect; 4.24 Larmor's Theorem; 4.25 Quantum Mechanical Explanation of the Normal Zeeman Effect; 4.26 Anomalous Zeeman Effect; 4.27 Paschen-Back Effect; 4.28 Stark Effect.

### **5. X-Rays** **90—107**

5.1 Production of X-Rays; 5.2 The Absorption of X-Rays; 5.3 Bragg's Law; 5.4 The Bragg X-Ray Spectrometer; 5.5 X-Ray Spectra; 5.6 Characteristic X-Ray Spectrum; 5.7 Auger Effect; 5.8 Satellites; 5.9 Moseley's Law; 5.10 Polarisation of X-Rays; 5.11 Scattering of

X-Rays (Thomson's Formula); 5.12 Determination of the Number of Electrons per Atom; 5.13 Compton Scattering.

## **6. Photoelectric Effect and Planck's Quantum Theory 108—120**

6.1 Introduction; 6.2 Lenard's Method to Determine e/m for Photoelectrons; 6.3 Richardson and Compton Experiment; 6.4 Experimental Investigations on the Photoelectric Effect; 6.5 Einstein's Photoelectric Equation; 6.6 Photoelectric Cells; 6.7 Planck's Quantum Theory.

# **PART – III : QUANTUM MECHANICS**

## **7. Wave Properties of Particles 122—147**

7.1 Inadequacy of Classical Mechanics; 7.2 Matter Waves; 7.3 Experimental Study of Matter Waves; 7.4 Electron Microscope; 7.5 Heisenberg's Uncertainty Principle.

## **8. Schrödinger Equation and its Applications 148—180**

8.1 Derivation of Time-dependent Form of Schrödinger Equation; 8.2 The Free Particle; 8.3 The Particle in a Box: Infinite Square Well Potential; 8.4 Finite Square Potential Well; 8.5 Rectangular Potential Well; 8.6 The Square Well in Three Dimensions; 8.7 Potential Step; 8.8 The Barrier Penetration Problem; 8.9 Linear Harmonic Oscillator; 8.10 Three Dimensional Harmonic Oscillator—(Spherically Symmetric Case); 8.11 The Rigid Rotator; 8.12 Derivation of Schrödinger's Wave Equation in Momentum Representation.

## **9. Quantum Theory of the Hydrogen Atom 181—196**

9.1 The Hydrogen Atom; 9.2 Electron Probability Density; 9.3 Radiative Transitions; 9.4 Selection Rules.

## **10. Operator Formalism of Quantum Mechanics 197—237**

10.1 Postulates of Quantum Mechanics; 10.2 Commuting Operators; 10.3 Wave Packet; 10.4 The Minimum Uncertainty Product; 10.5 Orthogonality of Eigenfunctions; 10.6 Dirac Delta Function; 10.7 System of Two Identical Particles; 10.8 Linear Harmonic Oscillator (Operator Method); 10.9 Hilbert Space; 10.10 Dirac's Bra and Ket Notation; 10.11 Total Angular Momentum; 10.12 Eigen Values of  $J^2$  and  $J_z$ .

## **11. Identical Particles and Spin 238—245**

11.1 Identical Particles; 11.2 Symmetric and Antisymmetric Wave Functions; 11.3 Construction of Symmetric and Antisymmetric Wave Functions from Unsymmetrised Functions; 11.4 Distinguishability of Identical Particles; 11.5 The Pauli Exclusion Principle; 11.6 Spin; 11.7 Pauli's Spin Matrices for Electron; 11.8 Commutation Relations Satisfied by Three Components  $\sigma_x$ ,  $\sigma_y$  and  $\sigma_z$ ; 11.9 Two Component Wave Functions.

## **12. Scattering Theory 246—258**

12.1 Introduction; 12.2 Scattering Experiment—Definitions of Cross-sections; 12.3 Reduction of the Two-Body Problem into One-Body Problem; 12.4 Quantum Scattering Theory; 12.5 Partial Wave Analysis.

**13. Perturbation Theory****259—263**

- 13.1 Time-Independent Perturbation Theory for a Non-Degenerate Energy Level;  
13.2 Application of Non-Degenerate Perturbation Theory Normal Helium Atom (Ground state of Helium).

**PART – IV : MOLECULAR PHYSICS****14. Molecular Spectra****266—307**

- 14.1 Introduction; 14.2 Theory of the Origin of Pure Rotational Spectrum of a Molecule; 14.3 Non-Rigid Rotator [Diatom Molecule as Non-Rigid Rotator]; 14.4 The Energy of a Diatomic Molecule; 14.5 Vibrating Diatomic Molecule as a Harmonic Oscillator; 14.6 The Anharmonic Oscillator; 14.7 Diatomic Vibrating Rotator—Theory of the Origin of the Vibration-Rotation Spectrum of a Molecule; 14.8 Vibration-Rotation Spectrum of Carbon Monoxide; 14.9 Infra-red Spectra; 14.10 Techniques and Instrumentation; 14.11 Double-beam Infrared Spectrometer; 14.12 Scattering of Light; 14.13 Experimental Study of Raman Effect; 14.14 Quantum Theory of Raman Effect; 14.15 Classical Theory of Raman Effect; 14.16 Vibrational Raman Spectra of Diatomic Molecules; 14.17 Applications of Raman Effect; 14.18 Techniques and Instrumentation; 14.19 Electronic Spectra of Diatomic Molecules.

**15. Spectra of Alkali and Alkaline Earth Elements****308—315**

- 15.1 Different Series in Alkali Spectra: Main Features; 15.2 Energy Levels in Alkali Spectra and Quantum Defect; 15.3 Nonpenetrating and Penetrating Orbitals; 15.4 Explanation of Salient Features of Alkali Spectra; 15.5 The Spectrum of Lithium and Other Hydrogen-like Species; 15.6 The Spectrum of Helium and the Alkaline Earths.

**16. X-Ray Spectra****316—320**

- 16.1 Fine Structure of X-Ray Levels; 16.2 Spin-relativity Doublets (Regular Doublets); 16.3 X-Ray Absorption Spectrum.

**PART – V : NUCLEAR PHYSICS****17. Introduction to the Nucleus****322—350**

- 17.1 Introduction; 17.2 Classification of Nuclei; 17.3 General Properties of Nucleus; 17.4 Binding Energy; 17.5 Nuclear Stability; 17.6 Theories of Nuclear Composition; 17.7 Nuclear Forces; 17.8 High Energy Electron Scattering Experiment; 17.9 Determination of Nuclear Radius—Mirror Nuclei Method; 17.10 The Liquid Drop Model; 17.11 Applications of Semi-Empirical Mass-Formula; 17.12 The Shell Model; 17.13 The Collective Model.

**18. Detectors of Nuclear Radiations****351—372**

- 18.1 Introduction; 18.2 Interaction between Energetic Particles and Matter; 18.3 Ionization Chamber; 18.4 Solid-state Detectors; 18.5 Proportional Counter; 18.6 Geiger-Muller Counter; 18.7 The Wilson Cloud Chamber; 18.8 Bubble Chamber; 18.9 Nuclear Emulsions; 18.10 Scintillation Counter; 18.11 Radiation Hazards; 18.12 Radiation Dosimeters.

**19. Particle Accelerators** **373—386**

19.1 Van de Graaff Generator; 19.2 The Linear Accelerator; 19.3 The Cyclotron; 19.4 Synchrocyclotron; 19.5 The Betatron; 19.6 The Synchrotrons; 19.7 The Proton Synchrotron (Bevatron, Cosmotron).

**20. Radioactivity** **387—418**

20.1 Introduction; 20.2 Determination of e/m of Alpha Particles; 20.3 Determination of the Charge of Alpha Particles; 20.4 Range of  $\alpha$ -particles; 20.5 Alpha Particle Disintegration Energy; 20.6 Alpha Particle Spectra; 20.7 Theory of Alpha Decay; 20.8 Gamow's Theory of Alpha Decay; 20.9 Determination of e/m of Beta Particles : Kaufmann's Experiment; 20.10 Beta Ray Spectrum; 20.11 Non-Conservation of Parity in Beta Decay; 20.12 Determination of the Wavelength of Gamma Rays; 20.13 Origin of  $\gamma$ -rays; 20.14 Nuclear Isomerism; 20.15 Internal Conversion; 20.16 Mossbauer Effect; 20.17 Soddy Fajan's Displacement Law; 20.18 Law of Radioactive Disintegration; 20.19 Units of Radioactivity; 20.20 Law of Successive Disintegration; 20.21 Radioactive Dating : The Age of the Earth.

**21. Artificial Transmutation of Elements** **419—441**

21.1 The Discovery of Artificial Transmutation; 21.2 The Q-Value Equation for a Nuclear Reaction (Nuclear Reaction Kinematics); 21.3 Nuclear Transmutations; 21.4 The Scattering Cross-Section; 21.5 Centre-of-Mass Coordinate System; 21.6 Discovery; 21.7 Preparation of Radioelements; 21.8 Applications of Radioisotopes; 21.9 The Discovery of the Neutron; 21.10 Basic Properties of the Neutron; 21.11 Neutron Sources; 21.12 Neutron Detection.

**22. Nuclear Fission and Fusion** **442—462**

22.1 Nuclear Fission; 22.2 Chain Reaction; 22.3 Nuclear Reactors; 22.4 Neutron Cycle in a Thermal Nuclear Reactor; 22.5 Transuranic Elements; 22.6 Nuclear Fusion; 22.7 Fusion Reactors.

**23. Cosmic Rays** **463—469**

23.1 Latitude Effect; 23.2 The East-West Effect (Azimuth Effect); 23.3 Altitude Effect; 23.4 Primary Cosmic Rays; 23.5 Secondary Cosmic Rays; 23.6 Cosmic Ray Showers; 23.7 Discovery of Positron; 23.8 The Mesons; 23.9 Van Allen Belts; 23.10 Origin of Cosmic Rays.

**24. Elementary Particles** **470—482**

24.1 Introduction; 24.2 Particles and Anti-particles; 24.3 Antimatter; 24.4 The Fundamental Interactions; 24.5 Elementary-particle Quantum Numbers; 24.6 Conservation Laws and Symmetry; 24.7 The Quark Model.

**PART – VI : SOLID STATE PHYSICS****25. Crystallography** **484—565**

25.1 Types of Solids; 25.2 Bravais Lattices; 25.3 Miller Indices; 25.4 Spacing between Three Dimensional Lattice Planes; 25.5 Elements of Symmetry; 25.6 Definitions; 25.7

Simple Cubic (SC) Structure; 25.8 Body Centred Cubic (BCC) Structure; 25.9 Face Centred Cubic (FCC) Structure; 25.10 Calculation of Lattice Constant; 25.11 Hexagonal Close Packed (HCP) Structure; 25.12 Diamond Cubic Structure; 25.13 Simple Crystal Structures; 25.14 Glass; 25.15 Quasicrystals; 25.16 Rotating Crystal Method; 25.17 Powder Method (Debye-Scherrer method); 25.18 X-Ray Diffraction: Laue Equations; 25.19 Reciprocal Lattice; 25.20 Brillouin Zones; 25.21 Reciprocal Lattice to Simple Cubic (SC) Lattice; 25.22 Reciprocal Lattice to BCC Lattice; 25.23 Reciprocal Lattice to FCC Lattice; 25.24 Atomic Scattering Factor; 25.25 The Geometrical Structure Factor; 25.26 Cohesive Energy of Inert Gas Crystals; 25.27 Ionic bond; 25.28 Covalent Bond; 25.29 Metallic Bond; 25.30 Hydrogen Bond; 25.31 Interatomic Potential; 25.32 Introduction; 25.33 Point Defects; 25.34 To Derive an Expression for the Number of Schottky Defects at a Given Temperature; 25.35 To Derive an Expression for the Number of Frenkel Defects at a Given Temperature; 25.36 To Derive an Expression for the Number of Vacancies at a Given Temperature; 25.37 Line Defects; 25.38 Surface Defects; 25.39 Effects of Crystal Imperfection.

## **26. Lattice Vibrations and Phonons** **566—586**

26.1 Introduction; 26.2 Vibrations of Monatomic Linear Lattice; 26.3 Vibrational Modes of a Finite One-dimensional Lattice of Identical Atoms; 26.4 Vibrational Modes of a Diatomic Linear Lattice; 26.5 Quantisation of Lattice Vibrations; 26.6 Inelastic Scattering of Neutrons by Phonons; 26.7 Dulong and Petit's law; 26.8 Einstein's Theory of Specific Heat Capacity; 26.9 Debye's Theory of Specific Heat Capacity of a Solid.

## **27. Free Electron Theory and Band Theory** **587—623**

27.1 Drude-Lorentz Theory of Electrical Conduction; 27.2 Expression for Electrical Conductivity (Derivation); 27.3 Expression for Thermal Conductivity; 27.4 Electrical Conductivity: Alternative Analysis; 27.5 Experimental Electrical Resistivity of Metals; 27.6 Sommerfeld's Free Electron Model of Metals; 27.7 Fermi-Energy; 27.8 Energy Levels in One Dimension; 27.9 Free Electron Gas in Three Dimensions; 27.10 Schottky Effect; 27.11 Paramagnetic Susceptibility of Conduction Electrons; 27.12 Hall Effect; 27.13 Band Theory of Solids; 27.14 Classification of Solids on the Basis of Band Theory; 27.15 Optical Properties of Solids; 27.16 Energy Bands: Alternative Analysis; 27.17 Bloch Theorem; 27.18 Kronig-Penney Model; 27.19 Effective Mass of Electron.

## **28. Superconductivity** **624—641**

28.1 Introduction; 28.2 Properties of Superconductors; 28.3 Classification of Superconductors; 28.4 BCS Theory of Superconductivity; 28.5 London Equation; 28.6 High-T<sub>c</sub> Superconductors (High-Temperature Superconductors); 28.7 Applications of Superconductors-Squid, Cryotron, Magnetic Levitation.

## **29. Dielectrics and Ferroelectrics** **642—662**

29.1 Introduction; 29.2 Different Types of Electric Polarisation; 29.3 Derivation of Clausius-Mossotti Equation; 29.4 Dielectric Loss; 29.5 Determination of Dielectric Constant of a Dielectric Material; 29.6 Dielectric Breakdown; 29.7 Properties and Applications of Different Types of Insulating Materials; 29.8 Introduction; 29.9 Properties of Ferroelectric Materials.

**30. Magnetism****663—694**

30.1 Introduction; 30.2 Langevin's Theory of Diamagnetism; 30.3 Properties of Paramagnetic Materials; 30.4 Langevin's Theory of Paramagnetism; 30.5 Weiss Theory of Paramagnetism; 30.6 Quantum Theory of Paramagnetism; 30.7 Cooling by Adiabatic Demagnetization of a Paramagnetic Salt; 30.8 Properties of Ferromagnetic Materials; 30.9 Domain Theory of Ferromagnetism; 30.10 Soft and Hard Magnetic Materials; 30.11 Weiss Theory of Ferromagnetism; 30.12 Nuclear Magnetic Resonance; 30.13 Heisenberg's Interpretation of Internal Field or Weiss Molecular Field; 30.14 Quantum Theory of Ferromagnetism; 30.15 Antiferromagnetic Materials; 30.16 Ferrites and their Applications.

**PART – VII : FRONTIERS OF PHYSICS****31. Astrophysics****696—709**

31.1 Introduction; 31.2 Classification of Stars — The Harvard Classification System; 31.3 Hertzsprung — Russel Diagram; 31.4 Luminosity of a Star; 31.5 Stellar Evolution; 31.6 White Dwarfs; 31.7 Chandrasekhar Limit; 31.8 Neutron Stars; 31.9 Black Holes; 31.10 Supernova Explosion; 31.11 Photon Diffusion Time; 31.12 Gravitational Potential Energy of a Star; 31.13 Internal Temperature of a Star; 31.14 Internal Pressure of a Star.

**32. Nanotechnology****710—733**

32.1 Introduction; 32.2 Nanoparticles; 32.3 Metal Nanoclusters; 32.4 Semiconducting Nanoparticles; 32.5 Rare Gas and Molecular Clusters; 32.6 Carbon Clusters; 32.7 Carbon Nanotubes; 32.8 Applications of Carbon Nanotubes; 32.9 Quantum Nanostructures; 32.10 Preparation of Quantum Nanostructures; 32.11 Size and Dimensionality Effects; 32.12 Applications of Quantum Nanostructures; 32.13 Photolithography; 32.14 Scanning Tunnelling Microscopy (STM); 32.15 Atomic Force Microscopy (AFM).

**PART – VIII : ELECTRONICS AND SOLID STATE DEVICES****33. Semiconductor Diode and Diode Applications****736—761**

33.1 Formation of  $p-n$  Junction Diode; 33.2 Forward and Reverse Biasing of a Junction Diode; 33.3 Derivation of Expressions for the Width of Depletion Layer and Height of Potential Barrier; 33.4 Derivation of Voltage Current (V-I) Equation for a  $p-n$  Junction Diode; 33.5  $p-n$  Junction Diode as Rectifier; 33.6 Filter Circuits; 33.7  $p-n$  Diode Clipping Circuits; 33.8  $p-n$  Diode Clamping Circuits.

**34. Special-Purpose Diodes****762—783**

34.1 The Zener Diode; 34.2 Optoelectronic Devices; 34.3 Photoconductors; 34.4 Photodiode; 34.5 Liquid Crystal Display (LCD); 34.6 Solar Cell; 34.7 Tunnel Diode; 34.8 Varactor Diode; 34.9 Shockley Diode; 34.10 PIN Diode.

**35. Bipolar Junction Transistor (BJT)****784—804**

35.1 Junction Transistor; 35.2 Common Emitter Characteristics of a Transistor; 35.3 Characteristics of a Transistor Connected in Common Base Configuration; 35.4 C – C

Configuration; 35.5 Transistor Circuit Configurations; 35.6 Transistor Biasing; 35.7 Thermal Runaway; 35.8 The D.C. Load Line; 35.9 A. C. Load Line; 35.10 Two-port Representation of a Transistor; 35.11 Photo Transistor.

## **36. BJT Amplifiers**

**805—835**

36.1 Common-Emitter Transistor Amplifier; 36.2 Common-Base Transistor Amplifier; 36.3 Hybrid Equivalent Circuit of Common Emitter Amplifier; 36.4 Power Amplifier; 36.5 Class B Push-Pull Amplifier; 36.6 Multistage Voltage Gain; 36.7 Direct-Coupled Multistage Amplifier; 36.8 Resistance-Capacitance (R-C) Coupled Amplifier; 36.9 Transformer Coupled (TC) Amplifier; 36.10 Distortion in Amplifiers; 36.11 Feedback Principle; 36.12 Emitter Follower.

## **37. Oscillators**

**836—863**

37.1 Introduction; 37.2 Hartley Oscillator; 37.3 Colpitt's Oscillator; 37.4 Wien-Bridge Oscillator; 37.5 Phase Shift Oscillator; 37.6 Crystal Oscillator; 37.7 Introduction; 37.8 Switching Action of a Transistor; 37.9 Astable Multivibrator; 37.10 Monostable Multivibrator; 37.11 Bistable Multivibrator or Flip-Flop Circuit; 37.12 Differentiating Circuit; 37.13 Integrating Circuit.

## **38. JFET, MOSFET, SCR and UJT**

**864—891**

38.1 Introduction; 38.2 Drain Characteristics of an n-Channel JFET; 38.3 Principle of n-Channel FET Operation; 38.4 Common source FET amplifier; 38.5 JFET Source-Follower (Common-Drain) Amplifier; 38.6 FET as Voltage-Variable Resistor (VVR); 38.7 The Depletion MOSFET; 38.8 The Enhancement MOSFET; 38.9 Cmos; 38.10 Silicon-Controlled Rectifier; 38.11 SCR Applications; 38.12 Unijunction Transistor (UJT); 38.13 Applications of UJT.

## **39. Operational Amplifiers**

**892—910**

39.1 Integrated Circuit (IC); 39.2 Introduction to Op Amps; 39.3 Inverting Amplifier; 39.4 Adder or Summing Amplifier; 39.5 Op-Amp as Integrator; 39.6 Op-Amp as Differentiator; 39.7 Difference Amplifier or Subtractor; 39.8 Logarithmic Amplifier; 39.9 Op-Amp as Comparator; 39.10 Non-inverting Operational Amplifier; 39.11 Voltage Follower (Unity Gain Buffer); 39.12 Introduction; 39.13 Fixed Voltage Regulators; 39.14 Adjustable Voltage Regulator.

## **40. Communications Electronics**

**911—944**

40.1 Introduction; 40.2 Amplitude Modulation; 40.3 Frequency Modulation; 40.4 Phase Modulation; 40.5 Block Diagram of AM Transmitter; 40.6 Demodulation or Detection (Linear Diode Detector); 40.7 Block Diagram of Superheterodyne AM Receiver; 40.8 An Integrated Circuit FM Transmitter; 40.9 Detection of FM Wave; 40.10 The Superheterodyne FM Receiver; 40.11 Television Camera — The Image Orthicon; 40.12 Colour Television; 40.13 Block Diagram of TV Transmitting System; 40.14 Block Diagram of TV Receiving System; 40.15 Constituents of Cathode Ray Oscilloscope; 40.16 Deflection Sensitivity of a CRO of Electrostatic Type; 40.17 Time Base; 40.18 Applications of the CRO; 40.19 Thevenin's Theorem; 40.20 Norton's Theorem.

**41. Number Systems and Logic Circuits****945—983**

41.1 Decimal Number System; 41.2 Binary Number System; 41.3 Octal Number System; 41.4 Hexadecimal Number System; 41.5 Binary Coded Decimal Code (BCD code); 41.6 ASCII CODE; 41.7 Data Selection Circuits; 41.8 Boolean Algebra; 41.9 De Morgan's Theorems; 41.10 The Basic Gates—NOT, OR, AND; 41.11 The NOR Gate; 41.12 The NAND Gate; 41.13 The Exclusive-OR Gate; 41.14 Half-Adder; 41.15 Full Adder; 41.16 Subtractor.

**PART – IX : THEORETICAL PHYSICS****42. Classical Mechanics****986—1014**

42.1 Mechanics of a Particle; 42.2 Mechanics of a System of Particles; 42.3 Basic Concepts; 42.4 Generalised Notations; 42.5 Derivation of Lagrange's Equations of Motion; 42.6 Applications of Lagrange's Equations; 42.7 Phase Space; 42.8 The Hamiltonian Function H; 42.9 Hamilton's Canonical Equations of Motion; 42.10 Physical Significance of the Hamiltonian Function; 42.11 Hamilton's Variational Principle; 42.12 Applications of Hamilton's Equations of Motion.

**43. Statistical Mechanics****1015—1037**

43.1 Introduction; 43.2 Phase Space; 43.3 Ensembles; 43.4 Liouville's Theorem; 43.5 Basic Concepts; 43.6 Maxwell-Boltzmann Distribution Law; 43.7 Bose-Einstein Distribution Law; 43.8 Fermi-Dirac Distribution Law; 43.9 Comparison of the Three Distribution Laws.

**44. Mathematical Physics****1038—1063**

44.1 Introduction; 44.2 Evaluation of Beta Function; 44.3 Evaluation of Gamma Function; 44.4 The Value of  $\Gamma\left(\frac{1}{2}\right)$  and Graph of the Gamma Function; 44.5 Other Forms of Gamma Function; 44.6 Relation Between Beta and Gamma Functions; 44.7 Bessel's Differential Equation; 44.8 Legendre's Differential Equation; 44.9 Hermite's Differential Equation; 44.10 Laguerre's Differential Equation; 44.11 Gauss' Divergence Theorem; 44.12 Stoke's Theorem; 44.13 Green's Theorem; 44.14 Special Types of Matrices; 44.15 Eigen Values, Eigen-Vectors: Characteristic Equation of a Matrix; 44.16 Cayley-Hamilton Theorem.

**Index****1064—1067**

**P A R T - I**

# **THEORY OF RELATIVITY**

## **1. Theory of Relativity**

# THEORY OF RELATIVITY

**After reading this chapter, you should be able to**

- ◆ Demonstrate an understanding of the basic principles of the special theory of relativity and general theory of relativity
- ◆ Discuss Michelson-Morley experiment and its significance
- ◆ List the consequences of Lorentz space-time transformation formulae.
- ◆ Explain why we don't notice length-contraction and time-dilation effects at everyday scales.
- ◆ Explain why the twin paradox is a false paradox.
- ◆ Explain when relativistic velocity addition should be used instead of classical addition of velocities
- ◆ Explain why the only mass it makes sense to talk about is rest mass.
- ◆ Illustrate the concept of mass-energy equivalence.
- ◆ Distinguish between Newtonian Mechanics and Relativistic Mechanics.

## 1.1 INTRODUCTION

Classical or Newtonian mechanics deals with the motions of bodies travelling at velocities that are very much less than the velocity of light. According to it, the three fundamental concepts of Physics, *viz.*, space, time and mass are all absolute and invariant.

**Concept of space.** Newton assumed that space is absolute and “exists in itself, without relation to anything external and remains unaffected under all circumstances”. This means to say that the length of an object is independent of the conditions under which it is measured, such as the motion of the object or the experimenter.

**Concept of time.** According to Newton, time is absolute ‘by its very nature flowing uniformly without reference to anything external’. Hence there is a universal time flowing at a constant rate, unaffected by the motion or position of objects and observers. This implies two things:

(1) The interval of time between two events has the same value for all observers, irrespective of their state of motion.

(2) If two events are simultaneous for an observer, they are simultaneous for all observers, irrespective of their state of motion, *i.e.*, *simultaneity is absolute*.

**Concept of mass.** In Newtonian mechanics,

(1) The mass of a body does not depend on the velocity of its motion.

(2) The mass of an isolated system of bodies does not change with any processes occurring within the system (*law of conservation of mass*).

## 1.2 FRAME OF REFERENCE

A system of co-ordinate axes which defines the position of a particle in two or three dimensional space is called a *frame of reference*. The simplest frame of reference is the familiar Cartesian system of co-ordinates, in which the position of the particle is specified by its three co-ordinates  $x, y, z$ , along the three perpendicular axes. In Fig. 1.1 we have indicated two observers  $O$  and  $O'$  and a particle  $P$ . These observers use frames of reference  $XYZ$  and  $X'Y'Z'$ , respectively. If  $O$  and  $O'$  are at rest, they will observe the same motion of  $P$ . But if  $O$  and  $O'$  are in relative motion, their observation of the motion of  $P$  will be different.

Unaccelerated reference frames in uniform motion of translation relative to one another are called *Galilean frames or inertial frames*.

Accelerated frames are called *non-inertial frames*.

### Definitions

#### (i) Inertial frame of reference

An *inertial frame of reference* is one in which Newton’s first law of motion holds. In such a frame, an object at rest remains at rest and an object in motion continues to move at constant velocity (constant speed and direction) if no force acts on it. Any frame of reference that moves at constant velocity relative to an inertial frame is itself an inertial frame.

*Special theory of relativity* deals with the problems that involve inertial frames of reference.

#### (ii) Non-inertial frame of reference

A *non-inertial frame of reference* is the one in which the Newton’s laws of motion are not valid *i.e.*, a body is accelerated when no external force acts on it.

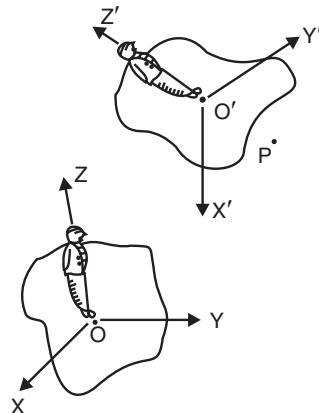


Fig. 1.1

### 1.3 NEWTONIAN PRINCIPLE OF RELATIVITY

**Statement.** *Absolute motion, which is the translation of a body from one absolute place to another absolute place, can never be detected. Translatory motion can be perceived only in the form of motion relative to other material bodies.*

**Explanation:** This implies that if we are drifting along at a uniform speed in a closed spaceship, all the phenomena observed and all the experiments performed inside the ship will appear to be the same as if the ship were not in motion. This means that the fundamental physical laws and principles are identical in all inertial frames of reference. This is the concept of Newtonian relativity.

### 1.4 GALILEAN TRANSFORMATION EQUATIONS

Let  $S$  and  $S'$  be two inertial frames (Fig. 1.2).

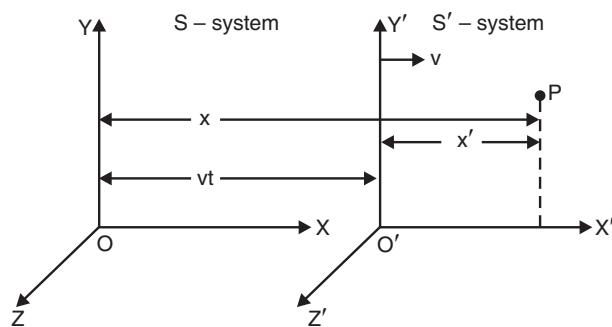


Fig. 1.2

Let  $S$  be at rest and  $S'$  move with uniform velocity  $v$  along the positive  $X$  direction. We assume that  $v \ll c$ . Let the origins of the two frames coincide at  $t = 0$ . Suppose some event occurs at the point  $P$ . The observer  $O$  in frame  $S$  determines the position of the event by the coordinates  $x, y, z$ . The observer  $O'$  in frame  $S'$  determines the position of the event by the coordinates  $x', y', z'$ . There is no relative motion between  $S$  and  $S'$  along the axes of  $Y$  and  $Z$ . Hence we have  $y = y'$  and  $z = z'$ . Let the time proceed at the same rate in both frames.

The distance moved by  $S'$  in the positive  $X$ -direction in time  $t = vt$ . So the  $X$  coordinates of the two frames differ by  $vt$ . Hence,  $x' = x - vt$ .

Then the transformation equations from  $S$  to  $S'$  are given by,

$$x' = x - vt \dots (1) \quad y' = y \dots (2) \quad z' = z \dots (3) \quad t' = t \dots (4)$$

#### Notes

(1) The inverse transformation equations (from  $S'$  to  $S$ ) are

$$x = x' + vt, y = y', z = z' \text{ and } t = t'.$$

(2) In general, the transformation of velocities from one to the other system is obtained by taking time derivatives. When the relative motion of the two frames is confined to the  $X$ -direction, the transformation becomes,

$$\frac{dx'}{dt} = \frac{dx}{dt} - v \quad i.e., \quad u' = u - v; \quad \frac{dy'}{dt} = \frac{dy}{dt}; \quad \frac{dz'}{dt} = \frac{dz}{dt}.$$

(3) Let  $a$  and  $a'$  be the accelerations of the particle in  $S$  and  $S'$ . We have  $a = \frac{du}{dt}$  and  $a' = \frac{du'}{dt}$ .

We have seen above that  $u' = u - v$ .

$$\therefore \frac{du'}{dt} = \frac{du}{dt} \quad (\text{since } v \text{ is constant})$$

or  $a' = a$ . i.e., the accelerations, as measured by the two observers in the two frames, are the same.  
Hence we say that acceleration is invariant under Galilean transformation.

## 1.5 THE ETHER HYPOTHESIS

A material medium is a necessity for the propagation of waves. It was considered that light propagates through *ether* as the sound waves propagate through air. Ether pervades all space. An interesting question is whether relative motion between the earth and ether can be detected. If such a motion can be detected, we can choose a fixed frame of reference in a stationary ether. Then we can express all motion relative to this frame. In 1887, Michelson and Morley, set out to measure the relative velocity of earth with respect to the ether. The principle of the experiment lies in noting the shift in fringes in the Michelson interferometer due to the difference in time taken by light to travel along and opposite the direction of motion of the earth. The time taken by a beam of light to travel along the direction of motion of earth is greater than that to travel distance opposite to the direction of motion of the earth. Surprisingly, despite best efforts, the presence of ether could not be detected.

## 1.6 THE MICHELSON-MORLEY EXPERIMENT

A beam of light from a monochromatic source  $S$  falls on a half-silvered glass plate  $P$ , placed at an angle of  $45^\circ$  to the beam. The incident beam is split up into two parts by  $P$  (Fig. 1.3). The reflected portion travels in a direction at right angles to the incident beam, falls normally at  $B$  on the plane mirror  $M_1$  and is reflected back to  $P$ . It gets refracted through  $P$  and enters the telescope  $T$ . The transmitted portion travels along the direction of the initial beam, falls normally on mirror  $M_2$  at  $A$  and is reflected back to  $P$ . After reflection from the back surface of  $P$ , it enters the telescope  $T$ . The two reflected beams interfere and the interference fringes are viewed with the help of the telescope  $T$ . The beam reflected upwards to  $M_1$  traverses the thickness of plate  $P$  thrice whereas the beam refracted on to mirror  $M_2$  traverses  $P$  only once. The effective distance of the mirrors  $M_1$  and  $M_2$  from the plate  $P$  is made to be the same by the use of a compensating plate not shown in figure.

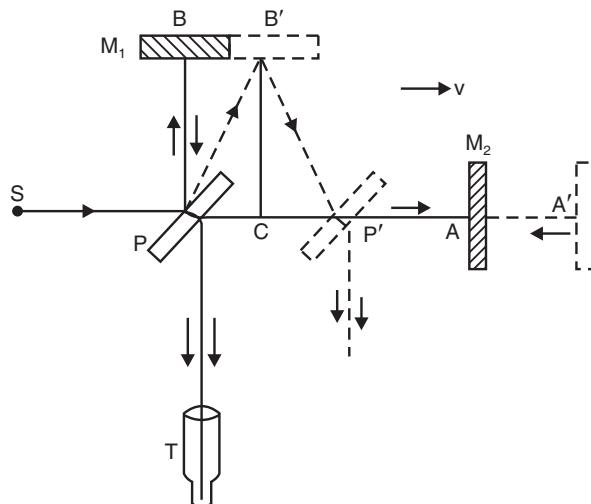


Fig. 1.3

The whole apparatus was floating on mercury. One arm ( $PA$ ) was pointed in the direction of earth's motion round the sun and the other ( $PB$ ) was pointed perpendicular to this motion. The paths of the two beams and the positions of their reflections from  $M_1$  and  $M_2$  will be as shown by the dotted lines.

Assume that the velocity of the apparatus (or earth) relative to fixed ether is  $v$  in the direction  $PA$ . The relative velocity of a light ray travelling along  $PA$  is  $(c - v)$  while its value would be  $(c + v)$  for the returning ray. Let  $PA = PB = d$ .

Time taken by light to travel from  $P$  to  $A = d/(c - v)$ .

Time taken by light to travel from  $A$  to  $P = d/(c + v)$ .

$\therefore$  Total time taken by light to travel from  $P$  to  $A$  and back

$$t = \frac{d}{c-v} + \frac{d}{c+v} = \frac{2cd}{c^2-v^2} = \frac{2d}{c} \left( 1 + \frac{v^2}{c^2} \right) \quad \dots(1)$$

Now, consider the ray moving upwards from  $P$  to  $B$ . It will strike the mirror  $M_1$  not at  $B$  but at  $B'$  due to the motion of the earth. If  $t_1$  is the time taken by the ray starting from  $P$  to reach  $M_1$ , then  $PB' = ct_1$  and  $BB' = vt_1$ .

The total path of the ray until it returns to  $P = PB' P'$ .

Now  $PB' P' = PB' + B' P' = 2PB'$ , since  $PB' = B' P'$ .

$$(PB')^2 = PC^2 + (CB')^2 = (BB')^2 + PB^2$$

$$\text{i.e., } c^2 t_1^2 = v^2 t_1^2 + d^2.$$

$$\therefore t_1 = \frac{d}{\sqrt{c^2 - v^2}}$$

$\therefore$  Total time taken by the ray to travel the whole path  $PB' P'$

$$t' = 2t_1 = \frac{2d}{\sqrt{c^2 - v^2}} = \frac{2d}{c\sqrt{1-(v^2/c^2)}} = \frac{2d}{c} \left( 1 + \frac{v^2}{2c^2} \right) \quad \dots(2)$$

Clearly,  $t' < t$ . The time difference

$$\Delta t = t - t' = \frac{2d}{c} \left( 1 + \frac{v^2}{c^2} \right) - \frac{2d}{c} \left( 1 + \frac{v^2}{2c^2} \right) = \frac{2d}{c} \times \frac{v^2}{2c^2} = \frac{dv^2}{c^3}.$$

The distance travelled by light in time  $\Delta t = c \times \Delta t = dv^2/c^2$ .

This is the path difference between the two parts of the incident beam. If the apparatus is turned through  $90^\circ$ , the path difference between the two beams becomes  $2dv^2/c^2$ . Michelson and Morley expected a fringe shift of about 0.4 in their apparatus when it was rotated through  $90^\circ$  and they believed that they could detect a shift as small as 0.01 of a fringe. But, in the experiment no displacement of the fringes was observed. They repeated the experiment at different points on the earth's surface and at different seasons of the year without detecting any measurable shift in fringes. *This negative result suggests that it is impossible to measure the speed of the earth relative to the ether.* Therefore, the effects of ether are undetectable. Thus, all attempts to make ether as a fixed frame of reference failed.

**Explanation of the negative result.** The negative result of the Michelson-Morley experiment can be explained by the following three explanations.

(1) The earth dragged along with it the ether in its immediate neighbourhood. Thus, there was no relative motion between the earth and ether. This is the explanation proposed by Michelson himself.

(2) Lorentz and Fitzgerald put forth the suggestion that there was contraction of bodies along the direction of their motion through the ether. Let  $L_0$  be the length of the body when at rest. If it is moving with a speed  $v$  parallel to its length, the new length  $L$  is  $L_0 \sqrt{1-(v^2/c^2)}$ . In the experiment

discussed above, the distance  $PB$  will remain unchanged. Distance  $PA$  will get shortened to  $d\sqrt{1-(v^2/c^2)}$ . If  $d$  were replaced by  $d\sqrt{1-(v^2/c^2)}$  in equation (1),  $t$  and  $t'$  will be the same and there will be no time difference expected. This contraction hypothesis easily explains why the Michelson-Morley experiment gave a negative result.

(3) The proper explanation for the negative result of the Michelson-Morley experiment was given by Einstein. He concluded that the velocity of light in space is a universal constant. This statement is called the *principle of constancy of the speed of light*. The speed of light is  $c$  rather than  $|(\mathbf{c} + \mathbf{v})|$  in any frame.

## 1.7 SPECIAL THEORY OF RELATIVITY

Einstein propounded the special theory of relativity in 1905. The special theory deals with the problems in which one frame of reference moves with a *constant linear velocity* relative to another frame of reference.

### Postulates of special theory of relativity

(1) *The laws of Physics are the same in all inertial frames of reference.*

(2) *The velocity of light in free space is constant. It is independent of the relative motion of the source and the observer.*

### Explanation :

(1) The first postulate expresses the fact that, since it is impossible to perform an experiment which measures motion relative to a stationary ether, no unique stationary frame of reference can be discovered. Since there is no favoured ‘rest’ frame of reference, all systems moving with constant velocity must be on equal footing. We cannot discuss absolute motion. We can discuss only relative motion.

(2) We know that the velocity of light is not constant under Galilean transformations. But according to the second postulate, the velocity of light is the same in all inertial frames. Thus the second postulate is very important and only this postulate is responsible to differentiate the classical theory and Einstein’s theory of relativity.

## 1.8 THE LORENTZ TRANSFORMATION EQUATIONS

We have to introduce new transformation equations which are consistent with the new concept of the invariance of light velocity in free space. The new transformation equations were discovered by Lorentz, and are known as “Lorentz transformations”.

**Derivation.** Consider two observers  $O$  and  $O'$  in two systems  $S$  and  $S'$ . System  $S'$  is moving with a constant velocity  $v$  relative to system  $S$  along the positive  $X$ -axis. Suppose we make measurements of time from the instant when the origins of  $S$  and  $S'$  just coincide, i.e.,  $t = 0$  when  $O$  and  $O'$  coincide. Suppose a light pulse is emitted when  $O$  and  $O'$  coincide. The

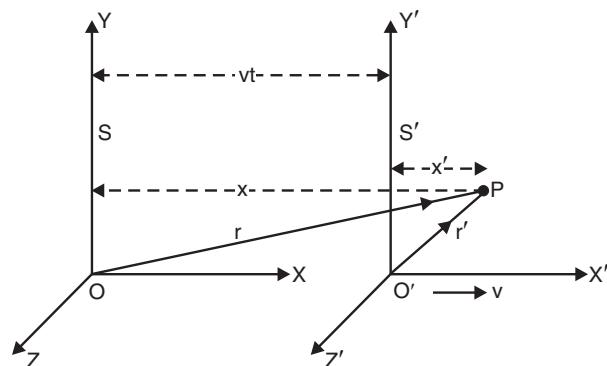


Fig. 1.4

light pulse produced at  $t = 0$  will spread out as a growing sphere. The radius of the wave-front produced in this way will grow with speed  $c$ . After a time  $t$ , the observer  $O$  will note that the light has reached a point  $P(x, y, z)$  as shown in Fig. 1.4. For him, the distance of the point  $P$  is given by  $r = ct$ . From figure,  $r^2 = x^2 + y^2 + z^2$ .

$$\text{Hence, } x^2 + y^2 + z^2 = c^2 t^2. \quad \dots(1)$$

Similarly, the observer  $O'$  will note that the light has reached the same point  $P$  in a time  $t'$  with the same velocity  $c$ . So  $r' = ct'$ .

$$\therefore x'^2 + y'^2 + z'^2 = c^2 t'^2 \quad \dots(2)$$

Now, equations (1) and (2) must be equal since both the observers are at the centre of the same expanding wavefront.

$$\therefore x^2 + y^2 + z^2 - c^2 t^2 = x'^2 + y'^2 + z'^2 - c^2 t'^2 \quad \dots(3)$$

Since there is no motion in the  $Y$  and  $Z$  directions,  $y' = y$  and  $z' = z$ .

$\therefore$  Equation (3) becomes,

$$x^2 - c^2 t^2 = x'^2 - c^2 t'^2 \quad \dots(4)$$

The transformation equation relating to  $x$  and  $x'$  can be written as,

$$x' = k(x - vt) \quad \dots(5)$$

Here,  $k$  is a constant.

The reason for trying the above relation is that, the transformation must reduce to Galilean transformation for low speed ( $v \ll c$ ).

Similarly, let us assume that

$$t' = a(t - bx) \quad \dots(6)$$

Here,  $a$  and  $b$  are constants.

Substituting these values for  $x'$  and  $t'$  in equation (4), we have,

$$x^2 - c^2 t^2 = k^2 (x - vt)^2 - c^2 a^2 (t - bx)^2$$

$$\text{i.e., } x^2 - c^2 t^2 = (k^2 - c^2 a^2 b^2) x^2 - 2(k^2 v - c^2 a^2 b) xt - \left(a^2 - \frac{k^2 v^2}{c^2}\right) c^2 t^2 \quad \dots(7)$$

Equating the coefficients of corresponding terms in equation (7),

$$k^2 - c^2 a^2 b^2 = 1 \quad \dots(8)$$

$$k^2 v - c^2 a^2 b = 0 \quad \dots(9)$$

$$a^2 - \frac{k^2 v^2}{c^2} = 1 \quad \dots(10)$$

Solving the above equations for  $k$ ,  $a$  and  $b$ , we get,

$$k = a = \frac{1}{\sqrt{1-(v^2/c^2)}} \quad \dots(11)$$

and

$$b = v/c^2 \quad \dots(12)$$

Substituting these values of  $k$ ,  $a$  and  $b$  in (5) and (6) we have,

$$x' = \frac{x - vt}{\sqrt{1-(v^2/c^2)}} \text{ and } t' = \frac{t - (vx/c^2)}{\sqrt{1-(v^2/c^2)}}$$

Therefore, the Lorentz transformation equations are

$$x' = \frac{x - vt}{\sqrt{1-(v^2/c^2)}}; y' = y; z' = z \text{ and } t' = \frac{t - (vx/c^2)}{\sqrt{1-(v^2/c^2)}} \quad \dots(13)$$

The inverse Lorentz transformation equations are obtained by interchanging the coordinates and replacing  $v$  by  $-v$  in the above.

$$x = \frac{x' + vt'}{\sqrt{1 - (v^2/c^2)}}, y = y'; z = z' \text{ and } t = \frac{t' + (vx'/c^2)}{\sqrt{1 - (v^2/c^2)}} \quad \dots(14)$$

These equations convert measurements made in frame  $S'$  into those in frame  $S$ .

**EXAMPLE 1.** Show that for values of  $v \ll c$ , Lorentz transformation reduces to the Galilean transformation.

**SOL.** When  $v \ll c$ ,  $\frac{v}{c} \rightarrow 0$ ;  $\therefore \frac{1}{\sqrt{1 - (v^2/c^2)}} \approx 1$ .

We have from (13),  $x' = x - vt$ ;  $y' = y$ ;  $z' = z$  and  $t' = t$  which are Galilean transformations.

**EXAMPLE 2.** Show that if  $(x_1, y_1, z_1, t_1)$  and  $(x_2, y_2, z_2, t_2)$  are the coordinates of one event in  $S_1$  and the corresponding event in  $S_2$  respectively, then the expression

$$ds_1^2 = dx_1^2 + dy_1^2 + dz_1^2 - c^2 dt_1^2$$

is invariant under a Lorentz transformation of coordinates.

**SOL.** The inverse Lorentz transformation equations are:

$$x_1 = \frac{x_2 + vt_2}{\sqrt{1 - \beta^2}}; y_1 = y_2; z_1 = z_2 \text{ and } t_1 = \frac{t_2 + vx_2/c^2}{\sqrt{1 - \beta^2}} \text{ (where } v/c = \beta).$$

Differentiating,

$$dx_1 = \frac{dx_2 + v dt_2}{\sqrt{1 - \beta^2}}; dy_1 = dy_2; dz_1 = dz_2 \text{ and } dt_1 = \frac{dt_2 + (\beta/c) dx_2}{\sqrt{1 - \beta^2}}$$

$$ds_1^2 = \left( \frac{dx_2 + v dt_2}{\sqrt{1 - \beta^2}} \right)^2 + dy_2^2 + dz_2^2 - c^2 \left( \frac{dt_2 + (\beta/c) dx_2}{\sqrt{1 - \beta^2}} \right)^2$$

This simplifies to

$$ds_1^2 = dx_2^2 + dy_2^2 + dz_2^2 - c^2 dt_2^2 = ds_2^2$$

## 1.9 LENGTH CONTRACTION

Consider two coordinate systems  $S$  and  $S'$  with their  $X$ -axes coinciding at time  $t = 0$ .  $S'$  is moving with a uniform relative speed  $v$  with respect to  $S$  in the positive  $X$ -direction. Imagine a rod ( $AB$ ), at rest relative to  $S'$  (Fig. 1.5).

Let  $x'_1$  and  $x'_2$  be the coordinates of the ends of the rod at any instant of time in  $S'$ . Then,

$$l_0 = x'_2 - x'_1 \quad \dots(1)$$

since the rod is at rest in frame  $S'$ .

Similarly, let  $x_1$  and  $x_2$  be the

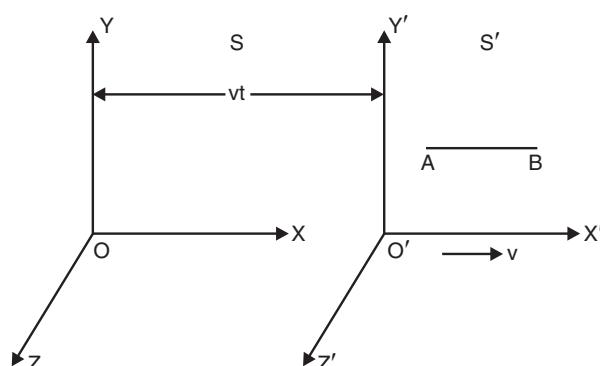


Fig. 1.5

coordinates of the ends of the rod at the same instant of time in  $S$ .

$$\text{Then } l = x_2 - x_1 \quad \dots(2)$$

$l$  is the length of the rod, measured relative to  $S$ .

According to Lorentz transformations,

$$x'_2 = \frac{x_2 - vt}{\sqrt{1-(v^2/c^2)}} \quad \dots(3)$$

and

$$x'_1 = \frac{x_1 - vt}{\sqrt{1-(v^2/c^2)}} \quad \dots(4)$$

Subtracting equation (4) from (3)

$$x'_2 - x'_1 = \frac{x_2 - x_1}{\sqrt{1-(v^2/c^2)}} \quad \text{or} \quad l_0 = \frac{l}{\sqrt{1-(v^2/c^2)}} \\ \therefore l = l_0 \sqrt{1-(v^2/c^2)} \quad \dots(5)$$

From equation (5) we see that  $l < l_0$ . Therefore, to the observer in  $S$  it would appear that the length of the rod (in  $S'$ ) has contracted by the factor  $\sqrt{1-(v^2/c^2)}$ .

For example, a body which appears to be spherical to an observer at rest relative to it, will appear to be an oblate spheroid to a moving observer. Similarly, a square and a circle in one appear to the observer in the other to be a rectangle and an ellipse respectively (Fig. 1.6).

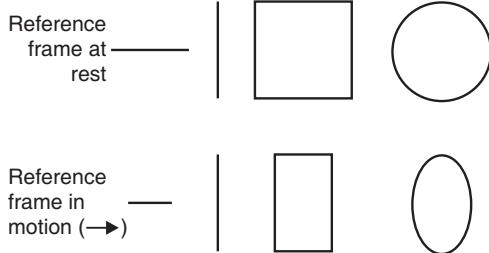


Fig. 1.6

### Notes

- (1) The *proper length* of an object is the length determined by an observer at rest with respect to the object. In the above case,  $l_0$  is the proper length.
- (2) The shortening or contraction in the length of an object along its direction of motion is known as the Lorentz-Fitzgerald contraction. There is no contraction in a direction perpendicular to the direction of motion.
- (3) The contraction becomes appreciable only when  $v \approx c$ .
- (4) The contraction is reciprocal, i.e., if two identical rods are at rest—one in  $S$  and the other in  $S'$ , each of the observers finds that the other is shorter than the rod of his own system.

**EXAMPLE 1.** A rod 1 metre long is moving along its length with a velocity  $0.6c$ . Calculate its length as it appears to an observer (a) on the earth (b) moving with the rod itself.

**SOL.** Here, 1 metre is the proper length ( $l_0$ ) of the rod in its own moving frame of reference. (a) Let  $l$  be the length of the rod as it appears to an observer in the stationary reference frame of the earth.

Here,

$$l_0 = 1 \text{ m}; v = 0.6 c; l = ?$$

$$l = l_0 \sqrt{1 - \frac{v^2}{c^2}} = 1 \sqrt{1 - \frac{(0.6c)^2}{c^2}} = 1 \sqrt{1 - 0.36} = 0.8 \text{ m}$$

Hence, the observer on the earth will estimate the length of the rod to be 0.8 metre.

(b) For an observer moving with the rod itself, the length of the rod is 1 metre.

**EXAMPLE 2.** How fast would a rocket have to go relative to an observer for its length to be contracted to 99% of its length at rest?

**SOL.** Here,

$$l = 0.99l_0; v = ?$$

We have,

$$l = l_0 \sqrt{1 - (v^2 / c^2)}$$

$$l^2 = l_0^2 \left(1 - \frac{v^2}{c^2}\right) \text{ or } \left(1 - \frac{v^2}{c^2}\right) = \frac{l^2}{l_0^2} \text{ or } \frac{v^2}{c^2} = 1 - \frac{l^2}{l_0^2}.$$

∴

$$v^2 = c^2 \left(1 - \frac{l^2}{l_0^2}\right) = c^2 (1 - 0.99^2)$$

∴

$$v = 0.1416 c = 0.1416 \times (3 \times 10^8) = 4.245 \times 10^7 \text{ ms}^{-1}$$

## 1.10 TIME DILATION

Imagine a gun placed at the position  $(x', y', z')$  in  $S'$  (Fig. 1.7).

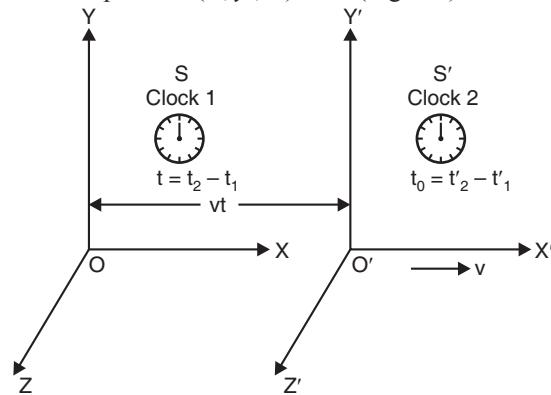


Fig. 1.7 A moving clock ticks more slowly than a clock at rest

Suppose it fires two shots at times  $t'_1$  and  $t'_2$  measured with respect to  $S'$ . In  $S'$  the clock is at rest relative to the observer. The time interval measured by a clock at rest relative to the observer is called the *proper time interval*. Hence,  $t_0 = t'_2 - t'_1$  is the time interval between the two shots for the observer in  $S'$ .

Since the gun is fixed in  $S'$ , it has a velocity  $v$  with respect to  $S$  in the direction of the positive  $X$ -axis. Let  $t = t_2 - t_1$  represent the time interval between the two shots as measured by an observer in  $S$ .

From inverse Lorentz transformations, we have

$$t_1 = \frac{t'_1 + vx'/c^2}{\sqrt{1 - (v^2/c^2)}} \quad \text{and} \quad t_2 = \frac{t'_2 + vx'/c^2}{\sqrt{1 - (v^2/c^2)}}$$

$$\therefore t_2 - t_1 = \frac{t'_2 - t'_1}{\sqrt{1 - (v^2/c^2)}} \quad \text{or} \quad t = \frac{t_0}{\sqrt{1 - (v^2/c^2)}}$$

$$\text{or} \quad t > t_0.$$

Thus, the time interval, between two events occurring at a given point in the moving frame  $S'$  appears to be longer to the observer in the stationary frame  $S$ ; i.e., a stationary clock measures a longer time interval between events occurring in a moving frame of reference than does a clock in the moving frame. This effect is called *time dilation*.

**The Twin Paradox.** Consider two exactly identical twin brothers. Let one of the twins go to a long space journey at a high speed in a rocket and the other stay behind on the earth. The clock in the moving rocket will appear to go slower than the clock on the surface of the earth (in accordance with  $t = \frac{t_0}{\sqrt{1-(v^2/c^2)}}$ ). Therefore, when he returns back to the earth, he will find himself younger than the twin who stayed behind on the earth!

**EXAMPLE 1.** A clock in a space ship emits signals at intervals of 1 second as observed by an astronaut in the space ship. If the space ship travels with a speed of  $3 \times 10^7 \text{ ms}^{-1}$ , what is the interval between successive signals as seen by an observer at the control centre on the ground?

**SOL.** Here,  $t_0 = 1\text{s}$ ;  $v = 3 \times 10^7 \text{ ms}^{-1}$ ; and  $c = 3 \times 10^8 \text{ ms}^{-1}$ ;  $t = ?$

$$\therefore t = \frac{t_0}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{1}{\sqrt{1-\frac{(3\times10^7)^2}{(3\times10^8)^2}}} = 1.005 \text{ s.}$$

**EXAMPLE 2.** A particle with a proper lifetime of  $1\mu\text{s}$  moves through the laboratory at  $2.7 \times 10^8 \text{ ms}^{-1}$ . (a) What is its lifetime, as measured by observers in the laboratory? (b) What will be the distance traversed by it before disintegrating?

**SOL.** Here,  $t_0 = 1\mu\text{s} = 10^{-6}\text{s}$ ;  $v = 2.7 \times 10^8 \text{ ms}^{-1}$ ;  $t = ?$

$$(a) t = \frac{t_0}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{10^{-6}}{\sqrt{1-\frac{(2.7\times10^8)^2}{(3\times10^8)^2}}} = 2.3 \times 10^{-6} \text{ s.}$$

(b) The average distance moved by the particle before disintegration =  $(2.7 \times 10^8) \times (2.3 \times 10^{-6})$   
 $= 620 \text{ m.}$

## 1.11 RELATIVITY OF SIMULTANEITY

Consider two events—the explosion of a pair of time bombs—that occur at the same time to an observer  $O$  in a reference frame  $S$ . Let the two events occur at different locations  $x_1$  and  $x_2$ . Consider another observer  $O'$  in  $S'$  moving with a uniform relative speed  $v$  with respect to  $S$  in the positive  $X$ -direction.

To  $O'$ , the explosion at  $x_1$  and  $t_0$  occurs at the time

$$t'_1 = \frac{t_0 - (v/c^2)x_1}{\sqrt{1-(v^2/c^2)}}$$

and the explosion at  $x_2$  and  $t_0$  occurs at the time

$$t'_2 = \frac{t_0 - (v/c^2)x_2}{\sqrt{1-(v^2/c^2)}}$$

$\therefore$  The time interval between the two events as observed by the observer  $O'$

$$= t' = t'_2 - t'_1 = \frac{(v/c^2)(x_1 - x_2)}{\sqrt{1-(v^2/c^2)}}$$

This is not zero. This indicates that two events at  $x_1$  and  $x_2$ , which are simultaneous to the observer in  $S$ , do not appear so to the observer in  $S'$ . Therefore, the concept of simultaneity has only a relative and not an absolute meaning.

## 1.12 ADDITION OF VELOCITIES

Suppose the system  $S'$  moves with a uniform velocity  $v$  relative to the system  $S$ . Suppose a particle is moving in the common direction of the  $x$  and  $x'$  axes. Let its velocity, as measured by an observer in the system  $S$ , be  $u$  and as measured by an observer in  $S'$  be  $u'$ .

$$\text{Then we have, } u = \frac{dx}{dt} \text{ and } u' = \frac{dx'}{dt}.$$

$$\text{Using inverse Lorentz transformations, we have, } x = \frac{x' + vt'}{\sqrt{1 - (v^2/c^2)}} \text{ and } t = \frac{t' + vx'/c^2}{\sqrt{1 - (v^2/c^2)}}.$$

$$\text{Differentiating, } dx = \frac{dx' + v dt'}{\sqrt{1 - (v^2/c^2)}} \text{ and } dt = \frac{dt' + v dx'/c^2}{\sqrt{1 - (v^2/c^2)}}$$

$$\text{Thus } \frac{dx}{dt} = \frac{dx' + v dt'}{dt' + \frac{v}{c^2} dx'} = \frac{\left(\frac{dx'}{dt'}\right) + v}{1 + \frac{v}{c^2} \left(\frac{dx'}{dt'}\right)}$$

$$\therefore u = \frac{u' + v}{1 + u' v / c^2}.$$

If we put  $u' = c$  and  $v = c$  i.e., photon is moving with a velocity  $c$  in frame  $S'$  and  $S'$  is moving with velocity  $c$  relative to  $S$ , then

$$u = \frac{c + c}{1 + (c^2/c^2)} = c.$$

*Thus, the addition of any velocity to the velocity of light  $c$  merely reproduces the velocity of light. Hence, the velocity of light is the maximum attainable velocity.*

### Notes

- (1) This law of addition of velocities applies only when the two velocities are in the same direction.
- (2) If  $v \ll c$ , we get the classical equation.
- (3) We can express the velocity  $u'$  in terms of  $u$  and  $v$ .

$$u' = \frac{u - v}{1 - \frac{uv}{c^2}}$$

**EXAMPLE 1.** An experimenter observes a radioactive atom moving with a velocity of  $0.25c$ . The atom then emits a  $\beta$  particle which has a velocity of  $0.9c$  relative to the atom in the direction of its motion. What is the velocity of the  $\beta$  particle, as observed by the experimenter?

**SOL.** Here,  $v = 0.25c$ ;  $u' = 0.9c$ ;  $u = ?$

$$u = \frac{u' + v}{1 + \frac{v}{c^2} u'} = \frac{0.9c + 0.25c}{1 + \frac{0.25c}{c^2} \times 0.9c} = 0.94c.$$

**EXAMPLE 2.** An electron is moving with a speed of  $0.85c$  in a direction opposite to that of a moving photon. Calculate the relative velocity of the photon with respect to the electron.

**SOL.** Let the photon be moving along the positive direction of  $X$ -axis and the electron along the negative direction of  $X$ -axis. Then, the speed of electron =  $-0.85c$ ; the speed of photon =  $c$ .

Consider that the electron is at rest in system  $S$ . Then, we may assume that the system  $S'$  (laboratory) is moving with velocity  $0.85c$  relative to system  $S$  (electron). i.e.,  $v = 0.85c$ ;  $u' = c$ ;  $u = ?$

$$u = \frac{u' + v}{1 + \frac{u'v}{c^2}} = \frac{c + 0.85c}{1 + \frac{(c)(0.85c)}{c^2}} = c.$$

### 1.13 VARIATION OF MASS WITH VELOCITY

**Derivation.** Consider two systems  $S$  and  $S'$ .  $S'$  is moving with a constant velocity  $v$  relative to the system  $S$ , in the positive  $X$ -direction (Fig. 1.8). Suppose that in the system  $S'$ , two exactly similar elastic balls  $A$  and  $B$  approach each other at equal speeds (i.e.,  $u$  and  $-u$ ). Let the mass of each ball be  $m$  in  $S'$ . They collide with each other and after collision coalesce into one body. According to the law of conservation of momentum,

Momentum of ball  $A$  + momentum of ball  $B$  = momentum of coalesced mass.

Or  $mu + (-mu) =$  momentum of coalesced mass = 0.

Thus the coalesced mass must be at rest in  $S'$  system.

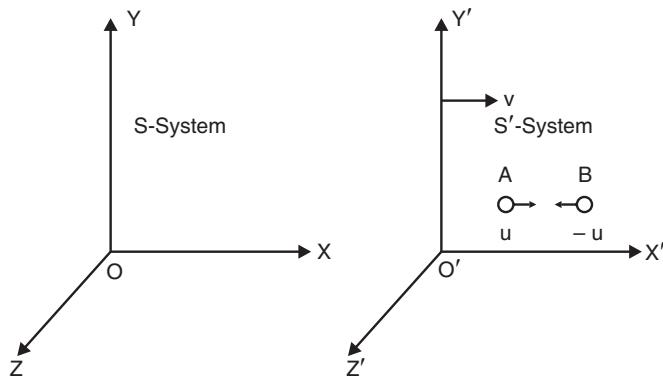


Fig. 1.8

Let us now consider the collision with reference to the system  $S$ .

Let  $u_1$  and  $u_2$  be the velocities of the balls relative to  $S$ . Then,

$$u_1 = \frac{u + v}{1 + uv/c^2} \quad \dots(1)$$

and

$$u_2 = \frac{-u + v}{1 - uv/c^2} \quad \dots(2)$$

After collision, velocity of the coalesced mass is  $v$  relative to the system  $S$ .

Let mass of the ball  $A$  travelling with velocity  $u_1$  be  $m_1$  and that of  $B$  with velocity  $u_2$  be  $m_2$  in the system  $S$ . Total momentum of the balls is conserved. Therefore,

$$m_1 u_1 + m_2 u_2 = (m_1 + m_2) v \quad \dots(3)$$

Substituting for  $u_1$  and  $u_2$  from equations (1) and (2), we have,

$$m_1 \left[ \frac{u + v}{1 + uv/c^2} \right] + m_2 \left[ \frac{-u + v}{1 - uv/c^2} \right] = (m_1 + m_2)v$$

or

$$m_1 \left[ \frac{u + v}{1 + \frac{uv}{c^2}} - v \right] = m_2 \left[ v - \frac{-u + v}{1 - \frac{uv}{c^2}} \right]$$

$$\text{or } m_1 \left[ \frac{u+v - uv^2/c^2}{1+uv/c^2} \right] = m_2 \left[ \frac{v-uv^2/c^2 + u-v}{1-uv/c^2} \right]$$

$$\text{or } m_1 \left[ \frac{u \left( 1 - \frac{v^2}{c^2} \right)}{1 + \frac{uv}{c^2}} \right] = m_2 \left[ \frac{u \left( 1 - \frac{v^2}{c^2} \right)}{1 - \frac{uv}{c^2}} \right]$$

$$\text{or } \frac{m_1}{m_2} = \frac{1+uv/c^2}{1-uv/c^2} \quad \dots(4)$$

Also,

$$\begin{aligned} 1 - \frac{u_1^2}{c^2} &= 1 - \frac{\left\{ \frac{u+v}{c} \right\}^2}{\left( 1 + \frac{uv}{c^2} \right)^2} \\ &= \frac{1 + \frac{u^2 v^2}{c^4} + \frac{2uv}{c^2} - \frac{u^2}{c^2} - \frac{v^2}{c^2} - \frac{2uv}{c^2}}{\left( 1 + \frac{uv}{c^2} \right)^2} \\ &= \frac{\left( 1 - \frac{u^2}{c^2} \right) - \frac{v^2}{c^2} \left( 1 - \frac{u^2}{c^2} \right)}{\left( 1 + \frac{uv}{c^2} \right)^2} \\ 1 - \frac{u_1^2}{c^2} &= \frac{\left( 1 - \frac{u^2}{c^2} \right) \left( 1 - \frac{v^2}{c^2} \right)}{\left( 1 + \frac{uv}{c^2} \right)^2} \end{aligned} \quad \dots(5)$$

Similarly,

$$1 - \frac{u_2^2}{c^2} = \frac{\left( 1 - \frac{u^2}{c^2} \right) \left( 1 - \frac{v^2}{c^2} \right)}{\left( 1 - \frac{uv}{c^2} \right)^2} \quad \dots(6)$$

Dividing equation (6) by equation (5),

$$\begin{aligned} \frac{1 - \frac{u_2^2}{c^2}}{1 - \frac{u_1^2}{c^2}} &= \frac{\left( 1 + \frac{uv}{c^2} \right)^2}{\left( 1 - \frac{uv}{c^2} \right)^2} \\ \text{or } \frac{\sqrt{1 - \frac{u_2^2}{c^2}}}{\sqrt{1 - \frac{u_1^2}{c^2}}} &= \frac{1 + \frac{uv}{c^2}}{1 - \frac{uv}{c^2}} \end{aligned} \quad \dots(7)$$

From equations (7) and (4),  $\frac{m_1}{m_2} = \frac{\sqrt{1-\frac{u_2^2}{c^2}}}{\sqrt{1-\frac{u_1^2}{c^2}}}$

or  $m_1 \sqrt{1-\frac{u_1^2}{c^2}} = m_2 \sqrt{1-\frac{u_2^2}{c^2}}$  ... (8)

Since the L.H.S. and R.H.S. of equation (8) are independent of one another, the above result can be true only if each is a constant. Therefore,

$$m_1 \sqrt{1-\frac{u_1^2}{c^2}} = m_2 \sqrt{1-\frac{u_2^2}{c^2}} = m_0.$$

The constant denoted by  $m_0$  is called the *rest mass* of the body and corresponds to *zero velocity*.

Thus,  $m_1 = \frac{m_0}{\sqrt{1-u_1^2/c^2}}$

In general, if  $m$  denotes the mass of a body when it is moving with a velocity  $v$ , then,

$$m = \frac{m_0}{\sqrt{1-v^2/c^2}} \quad \dots(9)$$

This is the relativistic formula for the variation of mass with velocity.

If we put  $v \rightarrow c$  in equation (9), we have  $m \rightarrow \infty$  i.e., an object travelling at the speed of light would have infinite mass. Thus, Eqn. (9) shows that no material body can have a velocity equal to, or greater than the velocity of light.

### Notes

(1) The first verification of the increase in mass with velocity came from the work of Kaufmann in 1906 and of Bucherer in 1909. While studying the  $\beta$ -rays emanating from radioactive materials, they found that their velocities were comparable to the velocity of light and also that their masses were found to be related to their velocities.

(2) The increase of mass with velocity has now been tested in "particle accelerators". Accelerators are used to accelerate the various charged particles of matter to very high velocities. It has been found that the electrons and protons accelerated in these machines to velocities close to the velocity of light acquire increased masses, exactly as predicted.

**EXAMPLE.** At what speed is a particle moving if the mass is equal to three times its rest mass?

**SOL.** We have,  $m = \frac{m_0}{\sqrt{1-v^2/c^2}}$ . Here,  $m = 3m_0$ ;  $v = ?$

$$\therefore 3m_0 = \frac{m_0}{\sqrt{1-v^2/c^2}} \text{ or } \sqrt{1-\frac{v^2}{c^2}} = \frac{1}{3} \text{ or } 1-\frac{v^2}{c^2} = \frac{1}{9}$$

$$\text{Or } \frac{v^2}{c^2} = \frac{8}{9} \text{ or } v^2 = \frac{8}{9}c^2 \text{ or } v = \sqrt{\frac{8}{9}}c = 0.94c.$$

## 1.14 MASS ENERGY EQUIVALENCE

**Derivation.** Force is defined as rate of change of momentum i.e.,

$$F = \frac{d}{dt}(mv) \quad \dots(1)$$

According to the theory of relativity, both mass and velocity are variable. Therefore,

$$F = \frac{d}{dt}(mv) = m \frac{dv}{dt} + v \frac{dm}{dt} \quad \dots(2)$$

Let the force  $F$  displace the body through a distance  $dx$ .

Then, the increase in the kinetic energy ( $dE_k$ ) of the body is equal to the work done ( $F dx$ ).

Hence,  $dE_k = F dx = m \frac{dv}{dt} dx + v \frac{dm}{dt} dx$

or  $dE_k = mv dv + v^2 dm \quad \dots(3)$

According to the law of variation of mass with velocity

$$m = \frac{m_0}{\sqrt{(1-v^2/c^2)}} \quad \dots(4)$$

Squaring both sides,  $m^2 = \frac{m_0^2}{1-(v^2/c^2)}$

or  $m^2 c^2 = m_0^2 c^2 + m^2 v^2$

Differentiating,  $c^2 2m dm = m^2 2v dv + v^2 2m dm$

or  $c^2 dm = mv dv + v^2 dm \quad \dots(5)$

From equations (3) and (5),  $dE_k = c^2 dm \quad \dots(6)$

Thus, a change in K.E.  $dE_k$  is directly proportional to a change in mass  $dm$ .

When a body is at rest, its velocity is zero, (K.E. = 0) and  $m = m_0$ . When its velocity is  $v$ , its mass becomes  $m$ . Therefore, integrating equation (6),

$$\begin{aligned} E_k &= \int_{m_0}^m dE_k = c^2 \int_{m_0}^m dm = c^2 (m - m_0) \\ \therefore E_k &= mc^2 - m_0 c^2 \quad \dots(7) \end{aligned}$$

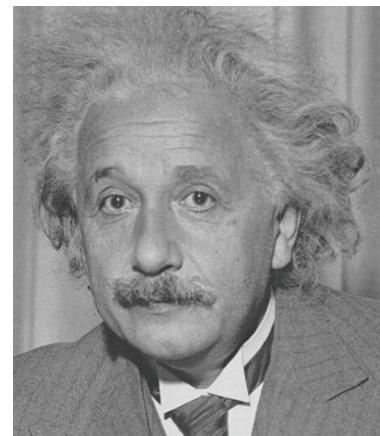
This is the relativistic formula for K.E.

When the body is at rest, the internal energy stored in the body is  $m_0 c^2$ .  $m_0 c^2$  is called the rest mass energy. The total energy ( $E$ ) of the body is the sum of K.E. ( $E_k$ ) and rest mass energy ( $m_0 c^2$ ).

$$\therefore E = E_k + m_0 c^2 = (mc^2 - m_0 c^2) + m_0 c^2 = mc^2.$$

$$\therefore E = mc^2$$

This is Einstein's mass-energy relation.



**Albert Einstein (1879-1955)**

Einstein almost single-handedly created "modern physics" and so occupies a special place in the subject.

Relativity and the quantum theory depend on Einstein.

### Notes

- (1) This relation states a universal equivalence between mass and energy. It means that mass may appear as energy and energy as mass.

**Example.** Consider the phenomenon of pair-annihilation or pair production. In this phenomenon, an electron and a positron can combine and literally disappear. In their place we find high energy radiation called  $\gamma$ -radiation, whose radiant energy is exactly equal to the rest mass plus kinetic energies of the disappearing particles. The process is reversible, so that a materialization of mass from radiant energy can occur when a high enough energy  $\gamma$ -ray, under proper conditions, disappears. In its place appears a positron-electron pair whose total energy (rest mass + K.E.) is equal to the radiant energy lost.

(2) The relationship ( $E = mc^2$ ) between energy and mass forms the basis of understanding nuclear reactions such as *fission* and *fusion*. These reactions take place in nuclear bombs and reactors. When a uranium nucleus is split up, the decrease in its total rest mass appears in the form of an equivalent amount of K.E. of its fragments.

(3) The formula for K.E. reduces to the classical formula for  $v \ll c$ .

$$E_k = mc^2 - m_0c^2 = (m - m_0)c^2 = m_0c^2 [(1 - v^2/c^2)^{-1/2} - 1]$$

Now, since  $v \ll c$ ,  $\left(1 - \frac{v^2}{c^2}\right)^{-1/2} = 1 + \frac{v^2}{2c^2} + \dots$  (neglecting higher order terms).

$$\therefore E_k = m_0c^2 \times \frac{v^2}{2c^2} = \frac{1}{2}m_0v^2.$$

**Unified mass unit.** The basic mass unit used in Atomic Physics is called the unified mass unit ( $u$ ). It is defined as (1/12)th of the rest mass of carbon—12 atom. Thus,  $1u = 1.66 \times 10^{-27}$  kg.

Now,

$$\begin{aligned} E &= m_0c^2 = (1.66 \times 10^{-27})(3 \times 10^8)^2 = 1.49 \times 10^{-10} \text{ J.} \\ &= \frac{1.49 \times 10^{-10}}{1.6 \times 10^{-19}} \text{ eV} = 931.3 \text{ MeV.} \end{aligned}$$

Hence the energy equivalent to unified mass unit is 931 MeV.

**EXAMPLE 1.** If 4 kg of a substance is fully converted into energy, how much energy is produced?

**SOL.** Here,

$$\begin{aligned} m &= 4 \text{ kg}, c = 3 \times 10^8 \text{ ms}^{-1}; E = ? \\ E &= mc^2 = 4 \times (3 \times 10^8)^2 = 3.6 \times 10^{17} \text{ J.} \end{aligned}$$

**EXAMPLE 2.** Calculate the rest energy of an electron in joules and in electron volts.

**SOL.** Here,  $m_0$  = rest mass of the electron =  $9.11 \times 10^{-31}$  kg;

$$\begin{aligned} c &= 3 \times 10^8 \text{ ms}^{-1} \\ \therefore E &= m_0c^2 = (9.11 \times 10^{-31})(3 \times 10^8)^2 = 8.2 \times 10^{-14} \text{ J.} \\ &= \frac{8.2 \times 10^{-14}}{1.6 \times 10^{-19}} \text{ eV} \quad (\text{since } 1 \text{ eV} = 1.6 \times 10^{-19} \text{ J).} \\ &= 0.51 \text{ MeV} \quad [\text{mega electron volt}.] \end{aligned}$$

**EXAMPLE 3.** Calculate the K.E. of an electron moving with a velocity of 0.98 times the velocity of light in the laboratory system.

**SOL.** Relativistic formula for K.E. is  $T = (m - m_0)c^2$ .

Here,  $m_0$  = rest mass of electron =  $9.11 \times 10^{-31}$  kg.

$$c = 3 \times 10^8 \text{ ms}^{-1}; v = 0.98c.$$

We can find  $m$  using the formula

$$\begin{aligned} m &= \frac{m_0}{\sqrt{(1 - v^2/c^2)}} = \frac{m_0}{\sqrt{1 - (0.98c/c)^2}} = 5.02m_0 \\ \therefore \text{K.E.} &= (5.02m_0 - m_0)c^2 = 4.02m_0c^2 \\ &= 4.02 \times (9.11 \times 10^{-31})(3 \times 10^8)^2 = 3.296 \times 10^{-13} \text{ J.} \end{aligned}$$

#### 1.14.1. Relationship Between the Total Energy, the Rest Energy, and the Momentum

The total relativistic energy (rest mass plus kinetic) of a particle is

$$E = mc^2 = m_0c^2 / (1 - v^2/c^2)^{1/2} \quad \dots(1)$$

The momentum of the particle is  $p = mv$

$$\text{or } v = p/m. \quad \dots(2)$$

$$\therefore E = m_0 \frac{c^2}{\sqrt{[1 - (p^2/c^2)^2]}} = \frac{m_0 c^2}{\sqrt{[1 - (p^2 c^2 / m^2 c^4)]}} = m_0 \frac{c^2}{\sqrt{[1 - (p^2 c^2 / E^2)]}}$$

$$\text{or } E^2 = m_0^2 \frac{c^4}{(1 - p^2 c^2 / E^2)} \quad \text{or} \quad E^2 - p^2 c^2 = m_0^2 c^4$$

$$\therefore E^2 = m_0^2 c^4 + p^2 c^2 \quad \dots(3)$$

### 1.14.2 Massless Particles

- In classical mechanics, a particle must have rest mass in order to have energy and momentum.
- But in relativistic mechanics, this requirement does not hold.

A particle can exist which has no rest mass but which exhibits such particlelike properties as energy and momentum.  
Relativistic momentum of a particle,

$$p = \frac{m_0 v}{\sqrt{1 - v^2/c^2}} \quad \dots(1)$$

Total energy of a particle,

$$E = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}} \quad \dots(2)$$

From Eqs. (1) and (2), when  $m_0 = 0$  and  $v \ll c$ , it is clear that  $E = p = 0$ .

A massless particle with a speed less than that of light can have neither energy nor momentum.

However when  $m_0 = 0$  and  $v = c$ ,  $E = 0/0$  and  $p = 0/0$ , which are indeterminate:  $E$  and  $p$  can have any values. Thus Eqs. (1) and (2) are consistent with the existence of massless particles that possess energy and momentum *provided that they travel with the speed of light*.

The relationship between  $E$ ,  $m_0 c^2$  and  $p$  is

$$E = \sqrt{(pc)^2 + (m_0 c^2)^2} \quad \dots(3)$$

- For a massless particle with  $m_0 = 0$ :

$$E = pc \quad \dots(4)$$

Fig. 1.9 shows the relativistic energy as a function of momentum for massive and massless particles.

The conclusion is not that massless particles necessarily occur, only that the laws of physics do not exclude the possibility as long as  $v = c$  and  $E = pc$  for them. In fact, a massless particle—the photon—indeed exists and its behavior is as expected.

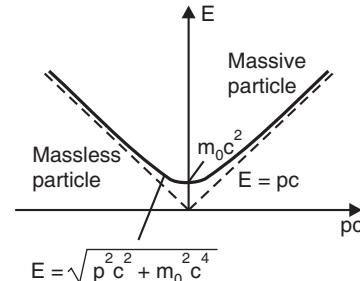
### 1.14.3 To show that the Rest Mass of a Photon is Zero

A photon travels with the velocity of light. Hence we must use the relativistic expression for its momentum,  $p$ .

Thus, for a photon

$$p = \frac{m_0 v}{\sqrt{1 - (v^2/c^2)}} \quad \dots(1)$$

Here,  $m_0$  is the rest mass and  $v$  the velocity of the photon.



**Fig. 1.9**

From quantum theory of radiation, the momentum of a photon of radiation of wavelength  $\lambda$  is

$$p = h/\lambda \quad \dots(2)$$

Here,  $h$  is the Planck's constant.

$$\therefore \frac{h}{\lambda} = \frac{m_0 v}{\sqrt{[1 - (v^2/c^2)]}} \quad \dots(3)$$

or  $m_0 = \frac{h}{v\lambda} \sqrt{[1 - (v^2/c^2)]}$ .

Since for photon,  $v = c$ , we have

$$m_0 = 0.$$

Hence the rest mass of photon is zero.

### 1.15 MINKOWSKI'S FOUR DIMENSIONAL SPACE-TIME CONTINUUM

Space is three dimensional in the sense that it is possible to describe the position of a point by three coordinate numbers as in Euclidean geometry. But the truth is that the world in which we live is a four-dimensional space-time continuum. This is revealed by the theory of relativity.

Any event in any frame of reference  $S$  can be represented by the four coordinates  $x, y, z, t$ . Suppose the same event is observed in another frame  $S'$  moving with uniform velocity  $v$  relative to  $S$  along the X-axis and coinciding with  $S$  at  $t = t' = 0$ . Then its coordinates in  $S'$  will be  $x', y', z', t'$ , satisfying the equation

$$x^2 + y^2 + z^2 - c^2 t^2 = x'^2 + y'^2 + z'^2 - c^2 t'^2 \quad \dots(1)$$

It is possible to characterise the Lorentz transformations more elegantly if we insert  $x_1 = x$ ,  $x_2 = y$ ,  $x_3 = z$  and  $x_4 = \sqrt{-1} ct = ict$  for the system  $S$  and similarly for  $S'$ . The resulting form of equation (1) is

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = x'_1^2 + x'_2^2 + x'_3^2 + x'_4^2 \quad \dots(2)$$

The imaginary time coordinate  $x_4$  appears here in exactly the same way as the space coordinates  $x_1, x_2$  and  $x_3$ . This four dimensional manifold which appears as a linking together of space and time is known as the "Minkowski world" or the 'four dimensional space-time continuum'. Here the *four coordinates* are  $x, y, z$  and  $ict$ , and time is equivalent to an imaginary length. The world of Minkowski involves space and time inextricably bound with each other. A point in such a space-time system is known as the "world point", and the motion of a particle in it may be represented by a "world line".

If  $s$  be the distance between two points with co-ordinates  $(x_1, x_2, x_3, x_4)$  and  $(x'_1, x'_2, x'_3, x'_4)$  in the four dimensional space, then

$$s^2 = (x'_1 - x_1)^2 + (x'_2 - x_2)^2 + (x'_3 - x_3)^2 + (x'_4 - x_4)^2$$

This can be shown to remain invariant *i.e.*, unaltered in form in all systems by the use of the transformation equations. For a "worldline element" this invariant form of equation is

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 = dx^2 + dy^2 + dz^2 - c^2 dt^2$$

This means that the distance between two points, *i.e.*, a length, is not defined by the three dimensions  $dx, dy, dz$  alone, but in addition, by a fourth dimension of time  $dt$ .

For  $ds = \text{minimum}$ , we get a straight line between two world points.

This straight world line is called a *geodesic*, for which

$$\delta \int_A^B ds = 0.$$

### 1.15.1. Four-vector

The four coordinates,  $x, y, z, ict$  define a vector in space-time.

This *four-vector* remains fixed in space-time regardless of any shift in point of view from one inertial frame  $S$  to another  $S'$ .

Another *four-vector* whose magnitude remains constant under Lorentz transformations has the components  $p_x, p_y, p_z, iE/c$ . Here  $p_x^2, p_y^2, p_z^2$  are the components of the linear momentum of a body whose total energy is  $E$ . Hence  $p_x^2 + p_y^2 + p_z^2 - (E^2/c^2)$  is invariant though  $p_x, p_y, p_z$  and  $E$  may be separately different.

## 1.16 THE GENERAL THEORY OF RELATIVITY

The general theory of relativity is based on the following two postulates.

**Postulate 1.** “*The laws of physics may be expressed in equations having the same form in all frames of reference, regardless of their states of motion*”.

Thus the general theory covers accelerated as well as uniform motion. Hence it is able to describe gravitational phenomena.

**Postulate 2.** *There is no way for an observer in a closed laboratory to distinguish between the effects produced by a gravitational field and those produced by an acceleration of the laboratory.*

This postulate is known as the *principle of equivalence*. This postulate follows from the experimental observation that the inertial mass of a body is always exactly equal to its gravitational mass.

Einstein generalised the principle of equivalence and proposed that the physical laws have similar mathematical forms not only in inertial systems but in accelerated systems also. According to him, the idea of action at a distance for exploration of gravitation is false and by selecting a suitable coordinate system, the effect of gravitation can be eliminated. In this system, the inertial mass of a body is equal to its gravitational mass. To explain these ideas, let us consider the weight of a body in a lift. If the lift goes upwards with an acceleration  $a$ , the weight of the body is  $m(g + a)$ . If the lift goes down with acceleration  $a$ , the weight of the body is  $m(g - a)$ . If the lift goes down with acceleration  $g$ , the weight of the body will be zero. Hence the gravitational effect on the body depends on the state of the lift.

## 1.17 PREDICTIONS OF GENERAL RELATIVITY

### 1. Deflection of light in a gravitational field

Inertial mass  $hv/c^2$  is associated with a light pulse of energy  $hv$ . From the principle of equivalence, the gravitational mass of this quantum is  $hv/c^2$  so that light ought to be subject to gravitational forces. For example, light rays that pass near the sun ought to be deflected towards it, just as the paths of the planets are deflected by the sun's gravitational field. This deflection for a ray of light is shown in Fig. 1.10. The deflection is known as '*Einstein effect*'. This effect can be detected at the time of total solar eclipse. The experimental value of the observed deflection is almost equal to its theoretical value which is 1.75 seconds.

### 2. Advance of the perihelion of Mercury's orbit

The general theory predicts that the path of a planet about the sun will not be precisely an ellipse. The orbit will slowly rotate in the plane of its motion, so that the axes turn in space in the same sense as the planet's rotation. For the planets in our solar system, this effect should be

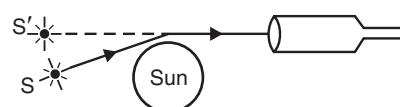


Fig. 1.10

just sufficiently great to be observed in the case of Mercury, which passes closest to the sun. The point of the orbit at which the planet is nearest to the sun is called the '*perihelion*'. The perihelion itself is found to be very slowly rotating around the sun, so that the orbit is as shown in Fig. 1.11. This effect is called the '*advance of the perihelion*' of Mercury. Observed and predicted values of this rotation of the planet's orbit ( $2 \times 10^{-4}$  rad. per century) agree to within the accuracy of astronomical measurement.

### 3. Gravitational red shift

The third prediction of general relativity is that *a stationary clock runs more slowly in a strong gravitational field than in a weak one*.

Suppose a clock has an interval of  $t$  between ticks when it is infinitely far from all matter. When the clock is at a distance  $R$  from a mass  $M$ , then it will have an interval of

$$t = t_0 / \sqrt{1 - 2GM/R}$$

between ticks.

An atom emitting light of a definite frequency is equivalent to a clock (atomic clock). The third prediction means that the light emitted by an atom on a star must be of frequency lower than that emitted by the same atom on earth (Since the gravitational field of star is much more stronger than that of earth). In other words, the light from the star must be shifted towards the red end of the spectrum (apart from any Doppler's shift). This effect is called the "*gravitational red shift*". The gravitational red shift has been actually observed in the spectral lines of certain stars called '*white dwarfs*', and also in a laboratory experiment involving '*Mossbauer effect*' and has verified the prediction.

## 1.18 PHOTONS AND GRAVITY

The second postulate of general relativity states that *there is no way for an observer in a closed laboratory to distinguish between the effects produced by a gravitational field and those produced by an acceleration of the laboratory*. This postulate is known as the *principle of equivalence*. This postulate follows from the experimental observation that the inertial mass of a body is always exactly equal to its gravitational mass.

For instance, if a man is inside a closed elevator car in outer space (Fig. 1.12), and if the car is suddenly accelerated upwards, he cannot decide whether a gravitational force had pulled him downwards or the car had been accelerated upwards.

The most interesting prediction of the general theory of relativity is that light passing near a large mass will be bent. The following two lines of reasoning lead to the conclusion that light should be affected by gravity.

1. The existence of momentum for a photon implies that it must have an effective mass  $m$  also. This mass can be computed by mass-energy relation ( $E = mc^2$ ).

$$m = \frac{E}{c^2} = \frac{h\nu}{c^2}.$$

Because gravitational mass is inseparable from inertial mass, light ought to be subject to gravity.

2. Consider a spaceship with a hole in the side allowing a ray of light to enter. If it is accelerating upward as shown in Fig. 1.13 (a), an observer inside will say that the light is bent down in a parabolic path as shown. The bending of the light is due to the motion of the rocket ship during

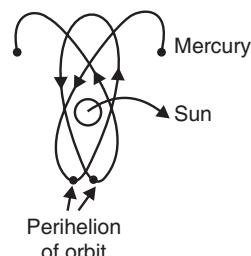


Fig. 1.11

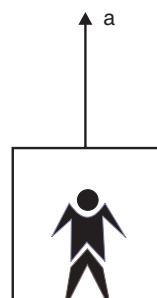


Fig. 1.12

the period of time it takes the light to travel from the entrance hole to the far side. The principle of equivalence indicates that exactly the same observation should be made in a rocket ship at rest on the surface of a large planet [Fig. 1.13 (b)]. The curvature of space-time will cause a light ray to be bent in the direction of the acceleration due to gravity.

The bending of light in a gravitational field is observed in the deflection of light from stars, moving past the sun. During a solar eclipse a star  $S$  (Fig. 1.14) near the sun appears to be displaced to  $S'$ . This is because the light from the star passing near the sun is bent in sun's gravitational field. The experimental value of the observed deflection is almost equal to its theoretical value which is 1.75 seconds. Thus *the theoretical prediction of the bending of light rays in the gravitational field agrees with the observations.*

The gravitational behaviour of light can also be demonstrated in the laboratory. Consider a photon falling through a height  $H$ . The photon gains the energy  $mgH$ . This gain in energy is manifested as an increase in frequency from  $\nu$  to  $\nu'$ .

Final photon energy ( $h\nu'$ ) = Initial photon energy ( $h\nu$ ) + increase in energy ( $mgH$ )

$$h\nu' = h\nu + \frac{h\nu}{c^2}gH$$

This result has been verified by Pound and Rebka (1960) using Mossbauer effect.

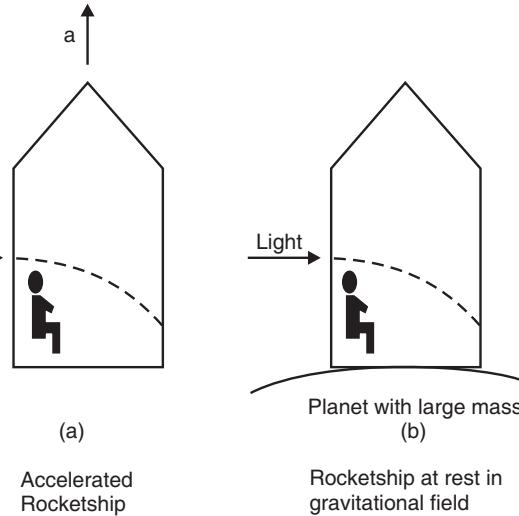


Fig. 1.13

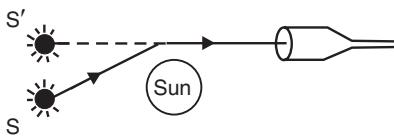


Fig. 1.14

### 1.19 GRAVITATIONAL RED SHIFT

The general theory of relativity predicts that the frequency of a spectral line due to an atomic transition will appear to be less if the light travels against the gravitational field before being measured by an observer at some distance from the source emitting the light. Hence the measured wavelength will shift towards the red. Let us derive an expression for gravitational red shift.

Consider a star of mass  $M$  and radius  $R$  (Fig. 1.15). Suppose a photon of initial frequency  $\nu$  is emitted by the star.

$$\text{Mass of the photon} = m = h\nu/c^2.$$

The potential energy of the photon on the star's surface is

$$V = -\frac{GMm}{R} = -\frac{GMh\nu}{c^2R}$$

Quantum energy of the photon =  $h\nu$

The total energy  $E$  of the photon is

$$E = h\nu - \frac{GMh\nu}{c^2R} = h\nu \left(1 - \frac{GM}{c^2R}\right)$$

When the photon arrives at the earth, it is beyond the star's gravitational field but its total energy remains the same. The photon's energy is now entirely electromagnetic. Let  $v'$  be the frequency of the photon at the earth. Then,

$$\begin{aligned} E &= hv' \\ \therefore hv' &= hv\left(1 - \frac{GM}{c^2 R}\right) \\ \frac{v'}{v} &= 1 - \frac{GM}{c^2 R} \end{aligned}$$

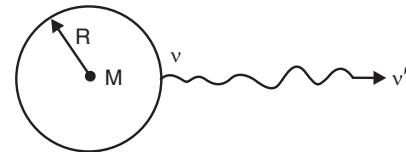


Fig. 1.15

The relative frequency change is

$$\frac{\Delta v}{v} = \frac{v - v'}{v} = 1 - \frac{v'}{v} = \frac{GM}{c^2 R} \quad \dots(1)$$

Thus there is a red shift for the spectral lines of light emitted by an atom situated in a gravitational field when this light is observed in a gravitation free part of space.

**EXAMPLE 1.** The sun's mass is  $2 \times 10^{30}$  kg and its radius is  $7 \times 10^8$  m. Find the gravitational red shift in light of wavelength 500 nm emitted by the sun.

**SOL.** The relative frequency change is

$$\begin{aligned} \frac{\Delta v}{v} &= \frac{GM}{c^2 R} \text{ or } \Delta v = v \cdot \frac{GM}{c^2 R} \\ \Delta \left( \frac{c}{\lambda} \right) &= \frac{c}{\lambda} \frac{GM}{c^2 R} \text{ or } -\frac{c}{\lambda^2} d\lambda = \frac{c}{\lambda} \frac{GM}{c^2 R} \text{ or } |d\lambda| = \lambda \cdot \frac{GM}{c^2 R} \\ \therefore d\lambda &= \frac{500 \times 10^{-9} \times 6.7 \times 10^{-11} \times 2 \times 10^{30}}{(3 \times 10^8)^2 \times 7 \times 10^8} \text{ m} = 1.06 \times 10^{-12} \text{ m.} \end{aligned}$$

**Black Hole.** Another prediction of the general theory that was pointed out in 1939 by Oppenheimer and Snyder was the possibility of what is called a *black hole*. Consider a star so dense that  $GM/(c^2 R) \geq 1$ . Then from Eq. (1) we see that no photon can ever leave the star. Such a star cannot radiate and so would be invisible—a *black hole* in space.

Now the question is how large would  $M$  have to be and how small would  $R$  have to be in order that the escape velocity would be equal to  $c$ ? According to the general theory, if an object of mass  $M$  has a radius given by

$$R_s = \frac{2GM}{c^2} \quad \dots(2)$$

the escape velocity would be  $c$ . This expression was derived by Schwarzschild.  $R_s$  is called *Schwarzschild radius*.

Since a black hole cannot obviously be seen in the usual sense, evidence for its existence can only be indicated by the effect on matter in its vicinity. A black hole that is a member of a double-star system will manifest its presence by its gravitational effect on the other member. Another possibility would be to observe  $\gamma$  radiation given off by charged particles as they are accelerated toward a black hole.

**EXAMPLE 2.** Calculate the Schwarzschild radius of a black hole of mass  $15 \times 10^6 M_\odot$

**SOL.**  $M = 15 \times 10^6 M_\odot = 15 \times 10^6 \times (\text{sun's mass}) = 15 \times 10^6 \times 1.9599 \times 10^{30} \text{ kg}$ ,

$G = 6.673 \times 10^{-11} \text{ Nm}^2 \text{ kg}^{-2}$ ,  $c = 3 \times 10^8 \text{ ms}^{-1}$ ; Schwarzschild radius,  $R_s = ?$

$$R_s = \frac{2GM}{c^2} = \frac{2 \times 6.673 \times 10^{-11} \times 15 \times 10^6 \times 1.9599 \times 10^{30}}{(3 \times 10^8)^2} = 4.425 \times 10^7 \text{ m}$$

## EXERCISES

**Part A: Choose the Correct Answer**

1. The length of the rod moving with a velocity  $v$  relative to the observer at rest is contracted by a factor

$$(a) \sqrt{1 - \frac{v^2}{c^2}} \quad (b) \sqrt{1 + \frac{v^2}{c^2}} \quad (c) \sqrt{1 - \frac{c^2}{v^2}} \quad (d) \sqrt{1 + \frac{c^2}{v^2}}$$

(Bharathiar; Nov. 2013)

2. The relativistic mass of a particle is

$$(a) m = \frac{m_0}{\sqrt{1 - \frac{v}{c}}} \quad (b) m = \frac{m_0}{\left(1 - \frac{v}{c}\right)} \quad (c) m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (d) m = \frac{m_0}{\sqrt{1 + \frac{v^2}{c^2}}}$$

(Bharathiar; April 2015)

3. Mass – Energy relation  $E = \underline{\hspace{2cm}}$

$$(a) mc \quad (b) mc^2 \quad (c) m^2c \quad (d) m^2c^2$$

(Bharathiar; April 2011)

4. The total relativistic energy is

$$(a) E = c\sqrt{p^2 + m_0^2c^2} \quad (b) E = c\sqrt{p^2 - m_0^2c^2} \quad (c) E = c\sqrt{p^2 + m_0c^2} \quad (d) E = c\sqrt{p^2 - m_0c^2}$$

(Bharathiar; April 2015)

5. The rest mass of photon is

$$(a) \text{positive} \quad (b) \text{finite} \quad (c) \text{infinite} \quad (d) \text{zero}$$

(Bharathiar; Nov. 2013)

**Ans.** 1. (a), 2. (c), 3. (b), 4. (a), 5. (d).

**Part B: Short Answer Questions**

6. What is meant by frame of reference? Define inertial and non-inertial frames. (Madras, Nov. 2006)

7. Deduce the Galilean transformation equations. (Bharathiar; April 2011)

8. Explain Ether hypothesis. (Bharathiar; April 2013)

9. State and explain the postulates of special theory of relativity. (Madras, April 2006)

10. Explain relativistic length contraction. (Bharathiar; April 2013)

11. Explain the relativistic time dilation. (Bharathiar; April 2014)

12. What is twin paradox? (Madras, April 2006)

13. If a particle could move with the velocity of light, how much kinetic energy would it possess?

(Madras, Nov. 2006)

[Ans. Since the mass of the particle moving with the velocity of light  $\left(m = \frac{m_0}{\sqrt{1 - v^2/c^2}}\right)$  becomes infinite, the kinetic energy would be also infinite. But it is impossible.]

14. Define unified mass unit and find its energy equivalence. (Madras, April 2006)

15. A particle of rest mass  $m_0$  moves with speed  $c/\sqrt{2}$ . Calculate its mass, momentum, total energy and kinetic energy. (Madras, April 2006)

**Hint.** (i)  $m = \frac{m_0}{\sqrt{1 - v^2/c^2}} = \frac{m_0}{\sqrt{1 - \frac{1}{2}}} = 1.41 m_0$

(ii)  $p = mv = \sqrt{2} m_0 \times c / \sqrt{2} = m_0 c$

(iii)  $E = mc^2 = 1.41 m_0 c^2$ .

(iv) K.E. =  $E - m_0 c^2 = 0.41 m_0 c^2$ . ]

16. What is the General Theory of Relativity? Discuss the important conclusions derived from it. What are the experimental observations in favour of these conclusions?
17. Is a laboratory at rest on the earth's surface really an inertial frame of reference?  
[Ans. No, because it is accelerated as a consequence of the earth's rotation on its axis and revolution around the sun].
18. An astronomer on earth flashes a pulsed laser, and 1.3 sec. later the pulse reaches the moon  $3.9 \times 10^8$  m away. An observer travelling in the same direction as the pulse sees the two events (*i.e.*, the flash and the arrival at the moon) as one event. What is the speed of this observer?  
[Ans. The observer must be moving at the speed of light in the same direction as the pulse.]
19. Does the mass of a substance increase on melting? Why?  
[Ans. Yes. Because an amount of energy equal to the specific latent heat of fusion has been added to the substance.]
20. If photons have a speed  $c$  in one reference frame, can they be found at rest in any other frame?  
[Ans. No. When a body moves with a speed less than  $c$  in one reference frame, we can find another reference frame in which it is at rest. But when the body moves with speed  $c$  in one reference frame, it will move with the same speed  $c$  in all reference frames].

**Part C: Long Answer Questions**

21. Describe the Michelson-Morley experiment and explain the physical significance of negative results.  
(Madras, Nov. 2006)
22. Derive the Lorentz space-time transformation formula. Discuss the length contraction and time dilation.  
(Madras, April 2006)
23. Deduce the formula for relativistic variation of mass with velocity. Briefly explain its significance.  
(Madras, April 2006)
24. What is the meaning of mass-energy equivalence? Obtain Einstein's mass energy relation. Show that 1 amu = 931 MeV.  
(Madras, Nov. 2006)

## PART-II

# ATOMIC PHYSICS

- 2. The Electron
- 3. Positive Rays
- 4. Structure of the Atom
- 5. X-Rays
- 6. The Photoelectric Effect and Planck's Quantum Theory

CHAPTER

# 2

## THE ELECTRON

**After reading this chapter, you should be able to**

- ◆ Describe the setting and purpose of Millikan's oil drop experiment
- ◆ Describe Dunnington's method for the precisional determination of  $e/m$  of electrons
- ◆ Develop good experimental technique, including proper setup and care of equipment, conducting experiments and analyzing results in order to observe physical phenomena.

## 2.1 DETERMINATION OF THE ELECTRONIC CHARGE : MILLIKAN'S OIL-DROP METHOD

**Experimental arrangement.** The apparatus consists of two optically plane parallel brass plates  $A$  and  $B$  (Fig. 2.1), separated by insulating rods of glass. The lower plate  $B$  is earthed. The upper plate  $A$  can be charged to a positive potential of the order of 10000 volts, from a high-tension battery. This observation chamber is situated in a bigger chamber  $C$  which is completely surrounded by a constant temperature bath of oil  $E$ . Tiny drops of a heavy non-volatile oil are sprayed into the chamber by an atomizer  $D$ . Some of the oil drops enter the space between the plates through the pinhole  $H$  in the top plate  $A$ . These drops are charged, due to the frictional effect at the nozzle of the atomizer. The air between the plates can be ionized by allowing X-rays to pass through it. Then the drops may pick up additional charges. The drops are illuminated by light from an arc lamp  $S$ . The oil-drop is observed by means of a short-focus telescope  $T$  provided with a millimetre scale in the eyepiece.

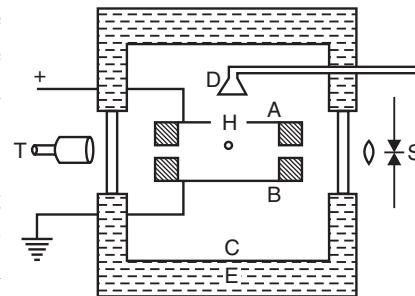


Fig. 2.1

**Experimental Procedure.** The plates are first connected together, so that the electric field between them is zero. A single drop is observed through the microscope as it descends under the action of gravity. As the drop is moving in a viscous medium (air), it will move with a constant terminal velocity  $v$ . Using a stop watch, the terminal velocity of the drop is measured. Let

$v$  = terminal velocity of the free fall,

$a$  = radius of the drop,

$\rho$  = density of oil,

$\sigma$  = density of air,

$\eta$  = coefficient of viscosity of air,

$m$  = effective mass of the drop and

$g$  = acceleration due to gravity.

Then by Stokes law,  $mg = 6\pi\eta av$  ... (1)

or  $\frac{4}{3}\pi a^3 (\rho - \sigma)g = 6\pi\eta av$  ... (2)

or  $a = \left[ \frac{9\eta v}{2(\rho - \sigma)g} \right]^{\frac{1}{2}}$  ... (3)

$\therefore$  The effective mass of the drop =  $m = \frac{4}{3}\pi a^3 (\rho - \sigma) = \frac{4}{3}\pi \left[ \frac{9\eta v}{2(\rho - \sigma)g} \right]^{\frac{3}{2}} (\rho - \sigma)$  ... (4)

Thus, by measuring  $v$ , we can calculate the effective mass  $m$  of the chosen oil drop.

Now, an electrostatic field  $X$  is applied steadily, so that the drop moves upwards with a steady velocity  $v_1$ . If the total charge on the drop is  $n_1 e$  (where  $n_1$  is an integer),

the resultant upward force on the drop =  $Xn_1e - mg$ .

According to Stokes law,  $Xn_1e - mg = 6\pi\eta av_1$  ... (5)

Dividing (5) by (1),  $\frac{Xn_1e - mg}{mg} = \frac{v_1}{v}$  or  $\frac{Xn_1e}{mg} = \frac{v_1 + v}{v}$

$$\therefore n_1 e = \frac{mg(v_1 + v)}{Xv} \quad \dots(6)$$

Now, the air between the plates is irradiated with X-rays. X-rays ionize the air molecules. The oil drop under observation may pick up one or more ionic charges. Due to the acquisition of charges, the velocity of the drop changes suddenly, when  $X$  is constant. Let the new velocity of the drop be  $v_2$  and the corresponding charge on the drop  $n_2 e$ . Then,

$$\begin{aligned} n_2 e &= \frac{mg(v_2 + v)}{Xv} & \dots(7) \\ \therefore (n_2 - n_1)e &= \frac{mg}{Xv}(v_2 - v_1) \end{aligned}$$

If  $v_2 > v_1$ ,  $(n_2 - n_1)e$  is positive and if  $v_2 < v_1$ ,  $(n_2 - n_1)e$  is negative. Both are possible, as both positive and negative ions are present.

Now for the same drop under the same electric field ( $X$ ),  $mg/Xv$  is constant. Hence  $v_2 - v_1 \propto (n_2 - n_1)e$ . i.e., any change in the velocity of the drop is proportional to the quantity of the acquired charge. Millikan made thousands of observations on a single drop. He found that there was a minimum value of  $(v_2 - v_1)$  and that other values of  $(v_2 - v_1)$  were simple integral multiples of this  $(v_2 - v_1)_{minm}$ .  $(v_2 - v_1)_{minm}$  corresponds to the addition or subtraction of one unit of charge. i.e.,  $(n_2 - n_1) = 1$  when  $(v_2 - v_1)$  is minimum.

$$\therefore e = \pm \frac{mg}{Xv}(v_2 - v_1)_{minm} \quad \dots(8)$$

Substituting the value of  $m$  from (4),

$$e = \pm \frac{4}{3} \pi \left( \frac{9\eta}{2} \right)^{\frac{3}{2}} \left[ \frac{v}{(\rho - \sigma)g} \right]^{\frac{1}{2}} \frac{(v_2 - v_1)_{minm}}{X} \quad \dots(9)$$

The value of electronic charge  $e$  found by Millikan was  $1.591 \times 10^{-19}$  C. The value of  $e$  now generally accepted is  $1.602 \times 10^{-19}$  C.

**Balanced Oil-Drop Method.** In this method, the intensity of the electric field is adjusted to a suitable value  $X$  such that the drop remains at rest. Then the effective weight of the drop ( $mg$ ) exactly balances the force due to the electric field ( $Xe$ ).

$\therefore Xe = mg$  or  $e = mg/X$ . (We have assumed that the drop carries one charge).

Substituting the value of  $m$  from (4),

$$e = \frac{4}{3} \frac{\pi}{X} \left( \frac{9\eta v}{2} \right)^{\frac{3}{2}} \frac{1}{[(\rho - \sigma)g]^{1/2}}$$

## 2.2 DUNNINGTON'S METHOD FOR DETERMINING E/M

### Experimental Arrangement

The apparatus is shown in Fig. 2.2. An alternating voltage at a constant high frequency produced by a crystal oscillator  $O$  is applied simultaneously to the two pairs of electrodes  $AA_1$  and  $BB_1$ . Electrons from the hot filament  $F$  are accelerated towards  $B_1$  during the positive half-cycle and emerge through a fine opening in  $B_1$ . The electrons are then bent into a circular path by a magnetic field  $B$  applied normal to the plane of the figure. The radius of curvature  $r$  of the circular path is defined by the slits  $S_1$ ,  $S_2$  and  $S_3$ . Then only those electrons whose velocity satisfies the relation  $Bev = mv^2/r$  can pass through the slits  $S_1$ ,  $S_2$  and  $S_3$ .

$\therefore$  the speed of the electron in its circular path

$$v = Ber / m \quad \dots(1)$$

After the electrons are turned through an angle  $\theta$ , they enter the Faraday chamber  $A$  through a grid  $A_1$  and produce a deflection in the galvanometer  $G$ . The electrodes and the slits are enclosed in an evacuated glass envelope. The grid  $A_1$  is connected to the same oscillator which accelerates the electrons. Suppose the time taken by the electrons to be turned through the angle  $\theta$  is the period  $1/f$  of the oscillator or an integral multiple of it, say  $n/f$ . Then the electrons will lose all their energy in overcoming the opposing P.D. between  $A_1$  and  $A$  and will just fail to reach  $A$ . The galvanometer indicates zero deflection when this condition is satisfied. This can be brought about by varying the intensity of the magnetic field  $B$ .

### Calculation

Now, distance travelled by the electron in going from  $B_1$  to  $A_1 = r\theta$ .

The time taken for travelling this distance  $= n/f$ .

$$\therefore v = \frac{\text{Distance}}{\text{Time}} = \frac{r\theta}{n/f} = \frac{r\theta f}{n} \quad \dots(2)$$

Hence, from (1) and (2),

$$\begin{aligned} \frac{Ber}{m} &= \frac{r\theta f}{n} \\ \therefore \frac{e}{m} &= \frac{f\theta}{nB}. \end{aligned}$$

$\theta$  is measured using microscopes built into the apparatus. The magnetic field ( $B$ ) can be determined accurately.  $f$  can be measured with an accuracy of one part in one million. This method is, therefore, a precision one. Thus,  $e/m$  is calculated. The value obtained was  $e/m = 1.7597 \times 10^{11}$  coulomb per kg.

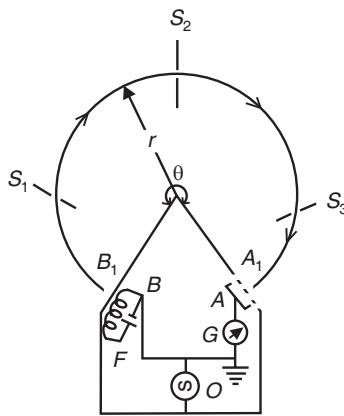


Fig. 2.2

### EXERCISE

1. Describe Millikan's method for the determination of electronic charge. Indicate the importance of this measurement.  
[Nagpur U., 2001]
2. With a neat diagram and necessary theory explain how  $e/m$  of an electron is obtained by Dunnington's method.  
[Bangalore, April 1992]
3. In a determination of  $e/m$  by Dunnington's method, frequency of A.C. = 20 MHz and  $\theta = 315^\circ$ . If the successive values of  $B$  for which the current is minimum are  $1.617 \times 10^{-4}$  and  $1.293 \times 10^{-4}$  weber/m<sup>2</sup>, find the value of  $e/m$ .  
(Ans.  $1.7 \times 10^{11}$  C kg<sup>-1</sup>)

## POSITIVE RAYS

**After reading this chapter, you should be able to**

- ◆ Identify the properties of positive rays
- ◆ Describe the apparatus used by Thomson to investigate the positive rays
- ◆ Explain how various mass spectrographs are used for the accurate determination of the masses of atomic nuclei
- ◆ List the various methods for separation of isotopes and recognize the general features of each method.

### 3.1 DISCOVERY

If the cathode in a discharge tube is perforated, “*luminous rays*” are observed behind the cathode. These rays travel in a direction opposite to the cathode rays, *i.e.*, from the anode to the cathode. These rays produce fluorescence when they strike a fluorescent screen present at the back of the cathode (Fig. 3.1). They are deflected by electric and magnetic fields. The direction of deflection indicates that they are *positively charged*. These rays are called *positive rays* or *canal rays*.

These rays consist of positively charged ions produced by the ionisation of the gas atoms by the fast moving electrons of the cathode rays. These positive ions travel towards the cathode and emerge out through the perforations in the cathode. The positive ions are gas atoms which have lost one or more electrons. Hence the mass of a positive ion is almost the same as that of a neutral gas atom.

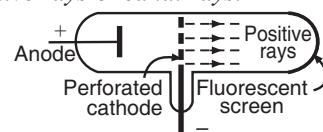


Fig. 3.1

### 3.2 PROPERTIES OF POSITIVE RAYS

- (1) These rays affect a photographic plate, produce fluorescence and penetrate thin aluminium foils.
- (2) They are deflected by electric and magnetic fields. The direction of deflection indicates that they are positively charged. For the same electric or magnetic field, the magnitude of deflection suffered by positive rays is much smaller than that of cathode rays. Therefore, the positive ray particles are much heavier than the electrons.
- (3) The velocities of all the positive ray particles is not the same. The velocities of these rays range from  $10^5$  to  $10^6 \text{ ms}^{-1}$ .

### 3.3 POSITIVE RAY ANALYSIS — THOMSON'S PARABOLA METHOD

**Apparatus.** Thomson determined the charge to mass ratio of positive ions by using the apparatus shown in Fig. 3.2. It consists of a discharge tube (*A*) in which the pressure of the gas is about  $10^{-5} \text{ m}$  of mercury. The anode is held in a side-tube. In order to ensure the supply of the gas under test, a steady stream of the gas is allowed to flow in through a capillary tube (*E*) and after circulating in *A* is pumped off at *F*. The cathode (*C*) is perforated with an extremely fine hole. The cathode is cooled by the water-jacket (*J*). The positive ions produced in *A* fly towards the cathode and those reaching it axially pass straight through the fine hole and emerge from the opposite end of the cathode as a narrow beam. This beam is then subjected to parallel electric and magnetic fields simultaneously. An electric field is applied between the plates *P* and *Q*. The electric field is perpendicular to the positive ray beam. *N* and *S* are the poles of a strong electromagnet. After passing through these fields, the beam enters a highly evacuated camera *G* and is received on a photographic plate *R*. A liquid air-trap (*T*) helps to keep the pressure in *G* quite low, even though pressure in *A* is comparatively large. The photographic plate, when developed, shows a series of parabolae.

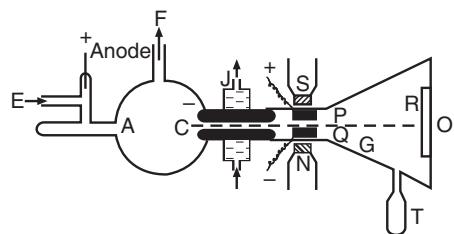


Fig. 3.2

**Theory.** Consider a positive ion of mass *M*, charge *E* and velocity *v*. When no electric or magnetic field is applied, the positive ion strikes the screen at *O*. This is called the *undeflected spot*.

**Action of Electric Field.** Let an electric field of strength *X* act over a length *l* of the path of the ion.

$$\left. \begin{array}{l} \text{Displacement of the ion in} \\ \text{passing through the electric field} \end{array} \right\} = s = \frac{1}{2} \left( \frac{XE}{M} \right) \left( \frac{l}{v} \right)^2.$$

After leaving the field, the ion moves in a straight line and finally strikes the plate at a distance  $x$  from  $O$ .  $x$  is proportional to  $s$  as well as to the distance between the field and the plate. Hence

$$\begin{aligned} x &\propto \frac{XEl^2}{2Mv^2} \\ \text{or} \quad x &= k_1 \frac{XE}{Mv^2} \end{aligned} \quad \dots(1)$$

Here  $k_1$  is a constant.

**Action of Magnetic Field.** Suppose a magnetic field of strength  $B$  is applied over the same length  $l$  in the same direction as that of the electric field. The positive ion will now be deflected by this field in a direction at right angles to that in which it was deflected by the electric field. It will strike the plate at a distance  $y$  from  $O$  such that  $oy$  is perpendicular to  $ox$  in the plane of the plate.

$$\left. \begin{array}{l} \text{Displacement of the ion just} \\ \text{emerging from the magnetic field} \end{array} \right\} = s' = \frac{1}{2} \left( \frac{BEv}{M} \right) \left( \frac{l}{v} \right)^2 = \frac{BEL^2}{2Mv}$$

On emerging from the field, the ion moves in a straight line and finally strikes the plate at a distance  $y$  from  $O$ .  $y$  is proportional to  $s'$  as well as to the distance between the field and the plate.

$$\begin{aligned} \text{Hence,} \quad y &\propto \frac{BEL^2}{2Mv} \\ \text{or} \quad y &= k_2 \frac{BE}{Mv} \end{aligned} \quad \dots(2)$$

Here  $k_2$  is another constant.

#### Action of Combined Electric and Magnetic Fields.

The combined effect of the two fields is found by eliminating  $v$  from (1) and (2).

Squaring (2) and dividing by (1),

$$\frac{y^2}{x} = \left( \frac{k_2^2}{k_1} \frac{B^2}{X} \right) \frac{E}{M} \quad \dots(3)$$

$B$  and  $X$  are constants. If  $E/M$  is constant, then by Eq. (3),  $\frac{y^2}{x} = \text{constant}$ . This is the equation of a parabola. As Eq. (3) is independent of  $v$ , particles of same  $E/M$  but of different velocities will fall on different points on the *same parabola*.

$\frac{y}{x} = \frac{k_2}{k_1} \frac{B}{X} v$ . i.e.,  $\frac{y}{x} \propto v$ . Thus, the position of any individual particle on the parabola will depend on the velocity of the particle.

The ions having different values of  $E/M$  will lie along the different parabolas.

For one direction of the magnetic field, one half of the parabola is traced. Reversing the magnetic field, the other half is also traced. When the full parabola is traced, it is easy to draw the axis of symmetry ( $X$ -axis).

**Determination of  $E/M$ .** The value of  $E/M$  can be calculated from Eq. (3) by measuring the coordinates  $x$  and  $y$  for a point on the parabola, evaluating the constants  $k_1$  and  $k_2$  for the apparatus and knowing  $B$  and  $X$ .

**Determination of Mass.** The mass of a positive ion is determined in terms of the mass  $M_1$  of the standard hydrogen ion. A small trace of hydrogen is always present in all samples of gases.

Hydrogen, being the lightest element, gives the outermost parabola. Let I and II represent the parabolic traces due to ions of the gas of mass  $M_2$  and the hydrogen ions of mass  $M_1$  respectively (Fig. 3.3). An ordinate  $a, b, c, d, e$  is drawn. It cuts the two parabolas at  $a, b, d, e$  and the  $X$ -axis at  $c$ . Let  $ac$  and  $bc$  represent the two values of  $y$  corresponding to a constant value of  $x$  on these parabolas. Let us assume that both ions have the same charge. Then from Eq. (3),

$$\frac{ac^2}{x} = \left( \frac{k_2^2}{k_1} \frac{B^2}{X} \right) \frac{E}{M_1}$$

and

$$\frac{bc^2}{x} = \left( \frac{k_2^2}{k_1} \frac{B^2}{X} \right) \frac{E}{M_2}$$

or

$$\frac{M_2}{M_1} = \frac{ac^2}{bc^2} = \left( \frac{ae}{bd} \right)^2$$

The lengths  $ae$  and  $bd$  can be measured on the photograph. Hence the parabolic traces enable us to compare the masses of different ions with hydrogen used as a standard.

$$M_2 = \left( \frac{ae}{bd} \right)^2 M_1.$$

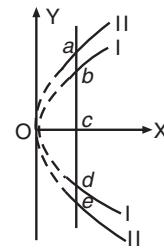


Fig. 3.3

**Discovery of Stable Isotopes.** Using neon gas in his apparatus, Thomson obtained two parabolas for the gas itself, a strong one corresponding to a mass 20 and a much weaker one corresponding to a mass 22. The intensity ratio of the two traces was 9 : 1 which gave the relative abundance of the two isotopes. Thomson, therefore, suggested that neon could exist in the form of two *isotopes*, chemically indistinguishable but with different masses 20 and 22. The actual observed atomic weight (20.2) of neon is the weighted mean of the masses of these two isotopes.

This established the existence of stable isotopes.

#### Limitations of the Parabola Method:

1. Due to the velocity dispersion, each parabolic trace is of very low intensity.
2. The traces on the photographic plate are blurred and have no definite edges. Hence accurate measurements are not possible.
3. The influence of secondary rays makes analysis difficult.

### 3.4 ASTON'S MASS SPECTROGRAPH

**Apparatus.** The apparatus used by Aston is shown in Fig. 3.4. The stream of positive ions obtained from a discharge tube is rendered into a fine beam by passing it between two narrow slits  $S_1$  and  $S_2$ . This beam enters the electric field between the metal plates  $P_1$  and  $P_2$ . Due to the action of the electric field ( $X$ ), all positive ions having the same value of  $E/M$  are not only deviated by an angle  $\theta$  from the original path but are dispersed by an angle  $d\theta$  due to their different velocities. The beam is then allowed to pass through a magnetic field  $M$  acting at right angles to the electric field so that it produces a deflection of the beam in the same plane. The magnetic field deviates the particles by an angle  $\phi$  and reconverges them by  $d\phi$ . The direction and magnitude of the field is so adjusted that it produces a deviation of the beam in the opposite direction and brings all ions having the same value of  $E/M$ , even though differing in velocities, to a focus at one point  $F$ . Ions having different values of  $E/M$  are brought to focus at different points on the photographic plate. The condition required for such a focusing may be derived as follows:

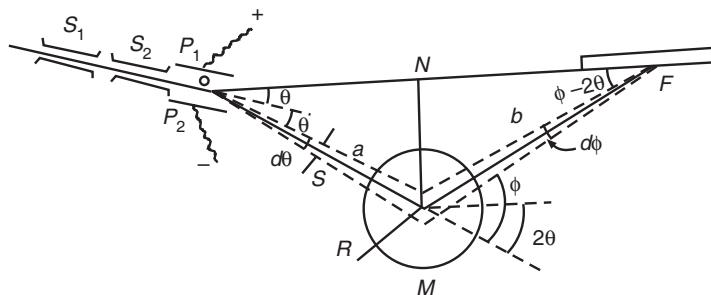


Fig. 3.4

**Theory: To obtain the condition of focusing.** Consider a group of ions having the same value of  $E/M$ , but moving with different velocities. Let  $\theta$  and  $\phi$  be the mean angles of deviation of the group of ions in the electric and magnetic fields respectively. Let  $d\theta$  be the dispersion angle due to the electric field and  $d\phi$  the convergence angle due to the magnetic field.

(i) Let

$X$  = The strength of the electric field,

$E$  = Charge of the ion.

$M$  = Mass of the ion and

$v$  = Velocity of the ion.

$l_1$  = the length of the electric field

and

$d_1$  = The linear displacement of an ion from its path due to the electric field.

then

$$d_1 = \frac{1}{2} \frac{XE}{M} \frac{l_1^2}{v^2}$$

$$\left. \begin{array}{l} \text{The angular deviation} \\ \text{of the beam} \end{array} \right\} = \theta = \frac{d_1}{l_1} = \frac{1}{2} \frac{XE}{Mv^2} l_1 \quad \dots(1)$$

$$\left. \begin{array}{l} \text{Angle of dispersion} \\ \text{of the beam} \end{array} \right\} = d\theta = \frac{-XEl_1}{M} \frac{dv}{v^3} \quad \dots(2)$$

$$\therefore \frac{d\theta}{\theta} = -\frac{2dv}{v} \quad \dots(3)$$

(ii) Similarly, if  $d_2$  = displacement of an ion from its path due to the magnetic field of strength  $B$ , and  $l_2$  = the length of path of the ion in the magnetic field,

$$d_2 = \frac{1}{2} \frac{BEv}{M} \frac{l_2^2}{v^2}$$

$$\left. \begin{array}{l} \text{The deviation produced by the magnetic} \\ \text{field in the opposite direction} \end{array} \right\} = \phi = \frac{d_2}{l_2} = \frac{1}{2} \frac{BE}{Mv} l_2 \quad \dots(4)$$

$$\therefore d\phi = -\frac{1}{2} \frac{BEL_2}{M} \frac{dv}{v^2} \quad \dots(5)$$

Hence

$$\frac{d\phi}{\phi} = -\frac{dv}{v} \quad \dots(6)$$

$$\text{Comparing (3) and (6), } \frac{d\theta}{\theta} = \frac{2d\phi}{\phi}$$

or

$$\frac{d\phi}{d\theta} = \frac{\phi}{2\theta} \quad \dots(7)$$

Let  $a = OR$  = The distance between the two fields. The width of the selected group of ions at  $R = a \ d\theta$ . If there had been no magnetic field, the width of the group after travelling a further distance  $b$  would be  $(a + b) d\theta$ . The magnetic field produces a convergence  $d\phi$  and brings the group of ions to a focus at a distance  $RF = b$ .

$\therefore$  The condition of focusing is given by

$$(a + b) d\theta = b d\phi \text{ or } \frac{a+b}{b} = \frac{d\phi}{d\theta} = \frac{\phi}{2\theta} \text{ [From Eq. 7]}$$

or

$$\frac{b}{a} = \frac{2\theta}{\phi - 2\theta} \quad \dots(8)$$

This expression represents a straight line drawn from  $O$ , making an angle  $2\theta$  with the direction of the beam deviated by the electric field. If  $\phi = 2\theta$ , the rays do not converge (since  $b = \infty$ ). However, if  $\phi = 4\theta$ ,  $b = a$  and the convergent beam can be easily photographed. The ions are focused at  $F$  such that  $OR = RF$ .

**Detection of isotopes.** To determine the masses of isotopes of an element, the element is mixed with a number of other elements whose atomic masses are known accurately. The traces of all of them are obtained on the same photographic plate. The distances of the traces of the known masses are measured from a given reference point on the plate. A calibration graph is drawn connecting the distances with the atomic masses. By measuring the distances of the traces of the isotopes of the element from the same reference point, their atomic masses are obtained from the calibration graph.

**Advantages :** 1. All the particles having the same value of  $E/M$  are focussed at a single point unlike Thomson's method in which they are spread out into a parabola. Hence, the intensities of the lines on the photographic plate in Aston's mass spectrograph are large, while the parabolic traces on the photographic plate in Thomson's arrangement are feeble.

2. The intensity of a line in the mass spectrum is proportional to the total number of particles of that mass. Hence a rough idea of the relative abundance of various isotopes of an element can be made by this method.

**Limitations :** 1. The mass scale is not linear.

2. Owing to the polarisation of the electrodes of the electric field, the traces are not quite straight but slightly curved.

**EXAMPLE.** The distance between traces corresponding to masses 12 and 16 in an Aston's mass spectrograph is 4.8 cm. Calculate the mass of the particle whose trace is at a distance of 8.4 cm from the trace of mass 16.

**SOL.** Let  $x$  be the distance of the fiducial mark from the trace of mass 16.

$$\therefore x \propto 16. \text{ Also, } x - 4.8 \propto 12.$$

$$\text{Hence } \frac{x}{x-4.8} = \frac{16}{12} \text{ or } x = 19.2 \text{ cm.}$$

If  $M$  be the mass of the particle whose trace is at a distance of 8.4 cm from the trace of mass 16, we have

$$19.2 \pm 8.4 \propto M \text{ and } 19.2 \propto 16.$$

$$\frac{M}{16} = \frac{19.2 \pm 8.4}{19.2} = 1 \pm \frac{8.4}{19.2} = 1 \pm \frac{7}{16}$$

$$\text{Hence } M = 16 \pm 7 = 23 \text{ or } 9.$$

### 3.5 BAINBRIDGE'S MASS SPECTROGRAPH

**Apparatus.** The apparatus is shown diagrammatically in Fig. 3.5. The beam of positive ions produced in a discharge tube is collimated by two slits  $S_1$  and  $S_2$  and enter a “velocity selector”. The velocity selector consists of (1) a steady electric field  $X$  maintained at right angles to the ion beam between two plane parallel plates  $P_1$  and  $P_2$  and (2) a magnetic field  $B$ . The magnetic field is produced by an electromagnet represented by the dotted circle. The magnetic field  $B$  is perpendicular to  $X$  and the ion beam. The electric field and magnetic field of the velocity selector are so adjusted that the deflection produced by one is nullified by the deflection produced by the other. If  $X$  and  $B$  are the electric intensity and magnetic induction, then  $Xe = Bev$  or  $v = X/B$ .

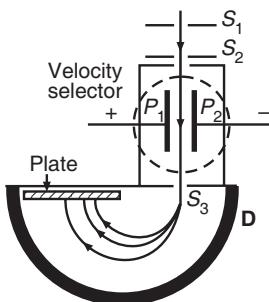


Fig. 3.5

Only those ions, having this velocity  $v$ , alone pass through the entry slit  $S_3$  to enter the evacuated chamber  $D$ . Thus all ions entering  $D$  must have the same velocity. The positive ions which enter into  $D$  are subjected to a strong uniform magnetic field of intensity  $B'$ , perpendicular to its path. The force acting on each ion will be  $B'e v$ .

**Theory.** Ions with different masses trace circular paths of different radii given by

$$R = \frac{Mv}{B'e} \left( \because B'e v = \frac{Mv^2}{R} \right)$$

$$\therefore \frac{e}{M} = \frac{v}{B'R} = \frac{X}{BB'R} \left( \because v = \frac{X}{B} \right).$$

Since  $v$  and  $B'$  are constant quantities,  $\frac{e}{M} \propto \frac{1}{R}$ .

After describing semicircles, the ions strike a photographic plate.

Now,  $M = \frac{B'eR}{v}$ . If  $e$  is the same for all ions, then  $M \propto R$ .

So we get a linear mass scale on the photographic plate. It will be seen that ions of different masses strike the photographic plate at different points, thus giving a typical mass spectrum.

**Advantages :** (1) Since a linear mass scale is obtained, accuracy of measurements is increased.

(2) The sensitivity depends on the strength of the deflecting magnetic field  $B'$  and the field area of the chamber  $D$ . Bainbridge used a magnetic field of 1.5 weber/m<sup>2</sup> over a semicircle of radius 0.2 m. He found a definite increase in resolving power over Aston's apparatus. The ten isotopes of tin were resolved by this instrument.

### 3.6 DEMPSTER'S MASS SPECTROGRAPH

**Experimental arrangement.** The experimental arrangement is shown in Fig. 3.6. The anode is a metal cylinder  $A$  with its front surface  $C$  coated with a salt of the element under test and heated electrically. A filament  $ff$ , if electrically heated by the battery  $B_1$ , emits electrons. By maintaining the filament at a P.D. of about 50 volts with respect to  $A$  by using another battery  $B_2$ , the electrons are made to bombard the heated salt with the result that the anode emits positively charged ions of

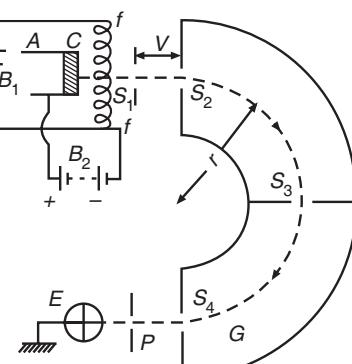


Fig. 3.6

the element. These ions are collimated into a narrow beam by the slit  $S_1$ . Then the positive ions are accelerated towards the slit  $S_2$  by a variable P.D.  $V$  maintained between  $S_1$  and  $S_2$ .

We know that when ions of mass  $M$  and charge  $e$  are accelerated through a P.D.  $V$ , they acquire a velocity  $v$  given by

$$\frac{1}{2} Mv^2 = eV$$

or  $v = \sqrt{\frac{2eV}{M}}$  ... (1)

The velocity of the ions  $v$  is very large compared with their initial speeds of emission. The ions then enter a vacuum chamber  $G$ . On entering the space  $G$ , the ions are subjected to a magnetic field of flux density  $B$  directed at right angles to the plane of the diagram. The ions are deflected through a semicircle towards the slit  $S_4$  defined by the three slits  $S_2$ ,  $S_3$  and  $S_4$ . The magnetic field of flux density  $B$  makes an ion of mass  $M$  entering with the velocity  $v$  traverse a circular path of radius  $r$  given by

$$Bev = \frac{Mv^2}{r}$$

or  $r = \frac{Mv}{Be}$  ... (2)

**Calculation.** From (1),  $v = (2eV/M)^{1/2}$

$$\therefore r = \frac{Mv}{Be} = \frac{M}{Be} \left( \frac{2eV}{M} \right)^{1/2} = \left( \frac{2MV}{B^2 e} \right)^{1/2}$$

or  $r^2 = \frac{2MV}{B^2 e}$

$$\therefore \frac{e}{M} = \frac{2V}{B^2 r^2}$$
 ... (3)

For this particular value of  $V$ , only ions with the specific charge given by equation (3) are collected by the electrode  $P$  and the current is registered by the electrometer  $E$ .

From the equation  $r = \left( \frac{2MV}{B^2 e} \right)^{1/2}$ , it is clear that if  $\left( \frac{M}{e} \right) \frac{V}{B^2}$  is kept

constant, the radius  $r$  of the path of these ions will remain unchanged. Hence ions of different  $e/M$  can be made to traverse a circular path of fixed radius by adjusting  $V$  and  $B$ .

**Determination of the masses of the isotopes:** In the actual experiment, the magnetic field ( $B$ ) is kept constant and the ionic current is measured by the electrometer for various values of  $V$ . A graph is plotted with this current against  $V$ . The graph will show sharp peaks at those values of  $V$  corresponding to the value of  $M$  given by Eq. (3). Fig. 3.7 gives the curves for potassium, showing the two isotopes of mass numbers 39 and 41. The relative abundance of the isotopes can be inferred from the relative heights of the peaks.

**EXAMPLE.** In a Dempster mass spectrograph, a 1000 volt accelerating potential brings  $^{12}\text{Mg}^{24}$  on the slit. What is the potential required to bring  $^{12}\text{Mg}^{25}$  on the slit, the magnetic field being kept constant?

**SOL.** The basic equations are:  $eV = \frac{1}{2} Mv^2$  and  $Bev = Mv^2/r$ .

Thus if  $V$  is the accelerating potential required to bring an ion of mass  $M$  on the slit, we have

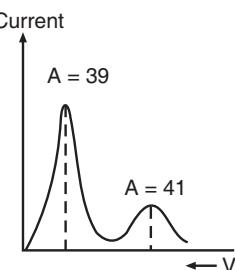


Fig. 3.7

$$V = \frac{B^2 r^2 e}{2M} = \frac{k}{M}, \text{ where } k \text{ is a constant.}$$

For  $M = 24$ ,  $V = 1000$  volts. Thus  $1000 = \frac{k}{24}$  or  $k = 24000$ .

For  $M = 25$ , we shall have  $V = \frac{k}{M} = \frac{24000}{25} = 960$  volts.

**Uses of Mass Spectrographs:** 1. Mass spectrographs have been used by many research workers to determine the isotopic masses of almost all elements on the periodic table. Their results show that there are more than 300 different isotopes occurring in nature.

2. Mass spectrographs have been employed to separate small quantities of pure isotopes. Large equipments have been built to separate large quantities of uranium -235 which is used in the manufacture of atomic bombs and production of nuclear power.

### 3.7 MASS DEFECT AND PACKING FRACTION

"Mass defect" of a given isotope is defined as the difference between the experimentally measured mass of the isotope ( $M$ ) and its mass number ( $A$ ).

$$\text{Mass defect} = \Delta = M - A.$$

The packing fraction of the isotope is defined as the ratio of its mass defect ( $M - A$ ) to its mass number ( $A$ ).

$$\text{Packing fraction} = (M - A)/A.$$

Since atomic masses are measure relative to C-12, the packing fraction for this isotope is zero.

### Binding Energy

**Definition.** When  $Z$  protons and  $N$  neutrons combine to make a nucleus, some of the mass ( $\Delta m$ ) disappears because it is converted into an amount of energy  $\Delta E = (\Delta m) c^2$ . This is called the binding energy (B.E.) of the nucleus.

$$\text{B.E.} = (Zm_p + Nm_n) c^2 - M_n c^2$$

where  $m_p$  = mass of the proton,  $m_n$  = mass of the neutron and  $M_n$  = nuclear mass.

**Atomic Masses.** Atomic masses refer to the masses of neutral atoms, not of bare nuclei. Thus an atomic mass always includes the masses of its  $Z$  electrons. Atomic masses are expressed in *mass units* ( $u$ ), which are so defined that the mass of a  $^{12}_6\text{C}$  atom is exactly  $12u$ .

The value of a mass unit is

$$1u = 1.66054 \times 10^{-27} \text{ kg.}$$

The energy equivalent of a mass unit is 931.49 MeV.

**Definition.** The atomic mass unit ( $u$ ) is defined as one-twelfth of the mass of the neutral carbon C-12 atom.

### 3.8 SEPARATION OF ISOTOPES

Since isotopes have the same chemical properties, they cannot be separated from one another by chemical methods. Isotopes can be separated from one another only by making use of the differences in their physical properties. All the methods of separation are based on the difference in the masses of the isotopes.

(1) *Mass spectrograph method.* The photographic plate in the mass spectrograph of Aston's type is replaced by different vessels where the different isotopes can be separately collected. Complete separation is possible by this method, but the yield is extremely small. The main difficulty is that

extremely high currents of ions or long intervals of time are required to separate any appreciable amount of an element into its isotopes.

(2) *Diffusion method.* The rate of diffusion of a gas through a porous partition is inversely proportional to the square root of the masses of the molecules. Hence different isotopes of a gas will diffuse with slightly different velocities.

Fig. 3.8 shows Hertz multiple-stage porous-tube apparatus for separating isotopes by diffusion.

The progress of the main volume of gas is from right to left, as it becomes enriched in the heavy fractions. The lighter fractions diffuse out through the porous tubes (crosshatched) and are returned by the recirculating pumps  $P$  to the previous stage, eventually collecting in the reservoir  $L$ . The heavier fractions progress from stage to stage, eventually collecting in the reservoir  $H$ . The over-all enrichment increases exponentially with the number of stages. Depending on the isotopes to be separated, some 10 to 50 or more stages may be used.

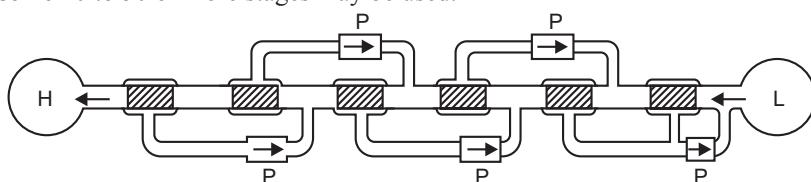


Fig. 3.8

(3) *Thermal diffusion method.* This is based on the fact that if a gaseous mixture of isotopes is placed in a vessel, one part of which is hotter than the other, the lighter molecules collect in the region of higher temperature.

Fig. 3.9 shows the Clusius-Dickel thermal-diffusion isotope separator. A cooled vertical glass tube with an electrically heated wire along its axis is used. By thermal diffusion the heavier fraction tends to concentrate at the cool outer wall.

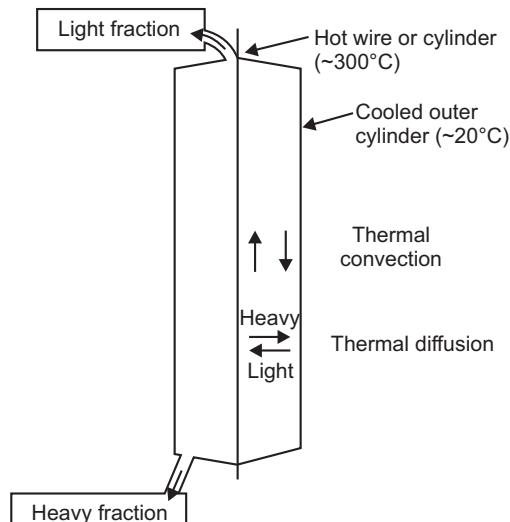


Fig. 3.9

The lighter fraction concentrates at the hot inner cylinder or wire. The action of gravity then causes thermal convection which provides an effective downward transport for the heavy fraction at the cool outer wall and an upward transport for the light fraction along the axis. The pressure is maintained at a low enough value to avoid turbulence in the thermal convective flow.

The process separation factor  $\alpha$ , in terms of the equilibrium mole concentrations in the heavy and light reservoirs, is

$$\alpha = e^{lA(M_2 - M_1)/(M_2 + M_1)}$$

Here,  $M_1$  and  $M_2$  are the masses of the light and heavy isotopic molecules being separated,  $l$  is the length of the column.

$A$  is a function of the viscosity, self-diffusion, and density of the gas, the temperatures and radii of the cylindrical walls, and the gravitational constant.

(4) *Pressure diffusion (or centrifugal) method.* This is based on the fact that if a gas or vapour flows into a rapidly rotating cylinder, the force acting on the molecules will result in an increased concentration of the heavier isotopes at the walls while the lighter ones collect near the axis of rotation. Isotopes of chlorine in carbon tetrachloride and of bromine in ethyl bromide have been separated by this method.

### EXERCISE

1. (i) Explain the Thomson's method for positive ray analysis.  
 (ii) Explain how Thomson's parabola method is helpful to determine the mass of the positive ion.  
*(Bharathiar, April 2011)*
2. Describe the construction of Aston's mass spectrograph with necessary theory. Show how it can be used in the detection of isotopes.  
*(Bharathiar, April 2012)*
3. Describe the principle, construction and working of Bainbridge's mass spectrograph. How is the nuclear mass determined using the spectrograph?  
*(Luck. U., 2007)*
4. Describe the construction, working and theory of a Dempster mass spectrograph.  
*(Bharathiar, April 2011)*

## STRUCTURE OF THE ATOM

**After reading this chapter, you should be able to**

- ◆ Describe Rutherford's gold foil experiment
- ◆ Describe the characteristics of Rutherford, Bohr, Sommerfeld and vector atomic models
- ◆ Explain the structure and spectra of simplest atoms using the quantum hypothesis
- ◆ Develop semi-classical models of the atom and show how these models lead to quantum mechanics
- ◆ Demonstrate that the energy is absorbed by atoms in discrete quanta
- ◆ Explain the Pauli exclusion principle and its application to the atom.
- ◆ Discuss the effect of magnetic field on atomic spectra
- ◆ Explain normal and anomalous Zeeman effect

### 4.1 RUTHERFORD'S EXPERIMENTS ON SCATTERING OF $\alpha$ -PARTICLES

**Experimental Arrangement.** The experimental arrangement is shown in Fig. 4.1.  $\alpha$ -particles are emitted by some radioactive material (polonium)  $R$ , kept inside a thick lead box.

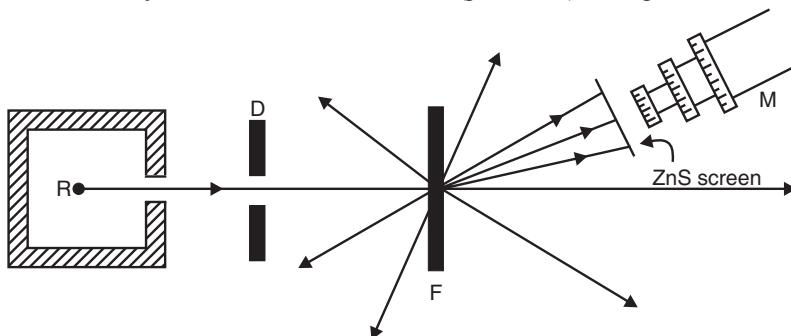


Fig. 4.1

$D$  is the lead screen with a small hole in it to select a very fine beam of  $\alpha$ -particles. This well collimated beam is then allowed to fall on a thin gold foil  $F$ . While passing through the gold foil, the  $\alpha$ -particles are scattered through different angles. A zinc sulphide screen was placed on the other side of the gold foil. This screen was movable, so as to receive the  $\alpha$ -particles, scattered from the gold foil at angles varying from  $0$  to  $180^\circ$ . When an  $\alpha$ -particle strikes the screen, it produces a flash of light and it is observed by the microscope  $M$ .

**Result:** It was found that:

- (1) Most of the  $\alpha$ -particles passed through the gold foil with small deviation.
- (2) There were a few particles that were scattered through large angles. Experimental results showed that one of about 8000 particles suffered angles of scattering  $> 90^\circ$ . A few of the  $\alpha$ -particles were even scattered directly in the backward direction (*i.e.*, angle of scattering =  $180^\circ$ ). Thomson's atom model failed to explain the experimental results. Therefore, Rutherford pictured an atom as follows.



Ernest Rutherford  
(1871-1937)

#### Rutherford Nuclear Atom Model.

- (1) The atom has a small, positively charged *nucleus*. All positive charges of an atom and most of the mass of the atom are concentrated in the nucleus. Electrons have no place inside the nucleus. The electrons revolve around this nucleus at some distance away. The dimensions of the nucleus and of the electrons are negligibly small as compared to the overall size of the atom, so that most of the volume occupied by an atom is actually an empty space. *Thus the discovery of the nucleus of the atom is due to Rutherford.*
- (2) Since the electron-distribution must give stability to the atom, the electrons could not be stationary. The electrons must be revolving around the nucleus in closed orbits, so that the centrifugal force due to the rotation of the electron is balanced by the force of electrostatic attraction between the nucleus and the electrons. Thus Rutherford proposed a dynamic planetary model in which nucleus plays the role of the sun and electrons correspond to the individual planets of the solar system.

Since most of the atom is empty, most of the  $\alpha$ -particles pass through the foil without appreciable deflection. However, if an  $\alpha$ -particle is moving in such a direction so as to have a head-on collision with the nucleus, the  $\alpha$ -particle will be scattered through large angles. Sometimes, when a head-on

collision occurs, the  $\alpha$ -particle is turned back. Thus the Rutherford's model of the nucleus enables us to account for very large deviations.

## 4.2 THEORY OF $\alpha$ -PARTICLE SCATTERING

We shall assume the following:

- (1) The  $\alpha$ -particle and the nucleus are point charges.
- (2) The scattering is due to coulomb electrostatic repulsive forces between the  $\alpha$ -particle and the positive charge of the nucleus ( $Ze$ ).
- (3) The nucleus is so massive compared with the  $\alpha$ -particle that it does not move during their interaction.
- (4) The  $\alpha$ -particles do not penetrate the nuclear region and the strong interaction nuclear forces are not involved.
- (5) Let  $r$  be the instantaneous separation between  $\alpha$ -particle and nucleus. Owing to the variation of the electric force with  $1/r^2$ , the  $\alpha$ -particle's path is a hyperbola with the nucleus at the outer focus (Fig. 4.2).

**Definitions.** (1) *The impact parameter  $b$  is the minimum distance to which the  $\alpha$ -particle would approach the nucleus if there were no forces between them.* For a head-on collision,  $b = 0$ . The impact parameter  $b$  should not be confused with the distance of closest approach  $D$ .

To determine  $D$ , consider an  $\alpha$ -particle at a great distance from the nucleus but approaching a head-on-collision with a kinetic energy  $K$ . At the point  $P$  in Fig. 4.2, the repulsive force of the nucleus stops the approaching  $\alpha$ -particle momentarily, and all of its K.E. is transformed into P.E. So we can write

$$K = \frac{1}{4\pi\epsilon_0} \frac{2Ze^2}{D} \quad \text{or} \quad D = \frac{1}{4\pi\epsilon_0} \frac{2Ze^2}{K}$$

(2) *The Scattering angle  $\theta$  is the angle between the asymptotic direction of approach of the  $\alpha$ -particle and the asymptotic direction in which it recedes.*

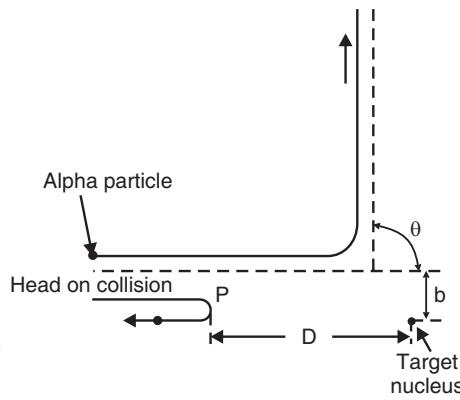


Fig. 4.2

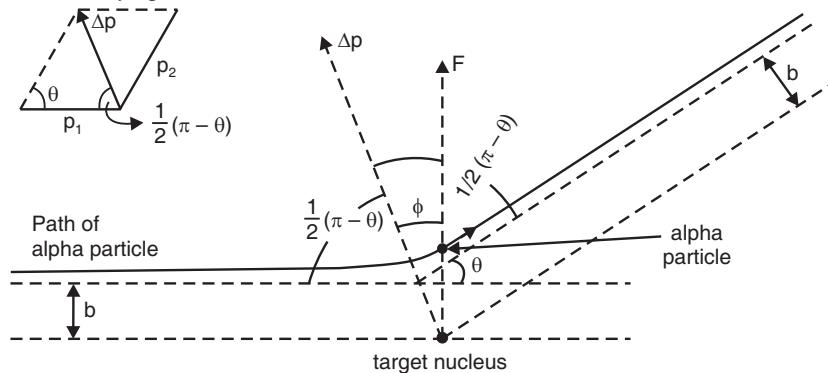


Fig. 4.3

**Relationship between  $b$  and  $\theta$ .** Let  $\vec{p}_1$  and  $\vec{p}_2$  be the linear momenta of the  $\alpha$ -particle when it is far from the nucleus before and after the interaction respectively. Then  $\frac{p_1^2}{2m} = \frac{p_2^2}{2m}$  or  $p_1 =$

$p_2 = mv$ . Here  $v$  is the  $\alpha$ -particle velocity far from the nucleus. The  $\alpha$ -particle moves under the action of a central force  $\vec{F} = \frac{1}{4\pi\epsilon_0} \left( \frac{2Ze^2}{r^2} \right)$  directed along the radius vector. From Newton's second law,

$$\Delta \vec{p} = \vec{p}_2 - \vec{p}_1 = \int_0^\infty \vec{F} dt \quad \dots(1)$$

From Fig. 4.3,  $\frac{\Delta p}{\sin \theta} = \frac{mv}{\sin(\pi-\theta)/2}$

or  $\Delta p = \frac{mv \sin \theta}{\sin(\pi-\theta)/2} = \frac{mv 2 \sin(\theta/2) \cos(\theta/2)}{\cos(\theta/2)} = 2mv \sin \theta / 2 \quad \dots(2)$

Let  $\phi$  be the instantaneous angle between  $\vec{F}$  and  $\Delta \vec{p}$  along the path of the  $\alpha$ -particle. Because the impulse  $\int \vec{F} dt$  is in the same direction as the momentum  $\Delta \vec{p}$ , its magnitude is  $\int F \cos \phi dt$ . Hence a scalar equation combining Equations (1) and (2) is

$$2mv \sin \frac{\theta}{2} = \int_{t=0}^{t=\infty} F \cos \phi dt$$

Changing variables from  $t$  to  $\phi$

$$2mv \sin \frac{\theta}{2} = \int_{-(\pi-\theta)/2}^{+(\pi-\theta)/2} F \cos \phi \frac{dt}{d\phi} d\phi \quad \dots(3)$$

$d\phi/dt = \omega$  is the angular velocity of the  $\alpha$ -particle about the nucleus. The electric force exerted by the nucleus on the  $\alpha$ -particle acts along the radius vector joining them. So there is no torque on the  $\alpha$ -particle and its angular momentum  $m\omega r^2$  is constant. Hence

$$m\omega r^2 = \text{constant} = mr^2 \frac{d\phi}{dt} = mvb \quad \text{or} \quad \frac{dt}{d\phi} = \frac{r^2}{vb}$$

Substituting  $F = \frac{1}{4\pi\epsilon_0} \left( \frac{2Ze^2}{r^2} \right)$  and  $\frac{dt}{d\phi} = \frac{r^2}{vb}$  in Eqn. (3),

$$\frac{4\pi\epsilon_0 mv^2 b}{Ze^2} \sin \frac{\theta}{2} = \int_{-(\pi-\theta)/2}^{+(\pi-\theta)/2} \cos \phi d\phi = 2 \cos \frac{\theta}{2}$$

or  $\cot \frac{\theta}{2} = \frac{2\pi\epsilon_0 mv^2}{Ze^2} b.$

Using  $K = mv^2/2$ ,

$$b = \frac{Ze^2}{4\pi\epsilon_0 K} \cot \frac{\theta}{2} \quad \dots(4)$$

This is the relation between the impact parameter  $b$  and the scattering angle  $\theta$ .

The scattering angle  $\theta$  decreases with increasing impact parameter (Fig. 4.4). In terms of  $D$ ,

$$b = \frac{D}{2} \cot \frac{\theta}{2} \quad \dots(5)$$

### Rutherford Scattering formula.

All  $\alpha$ -particles that approach the nucleus with an impact parameter  $\leq b$  will be scattered by an angle  $\geq \theta$ . The area around each nucleus (Fig. 4.4) with a radius equal to an impact parameter  $b$  is

called the *integral cross section*. This area is

$$\sigma = \pi b^2 \quad \dots(6)$$

Consider a gold foil with a surface area  $A$  and a thickness  $t$ . If  $n$  is the No. of nuclei per unit volume, the foil will contain  $nAt$  nuclei. Therefore the target area  $T$  offered by those nuclei in order to have scattering angles greater than  $\theta$  and impact parameter less than  $b$  will be

$$T = nAt\sigma \quad \dots(7)$$

Since  $A$  is the total target area, the fraction of  $\alpha$ -particles experiencing scattering angles greater than  $\theta$  will be

$$f = \frac{\text{target area offered by nuclei in the foil}}{\text{total target area}} = \frac{nAt\sigma}{A}$$

or

$$f = nt\sigma = nt \pi b^2$$

Substituting

$$b = \frac{Ze^2}{4\pi\epsilon_0 K} \cot \frac{\theta}{2},$$

$$f = \pi nt \left( \frac{Ze^2}{4\pi\epsilon_0 K} \right)^2 \cot^2 \frac{\theta}{2} \quad \dots(8)$$

In an actual experiment, a detector measures  $\alpha$ -particles scattered between  $\theta$  and  $\theta + d\theta$  (Fig. 4.5). The fraction of incident  $\alpha$ -particles so scattered is found by differentiating Eq. (8) with respect to  $\theta$ .

$$df = -\pi nt \left( \frac{Ze^2}{4\pi\epsilon_0 K} \right)^2 \cot \frac{\theta}{2} \cosec^2 \frac{\theta}{2} d\theta \quad \dots(9)$$

In the experiment, a fluorescent screen was placed at a distance  $r$  from the foil. The scattered  $\alpha$ -particles were detected by means of the scintillations they caused. Those  $\alpha$ -particles scattered between  $\theta$  and  $\theta + d\theta$  reached a zone of a sphere of radius  $r$  whose width is  $r d\theta$ . The zone radius  $= r \sin \theta$ . So the area of the screen struck by these particles is

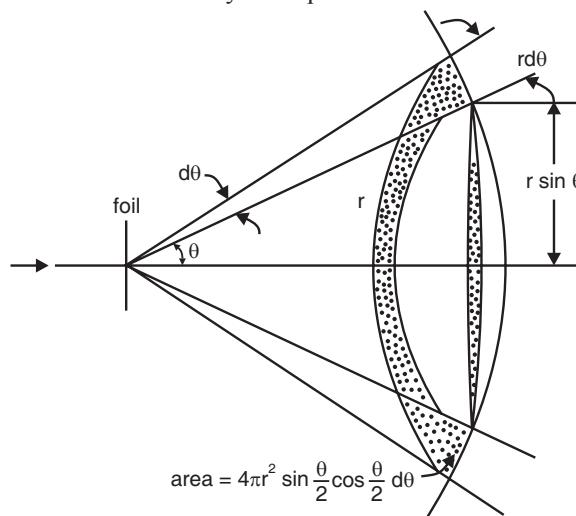


Fig. 4.5

$$ds = (2\pi r \sin \theta) (rd\theta) = 2\pi r^2 \sin \theta d\theta = 4\pi r^2 \sin \theta / 2 \cos \theta / 2 d\theta.$$

Let  $N_i$  be the total number of  $\alpha$ -particles directed against the target.

Then the No. scattered into  $d\theta$  at  $\theta = N_i df$

No. of particles per unit area striking the screen at  $\theta$

$$= N(\theta) = \frac{N_i |df|}{ds} = \frac{N_i \pi nt \left( \frac{Ze^2}{4\pi \epsilon_0 K} \right)^2 \cot \frac{\theta}{2} \cosec^2 \frac{\theta}{2} d\theta}{4\pi r^2 \sin \theta / 2 \cos \theta / 2 d\theta}$$

$$\therefore N(\theta) = \frac{N_i nt Z^2 e^4}{(8\pi \epsilon_0)^2 r^2 K^2 \sin^4 (\theta/2)} \quad \dots(10)$$

Eq. (10) is the Rutherford scattering formula.

#### Experimental Verification of the Rutherford Scattering Theory

The apparatus used by Geiger and Marsden is shown in Fig. 4.6a.  $R$  represents a radioactive substance (radon) which is the source of the  $\alpha$ -particles.  $D$  is a diaphragm.  $F$  is a very thin foil of scattering material.  $S$  is a zinc sulphide screen rigidly attached to a microscope  $M$ . The source and foil were held fixed. The screen and microscope could be rotated. The entire apparatus was enclosed in a metal box which could be evacuated.

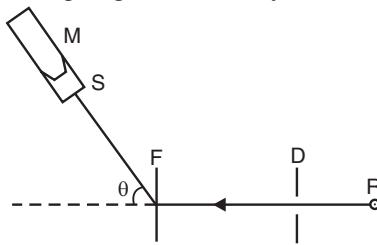


Fig. 4.6 (a)

(i) **Angular distribution of  $\alpha$ -particles scattered by thin foil.** The  $\alpha$ -particles from a radon source  $R$  passed through a diaphragm  $D$  and were scattered by the thin foil  $F$ . The  $\alpha$ -particles that were scattered through an angle  $\theta$  struck a zinc sulphide screen  $S$ . The scintillations were viewed through the microscope  $M$ . The number of particles striking the screen in a given time was measured at different angles over a range of angles from 5 to 150 degrees. Let  $N$  be the number of particles scattered per unit time through an angle  $\theta$ . They found that  $N \propto 1/\sin^4(\theta/2)$ .

(ii) **Dependence of scattering on the thickness of the foil.** The angle of scattering  $\theta$  was kept constant. Foils of different thicknesses and also of different materials were used. The number of particles scattered per minute,  $N$ , is plotted as ordinates and the thickness  $t$  of the scattering foil of a given material is plotted as abscissae (Fig. 4.6b). Here the thickness of each foil is expressed in terms of an equivalent length of air path. The straight lines show that for any one element  $N \propto t$ .

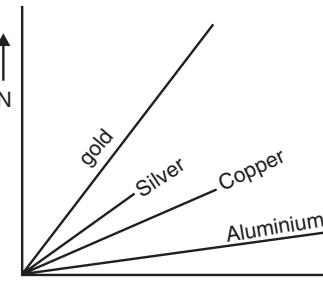


Fig. 4.6 (b)

(iii) **Dependence of scattering on the velocity of  $\alpha$ -particles.** Velocity of the incident  $\alpha$ -particles was changed by placing absorbing screens of mica between the source and the scattering foil. The velocities of the  $\alpha$ -particles were determined by first finding the range  $R$  of the  $\alpha$ -particles in air and then applying Geiger's law,  $R = av^3$ . Here  $a$  is a constant. It was found that  $Nv^4 = \text{constant}$ . This means  $N \propto 1/v^4$  or  $N$  is inversely proportional to the square of the kinetic energy ( $K$ ) of the  $\alpha$ -particles incident on the scatterer.

(iv) **The dependence of the scattering on the nuclear charge.** Chadwick determined the nuclear charge from scattering experiment. The arrangement is shown in Fig. 4.6 (c).  $AA'$  is the scattering foil

in the form of an annular ring.  $R$  is the source of  $\alpha$ -particles. The zinc sulphide screen  $S$  was placed such that  $RA = AS$ . The total number of  $\alpha$ -particles falling on the foil  $AA'$  could be determined by counting the number reaching  $S$  directly from  $R$ , since the areas of the screen  $S$  and the foil  $AA'$  were known. The  $\alpha$ -particles that were scattered through an angle  $\theta$  were counted by the scintillations they produced on the screen  $S$ . When the scattered rays were investigated, the direct rays from  $R$  were cut off completely by the thick lead plate  $L$ .

The values of  $Ze$  obtained for copper, silver and platinum foils were  $29.3\text{ e}$ ,  $46.3\text{ e}$  and  $77.4\text{ e}$ . These results are not precise enough to determine unique, integral values of  $Z$ . But they agree well with the values  $29$ ,  $47$  and  $78$  for the three elements, obtained by an entirely independent method. These results are in agreement with Rutherford's nuclear theory of the atom and provide the only direct measurement of the nuclear charge.

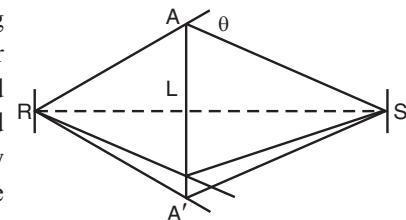


Fig. 4.6 (c)

### 4.3 BOHR ATOM MODEL

Bohr assumed basically Rutherford nuclear-model of the atom and tried to overcome the defects of the model. He proposed the following two postulates.

**Basic Postulates.** (1) *An electron cannot revolve round the nucleus in all possible orbits as suggested by the classical theory. The electron can revolve round the nucleus only in those allowed or permissible orbits for which the angular momentum of the electron is an integral multiple of  $\hbar$ .*

Here,  $\hbar$  is the Planck's constant—or rather, his original constant ( $h$ ) divided by  $2\pi$ :

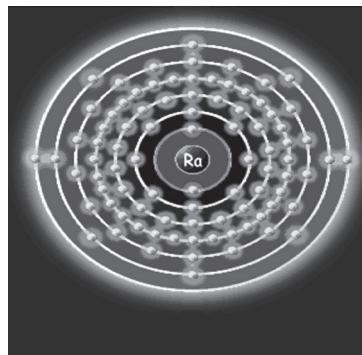
$$\hbar = \frac{h}{2\pi} = 1.054572 \times 10^{-34} \text{ Js.}$$

$$h = 6.64 \times 10^{-34} \text{ Js.}$$

These orbits are called stationary orbits and an electron revolving in these orbits does not radiate energy.

**Explanation.** For an electron of mass  $m$ , moving with a speed  $v$  in an orbit of radius  $r$ ,  $L = mvr = mr^2\omega = n\hbar$ .  $n$  is called the *Principal quantum number*. It takes integral values, excluding zero;  $n = 1, 2, 3, 4 \dots$

(2) *An atom radiates energy only when an electron jumps from a stationary orbit of higher energy to one of lower energy. If the electron jumps from an initial orbit of energy  $E_i$  to a final orbit of energy  $E_f$  ( $E_i > E_f$ ), a photon of frequency  $v = \frac{E_i - E_f}{h}$  is emitted.*



Bohr's Atomic Model of Radium.

**The Bohr formulae.** Based on these postulates, Bohr derived the formulae for (i) the radii of the stationary orbits and (ii) the total energy of the electron in the orbit.

Consider an atom whose nucleus has a positive charge  $Ze$  and mass  $M$  (Fig. 4.7). For hydrogen,  $Z = 1$ . Let an electron of charge  $(-e)$  and mass  $m$  move round the nucleus in an orbit of radius  $r$ . Since  $M \gg m$ , the nucleus is stationary. Hence the mass of the nucleus does not come into the calculations.

The electrostatic force of attraction  $\left. \right\} = \frac{1}{4\pi\epsilon_0} \frac{(Ze)(-e)}{r^2} \dots(1)$  between the nucleus and the electron

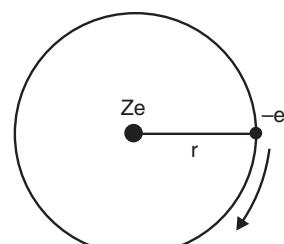


Fig. 4.7

The centrifugal force on the electron =  $mv^2/r$  ... (2)

The system will be stable if  $\frac{mv^2}{r} = \frac{1}{4\pi\epsilon_0} \frac{Ze^2}{r^2}$  ... (3)

According to Bohr's first postulate,  $mvr = n\hbar$  or  $v = \frac{n\hbar}{mr}$

$$\text{or } v^2 = \frac{n^2 \hbar^2}{m^2 r^2} \quad \dots(4)$$

Substituting this value of  $v^2$  in (3),

$$\frac{m}{r} \left( \frac{n^2 \hbar^2}{m^2 r^2} \right) = \frac{1}{4\pi\epsilon_0} \frac{Ze^2}{r^2}$$

$$\text{or } r = \frac{(4\pi\epsilon_0) \hbar^2 n^2}{Zme^2} \quad \dots(5)$$

$$\therefore \text{Radius of the } n^{\text{th}} \text{ permissible orbit for hydrogen} = r_n = \frac{4\pi\epsilon_0 \hbar^2}{me^2} n^2 \quad (\because Z = 1).$$

$$\therefore r_n = a_0 n^2.$$

Here the *Bohr radius*  $a_0$  is defined as

$$a_0 = \frac{4\pi\epsilon_0 \hbar^2}{me^2} = \frac{4\pi (8.854 \times 10^{-12}) (1.055 \times 10^{-34})^2}{(9.11 \times 10^{-31}) (1.6 \times 10^{-19})^2} \text{ m} = 0.053 \text{ nm.}$$

**Calculation of total energy.** The total energy of the electron in any orbit is the sum of its kinetic and potential energies. The potential energy of the electron is considered to be zero when it is at an infinite distance from the nucleus. P.E. of an electron in an orbit is given by the work done in bringing the electron from infinity to that orbit. This amount of work is obtained by integrating the electrostatic force of attraction between the nucleus and the electron from the limits  $\infty$  to  $r$ .

$$\text{P.E. of the electron} = \int_{\infty}^r \frac{Ze^2}{4\pi\epsilon_0 r^2} dr = \frac{-Ze^2}{4\pi\epsilon_0 r} \quad \dots(6)$$

$$\text{K.E. of the electron} = \frac{1}{2} mv^2 = \frac{Ze^2}{8\pi\epsilon_0 r} \quad \dots(7)$$

$$\left[ \because \text{from (3), } mv^2 = \frac{Ze^2}{4\pi\epsilon_0 r} \right]$$

$$\therefore \text{Total energy of the electron in the } n^{\text{th}} \text{ orbit} = E_n = \text{P.E.} + \text{K.E.}$$

$$= \frac{-Ze^2}{4\pi\epsilon_0 r} + \frac{Ze^2}{8\pi\epsilon_0 r} = \frac{-Ze^2}{8\pi\epsilon_0 r}$$

Substituting the value of  $r$  from (5),

$$E_n = \frac{-me^4 Z^2}{32\pi^2 \epsilon_0^2 \hbar^2} \frac{1}{n^2} \quad \dots(8)$$

As the value of  $n$  increases,  $E_n$  increases. Hence, the outer orbits have greater energies than the inner orbits.

**Bohr's interpretation of the Hydrogen spectrum.**

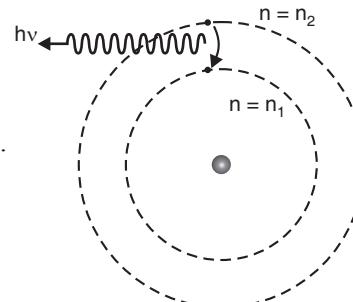
If an electron jumps from an outer initial orbit  $n_2$  of higher energy to an inner orbit  $n_1$  of lower energy, as in Fig. 4.8, the frequency of the radiation emitted is given by  $\nu = (E_{n_2} - E_{n_1}) / h$ .

$$E_{n_2} = -\frac{me^4}{32\pi^2 \epsilon_0^2 \hbar^2} \frac{1}{n_2^2}$$

and  $E_{n_1} = -\frac{me^4}{32\pi^2 \epsilon_0^2 \hbar^2} \frac{1}{n_1^2}$  [ $Z = 1$  for hydrogen].

$$\therefore \nu = \frac{E_{n_2} - E_{n_1}}{h}$$

$$\nu = \frac{me^4}{64\pi^3 \epsilon_0^2 \hbar^3} \left\{ \frac{1}{n_1^2} - \frac{1}{n_2^2} \right\} \quad \dots(9)$$

**Fig. 4.8**

The wavenumber  $\bar{\nu}$  of a radiation is defined as the reciprocal of its wavelength  $\lambda$  in vacuum and gives the number of waves contained in unit length in vacuum.  $\bar{\nu} = 1/\lambda = \nu/c$ .

$$\therefore \bar{\nu} = \frac{me^4}{64\pi^3 \epsilon_0^2 \hbar^3 c} \left( \frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \quad \dots(10)$$

$$\frac{me^4}{64\pi^3 \epsilon_0^2 \hbar^3 c} = R_\infty = \text{Rydberg constant for infinite nuclear mass} = 1.097 \times 10^7 \text{ m}^{-1}$$

$$\therefore \bar{\nu} = R_\infty \left( \frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \quad \dots(11)$$

**Spectral series of hydrogen atom.****(1) Lyman series.**

When an electron jumps from second, third, ... etc., orbits to the first orbit, we get the Lyman series which lies in the ultraviolet region. Here,  $n_1 = 1$  and  $n_2 = 2, 3, 4, 5, \dots$

$$\therefore \bar{\nu} = R_\infty \left( \frac{1}{1^2} - \frac{1}{n^2} \right); \quad n = 2, 3, 4, 5, \dots$$

**(2) Balmer series.**

When an electron jumps from outer orbits to the second orbit,  $n_1 = 2$  and  $n_2 = 3, 4, 5, \dots$  etc.

$$\bar{\nu} = R_\infty \left( \frac{1}{2^2} - \frac{1}{n^2} \right); \quad n = 3, 4, 5, \dots$$

This series is called Balmer series and lies in the visible region of the spectrum. The first line in the series ( $n = 3$ ) is called the  $H_\alpha$  line, the second ( $n = 4$ ) the  $H_\beta$  line and so on.

**(3) Paschen series.** Paschen series in the infrared region are given by  $n_1 = 3$  and  $n_2 = 4, 5, 6, \dots$  etc.

$$\therefore \bar{\nu} = R_\infty \left( \frac{1}{3^2} - \frac{1}{n^2} \right); \quad n = 4, 5, 6, 7, \dots$$

**(4) Brackett series.** If  $n_1 = 4$  and  $n_2 = 5, 6, 7, \dots$  etc., we get the Brackett series.

$$\therefore \bar{\nu} = R_\infty \left( \frac{1}{4^2} - \frac{1}{n^2} \right); \quad n = 5, 6, 7, \dots$$

(5) **Pfund series.** If  $n_1 = 5$  and  $n_2 = 6, 7, 8, \dots$  we get Pfund series.

$$\bar{v} = R_\infty \left( \frac{1}{5^2} - \frac{1}{n^2} \right); \quad n = 6, 7, 8, \dots$$

Brackett and Pfund series lie in the very far infrared region of the hydrogen spectrum. By putting  $n = \infty$  in each one of the series, we get the wavenumber of the *series limit*, i.e., the last line in the series. The electron jumps giving rise to the different series in hydrogen are shown in Fig. 4.9.

**The energy-level diagram.** The equation

$$E_n = -\frac{me^4 Z^2}{32\pi^2 \epsilon_0^2 \hbar^2} \frac{1}{n^2}$$

can be diagrammatically represented. Then, it is called the energy-level diagram. Substituting the values of  $m, e, \epsilon_0$  and  $\hbar$  we get for hydrogen ( $Z = 1$ ),

$$\begin{aligned} \frac{me^4 Z^2}{32\pi^2 \epsilon_0^2 \hbar^2} &= \frac{(9.11 \times 10^{-31})(1.6 \times 10^{-19})^4 (1^2)}{32\pi^2 (8.854 \times 10^{-12})^2 (1.055 \times 10^{-34})^2} = 21.76 \times 10^{-19} \text{ J} \\ &= \frac{21.76 \times 10^{-19}}{1.6 \times 10^{-19}} \text{ eV} = 13.6 \text{ eV}. \end{aligned}$$

$$\therefore E_n = (-13.6/n^2) \text{ eV} \text{ for } n = 1, 2, 3, \dots$$

The lowest energy level  $E_1$  is called the *normal* or the *ground state* of the atom and the higher energy levels  $E_2, E_3, E_4 \dots$  are called the *excited states*. As  $n$  increases,  $E_n$  increases. As  $n$  increases, the energy levels crowd and tend to form a continuum.

In the energy-level diagram, the discrete energy states are represented by horizontal lines, and the electronic jumps between these states by vertical lines (Fig. 4.10). Fig. 4.10 shows schematically how spectral lines are related to atomic energy levels.

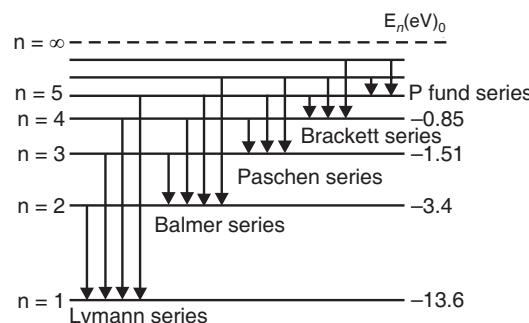


Fig. 4.10

#### 4.4 EFFECT OF NUCLEAR MOTION ON ATOMIC SPECTRA

In Bohr theory, we have assumed that the nucleus remains fixed at the centre of the circular orbit while the electron revolves round it. This is true if the nucleus has infinite mass. If the nucleus has a mass  $M$  that is not infinite, both the nucleus and its orbital electron revolve around a common centre of mass with same angular velocity  $\omega$ . Let  $N$  and  $e$  represent (Fig. 4.11) the nucleus and the electron of mass  $M$  and  $m$  respectively. Now both of them are rotating about their common centre of mass  $C$ , the nucleus moving in a circle of radius  $r_1$  and the electron in a circle of radius  $r_2$ .

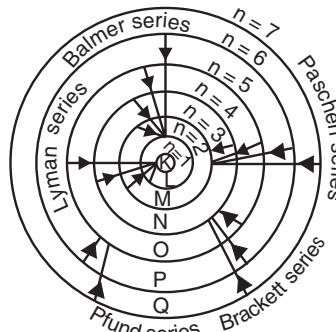


Fig. 4.9

Now according to centre of mass theory,

$$Mr_1 = mr_2 \quad \dots(1)$$

Let  $r$  represent the distance between the nucleus and electron.

Then,

$$r = r_1 + r_2 \quad \dots(2)$$

$\therefore$

$$r_1 = r - r_2 = r - \frac{Mr_1}{m}$$

or

$$r_1 \left(1 + \frac{M}{m}\right) = r$$

or

$$r_1 = \left(\frac{m}{M+m}\right)r \quad \dots(3)$$

Similarly

$$r_2 = \left(\frac{M}{M+m}\right)r \quad \dots(4)$$

The total angular momentum of the atom about the centre of mass is

$$\begin{aligned} L &= Mr_1^2 \omega + mr_2^2 \omega \\ &= M \left(\frac{mr}{M+m}\right)^2 \omega + m \left(\frac{Mr}{M+m}\right)^2 \omega \\ &= \left(\frac{Mm}{M+m}\right) r^2 \omega \\ &= \mu r^2 \omega \end{aligned} \quad \dots(5)$$

Here,  $\mu = \frac{Mm}{M+m} = \frac{m}{1+(m/M)}$  is called the reduced mass of the electron.

According to Bohr's first postulate,

$$\mu r^2 \omega = n \hbar \quad \dots(6)$$

In the absence of nuclear motion, the corresponding equation was

$$mr^2 \omega = n \hbar \quad \dots(7)$$

Eq. (6) is identical to Eq. (7), except that  $m$  has been replaced by  $\mu$ . Therefore to take into account the finite mass of the nucleus we must replace the mass of the electron ( $m$ ) in all the Bohr formulae by the reduced mass ( $\mu$ ).

#### Equation for the wavenumbers of spectral lines of the atom.

The energy-levels of the hydrogen atom become

$$E_n = -\frac{\mu e^4 Z^2}{32\pi^2 \epsilon_0^2 \hbar^2} \left(\frac{1}{n^2}\right) \quad \dots(8)$$

When an electron jumps from the Bohr orbit  $n_2$  to the Bohr orbit  $n_1$ , the frequency of the radiation emitted is

$$\nu = \frac{\mu e^4 Z^2}{64\pi^3 \epsilon_0^2 \hbar^3} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$$

or

$$\bar{\nu} = \frac{\mu e^4 Z^2}{64\pi^3 \epsilon_0^2 c \hbar^3} \left[\frac{1}{n_1^2} - \frac{1}{n_2^2}\right] \quad \dots(9)$$

The Rydberg constant for any element is given by,

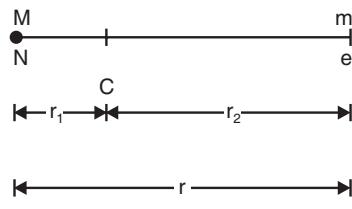


Fig. 4.11

$$\begin{aligned}
 R_Z &= \frac{\mu e^4}{64\pi^3 \epsilon_0^2 c \hbar^3} = \frac{me^4}{64\pi^3 \epsilon_0^2 c \hbar^3} \times \frac{M_z}{(M_z + m)} \\
 &= R_\infty \times \frac{1}{1 + m/M_z}
 \end{aligned} \quad \dots(10)$$

Here,  $R_\infty = \frac{me^4}{64\pi^3 \epsilon_0^2 c \hbar^3}$  = The value of the Rydberg constant when  $M_z = \infty$  i.e., when the nucleus is at rest and  $M_z$  = Mass of the nucleus of the element of atomic number  $Z$ .

The constant  $R_\infty$  can be estimated from the spectroscopically observed value of the Rydberg constant for hydrogen  $R_H = 10967770 \text{ m}^{-1}$  and the known value of  $m/M_H = 1/1840$  using the relation  $R_H = R_\infty \{1/(1 + m/M_H)\}$ .

The value of  $R_\infty$  thus found is  $10973740 \text{ m}^{-1}$ .

Equation (9) can be written as

$$\bar{v} = Z^2 R_Z \left( \frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \quad \dots(11)$$

## 4.5 EVIDENCES IN FAVOUR OF BOHR'S THEORY

### (1) The ratio of mass of electron to the mass of proton.

The Rydberg constants for hydrogen and helium are

$$\begin{aligned}
 R_H &= \frac{R_\infty}{1 + m/M_H} \text{ and } R_{He} = \frac{R_\infty}{1 + m/M_{He}} \\
 \therefore \frac{R_{He}}{R_H} &= \frac{1 + m/M_H}{1 + m/M_{He}}
 \end{aligned}$$

But

$$\begin{aligned}
 M_{He} &\approx 4M_H \\
 \therefore \frac{R_{He}}{R_H} &= \frac{1 + m/M_H}{1 + m/4M_H}
 \end{aligned}$$

$$\text{or } R_{He} + \frac{R_{He}}{4} \frac{m}{M_H} = R_H + R_H \frac{m}{M_H}$$

$$\text{or } R_{He} - R_H = \frac{m}{M_H} \left( R_H - \frac{R_{He}}{4} \right)$$

$$\therefore \frac{m}{M_H} = \frac{R_{He} - R_H}{R_H - R_{He}/4}$$

But from spectroscopic data,  $R_{He} = 10972240 \text{ m}^{-1}$  and  $R_H = 10967770 \text{ m}^{-1}$

$$\therefore \frac{m}{M_H} \approx \frac{1}{1837}.$$

This value is in excellent agreement with the value obtained by other methods.

### (2) Spectrum of singly ionised helium.

Singly ionised helium  $\text{He}^+$  (a helium atom which has lost a single electron) resembles a hydrogen atom, except that  $Z = 2$  and the nucleus is nearly four times as heavy. Putting  $Z = 2$  in Eqn. (11),

$$\bar{v} = 4R_H \left( \frac{1}{n_1^2} - \frac{1}{n_2^2} \right).$$



**Niels Bohr**  
**(1885 – 1962)**

Thus we see that singly-ionised helium should emit the same type of atomic spectrum as hydrogen, except that all wavenumbers are approximately four times as large. This conclusion from the theory agrees with observation except for a slight numerical discrepancy due to the fact that the reduced mass  $\mu$  is slightly greater for  $He$  than for  $H$ .

(3) **The discovery of deuterium.** Deuterium is an isotope of hydrogen whose atomic mass is double that of ordinary hydrogen owing to the presence of a neutron as well as a proton in the nucleus. According to theory, for atoms of the same value of  $Z$ , there should be lines of slightly different wavenumbers if their nuclei have different masses. Because of the greater nuclear mass, the spectral lines of deuterium are all shifted slightly to wavelengths shorter than those of ordinary hydrogen. For example,  $H_\alpha$  line of deuterium has a wavelength of 6561 Å, while that of hydrogen is 6563 Å. It was found by Urey and his co-workers that in the atomic spectrum of natural hydrogen, lines due to deuterium were slightly displaced from the corresponding lines due to hydrogen proper. It was also observed by them that the intensity of lines which are slightly shifted towards the short wavelength side is extremely less than the corresponding hydrogen lines. This is because the concentration of deuterium in natural hydrogen is only one atom in 5000.

#### 4.5.1. Ritz Combination Principle

**Statement.** By a combination of the terms that occur in the Rydberg or Balmer formula, other relations can be obtained holding good for new lines and new series.

By this principle, Ritz predicted new series of lines in the hydrogen spectrum before they were actually discovered.

**EXAMPLE.** Consider the first two lines,  $H_\alpha$  and  $H_\beta$  of the Balmer series. We may represent them by

$$\bar{v}_\alpha = R \left( \frac{1}{2^2} - \frac{1}{3^2} \right) \quad \text{and} \quad \bar{v}_\beta = R \left( \frac{1}{2^2} - \frac{1}{4^2} \right)$$

Combining these two formulae,

$$\bar{v}_\beta - \bar{v}_\alpha = R \left( \frac{1}{3^2} - \frac{1}{4^2} \right).$$

This represents a new line. It is the first line of a new series in the infrared, discovered by Paschen. Similarly, the second line of the Paschen series is obtained by forming the difference of  $H_\gamma$  and  $H_\alpha$  and so on.

Ritz combination principle may also be stated as follows:

If lines at frequencies  $v_{ij}$  and  $v_{jk}$  exist in a spectrum with  $j > i$  and  $k > j$ , then there will usually be a line at  $v_{ik}$  where  $v_{ik} = v_{ij} + v_{jk}$ .

However, not all combinations of frequencies are observed because certain selection rules operate which will be discussed later.

**EXAMPLE.** If lines of frequencies  $v_{12}$  and  $v_{23}$  can be represented as

$$v_{12} = T_1 - T_2; \quad v_{23} = T_2 - T_3$$

then a line of frequency  $v_{13}$  will exist, where

$$v_{13} = (T_1 - T_2) + (T_2 - T_3) = T_1 - T_3$$

#### 4.6 BOHR CORRESPONDENCE PRINCIPLE

**Statement.** Any new theory in Physics must reduce to well-established corresponding classical theory when the new theory is applied to the special situation in which the less general theory is known to be valid.

- Quantum theory must agree with classical theory in the limit in which classical theory is known to agree with experiment.
- Quantum theory must agree with classical theory in the limit of large quantum numbers.

**Explanation.** Bohr's theory gives only the frequencies of the spectral lines and says nothing about the nature (whether polarised or not) and intensity of lines, whereas classical theory is very successful in this respect. Also, according to classical theory, the frequency of the spectral line is the same as the orbital frequency of the electron ( $\nu = \omega/2\pi$ ). But in Bohr's theory, the frequency of the spectral line is determined by the difference in energy between two orbital states:  $\nu = (E_i - E_f)/h$ . But it can be shown that, for transitions between states whose quantum numbers are relatively high, the frequency of the spectral line coincides very nearly with the orbital frequency.

Let us consider an atom of effectively infinite mass. Then

$$E_n = \frac{-me^4}{32\pi^2 \epsilon_0^2 \hbar^2 n^2} \quad \dots(1)$$

If  $n$  is sufficiently great, the energy change  $\Delta E$  corresponding to a change of  $n$  by  $\Delta n$  is obtained by differentiating equation (1).

$$\Delta E_n = \frac{me^4}{16\pi^2 \epsilon_0^2 \hbar^2 n^3} \Delta n \quad \dots(2)$$

$$\therefore \nu = \frac{\Delta E_n}{h} = \frac{me^4}{32\pi^3 \epsilon_0^2 \hbar^3 n^3} \Delta n \quad \dots(3)$$

According to Bohr's first postulate,

$$\begin{aligned} n\hbar &= mr^2 \omega \\ \therefore \nu &= \frac{me^4}{32\pi^3 \epsilon_0^2 (mr^2 \omega)^3} \Delta n \\ &= \frac{me^4}{32\pi^3 \epsilon_0^2 m^3 r^6 \omega^3} \Delta n \end{aligned} \quad \dots(4)$$

Now for the equilibrium in the orbit, we have,

$$\begin{aligned} mr\omega^2 &= \frac{1}{4\pi\epsilon_0} \frac{e^2}{r^2} \quad \text{or} \quad \frac{1}{r^3} = \frac{4\pi\epsilon_0 m \omega^2}{e^2} \\ \therefore \nu &= \frac{me^4}{32\pi^3 \epsilon_0^2 m^3 \omega^3} \left( \frac{4\pi\epsilon_0 m \omega^2}{e^2} \right)^2 \Delta n \\ \text{or} \quad \nu &= \frac{\omega}{2\pi} \Delta n \end{aligned} \quad \dots(5)$$

$$\text{Now, if } \Delta n = 1, \nu = \frac{\omega}{2\pi}.$$

Thus the frequency given by the quantum theory for two very large quantum numbers and separated by unity becomes identical with the orbital frequency and hence with the classical frequency. Therefore, we may conclude that the behaviour of the atom tends asymptotically to that expected from the classical theory in the region of large quantum numbers.

This correspondence principle has proved to be of great value in the computation of the intensity, polarisation and coherence of spectral radiation, as well as in the formulation of the selection rules.

## 4.7 CRITICAL POTENTIALS

**Definitions.** (1) **Critical potential.** *The least energy, expressed in electron volts, required to excite a free neutral atom from its ground state to a higher state is called a critical potential of the atom.*

It is usual to distinguish two kinds of critical potentials, viz., excitation potential (radiation potential or resonance potential) and ionisation potential.

(2) **Excitation potential.** *The energy in electron volts required to raise an atom from its normal state into an excited state is called excitation potential of the state.*

(3) **Ionisation potential** of the atom may be defined as the energy required to remove an electron from a given orbit to an infinite distance from the nucleus.

**Example.** We have seen that the energy associated with an electron in the  $n$ th orbit of the hydrogen atom is  $E_n = -13.6/n^2$  volts. Thus the energies of the first, second, third ... orbits are respectively  $-13.6, -3.4, -1.51, \dots$  0 electron volts. The energy required to raise the atom from the ground state ( $n = 1$ ) to the first excited state is  $13.6 - 3.4 = 10.2$  eV. The energy required to raise it to the second excited state is  $(13.6 - 1.51) = 12.09$  eV and so on. The ionisation potential = 13.6 eV. It is clear that 10.2, 12.09 are excitation potentials, while 13.6 volts is the ionisation potential of the hydrogen atom.

## 4.8 ATOMIC EXCITATION

There are two main ways of exciting an atom.

(i) **Excitation by Collision.** One method of exciting the atoms is to bombard them with electrons having enough kinetic energy. To produce a luminous discharge in a rarefied gas, an electric field is established that accelerates electrons until their kinetic energies are sufficient to excite atoms they collide with. The electron gives up all or part of its energy in exciting the atom. The atom then emits a photon in returning to its ground state. Neon signs and mercury vapour lamps are familiar examples of how a strong electric field applied between electrodes in a gas-filled tube leads to the emission of the characteristic spectral radiation of that gas. We get reddish light in the case of neon and bluish light in the case of mercury vapour.

(ii) **Irradiation of atoms with light.** Atoms can be excited by energy supplied in the form of light. An atom absorbs a photon of light whose energy is just the right amount to raise the atom to a higher energy level. For example, a photon of wavelength 121.7 nm is emitted when a hydrogen atom in the  $n = 2$  state drops to the  $n = 1$  state. Absorbing a photon of wavelength 121.7 nm by a hydrogen atom initially in the  $n = 1$  state will therefore bring it up to the  $n = 2$  state. This process explains the origin of absorption spectra.

When white light, which contains all wavelengths, is passed through hydrogen gas, photons of those wavelengths that correspond to transitions between energy levels are absorbed. The resulting excited hydrogen atoms radiate their excitation energy almost at once. But these photons come off in random directions with only a few in the same direction as the original beam of white light. The dark lines in an absorption spectrum are therefore never completely black, but only appear so by contrast with the bright background. We expect the absorption spectrum of any element to be identical with its emission spectrum and this agrees with observations. For example, in the case of sodium, the two black absorption lines occupy exactly the same position in the spectrum as the two yellow  $D_1$  and  $D_2$  emission lines.

## 4.9 EXPERIMENTAL DETERMINATION OF CRITICAL POTENTIALS

### (1) Franck and Hertz's method

**Experimental arrangement.** The gas of the element under study (Franck and Hertz had studied mercury vapour) is filled in a glass tube ( $T$ ) at a pressure of about 1 mm of mercury (Fig. 4.12). Electrons are produced by heating the filament ( $F$ ) by a low tension battery ( $B$ ). These electrons are accelerated towards a grid  $G$  by the potential difference  $V$  between  $F$  and  $G$ .  $V$  can be varied between 0 and  $60V$  by a potentiometer arrangement.  $P$  is the collector plate which is kept at a slightly negative potential ( $V'$ ), about  $0.5 V$ , with respect to  $G$ . Thus, only those electrons from  $G$  can go to  $P$  which have kinetic energies greater than this potential difference. The milliammeter  $A$  measures the plate current.

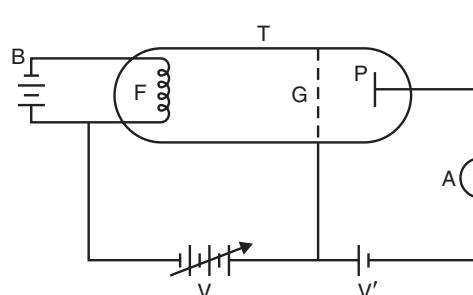


Fig. 4.12

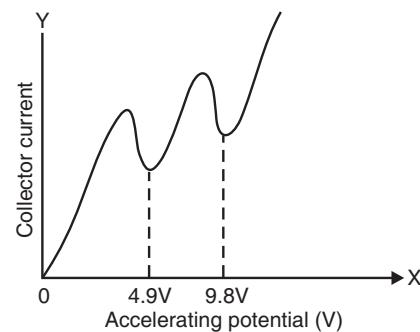


Fig. 4.13

Keeping  $V'$  constant,  $V$  is gradually increased in small steps from zero upwards. A plot of the collector current against the accelerating potential  $V$  is shown in Fig. 4.13. From this graph, we see that there is no collector current for  $V < 0.5 V$ . Above this, the collector current increases continuously. When the accelerating P.D. reaches a value 4.9 volt, the current suddenly dips to a minimum. Again when the P.D. is gradually increased above 4.9 volt, the current gradually increases, till another maximum is reached when the P.D. is just below 9.8 volts. At 9.8 volts, the current again dips steeply to a minimum. Similarly, a significant decrease in the collector current occurs each time the accelerating potential is increased by approximately  $5V$ .

**Explanation of the graph.** Consider the case of a beam of slow electrons travelling through mercury vapour at low pressure. If the electrons suffer no energy loss due to collision with the atoms of the gas, they will reach the collector plate. On increasing the accelerating potential, collector current increases. As  $V$  reaches the value of a critical potential (4.9 volt), an electron acquires  $4.9 eV$  of energy on reaching  $G$ . The electron loses all its energy in an inelastic collision with a mercury atom. Thus the electron is left with no energy to reach  $P$ . Consequently the collector current drops abruptly. This suggests that the atom has absorbed this  $4.9 eV$  energy to raise it from the ground state to a state of higher energy.

This current dipping to a minimum at 4.9 volts does not reach zero because statistically some electrons may succeed in reaching the collector plate, avoiding an inelastic collision with a mercury atom. Such electrons record minimum current on the plate. Then as  $V$  is further increased, the electrons start reaching  $P$  after the inelastic collisions, if they are left with energy to overcome  $V'$ . Hence, the plate current again increases. At  $V = 9.8$  volts, the current again dips to a second minimum. This can be explained if an electron of  $9.8 eV$  energy suffers two consecutive inelastic collisions with different mercury atoms before it reaches the plate. Such an electron excites both the mercury atoms to their first excited states, losing  $4.9 eV$  energy in each collision. This explains the second minimum.

Each time there is an inelastic collision, the mercury atoms will be excited and return to the ground state by the emission of photons. By using spectroscopic techniques, the wavelength of the radiation coming from the tube was found to be  $2536 \text{ \AA}$  corresponding to transitions from the first

excited state to the ground state. Let us see what is the energy of a photon of this wavelength.

$$\begin{aligned} E = h\nu &= \frac{hc}{\lambda} = \frac{(6.63 \times 10^{-34})(3 \times 10^8)}{2536 \times 10^{-10}} \text{ J} \\ &= \frac{(6.63 \times 10^{-34})(3 \times 10^8)}{(2536 \times 10^{-10})(1.6 \times 10^{-19})} \text{ eV} = 4.9 \text{ eV}. \end{aligned}$$

Thus, this experiment shows that the energy lost by the electron in its inelastic collision with the mercury atom reappears as a quantum of energy of wavelength  $hc/E$ . This experiment shows in a very convincing way the existence of discrete energy levels in the mercury atom.

The main drawback of this experiment is that one cannot distinguish between the excitation and ionisation potentials.

## (2) Davis and Goucher's method

**Experimental arrangement.** A glass bulb, containing the gas under study (mercury), has a filament  $F$ , cathode  $C$ , grids  $G_1$  and  $G_2$  and a plate  $P$  (Fig. 4.14).

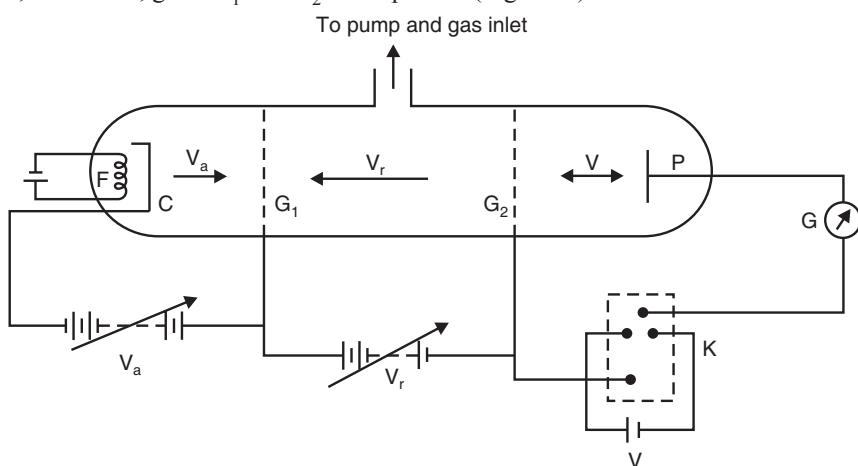


Fig. 4.14

An accelerating potential  $V_a$  is applied between  $C$  and  $G_1$ . The electrons from the filament  $F$  are accelerated by the P.D.  $V_a$  and enter the collision region between the two grids  $G_1$  and  $G_2$ . In the collision region  $G_1 G_2$ , the electrons experience a retarding P.D.  $V_r$  just greater than  $V_a$ . A small P.D.  $V$  is maintained between  $G_2$  and  $P$ . This potential can be reversed with the help of reversing key  $K$ . The grid  $G_2$  and the metal plate  $P$  are made of photosensitive material. The current between  $G_2$  and  $P$  is indicated by a galvanometer  $G$ .

Fig. 4.15 shows a plot of the galvanometer current against the accelerating potentials  $V_a$ .

**Explanation of the graph.** The accelerating potential  $V_a$  is gradually increased from zero. As  $V_r$  is always maintained just greater than  $V_a$ , no electrons are able to reach the plate whatever be the direction of  $V$ . The galvanometer current is, therefore, zero. When  $V_a$  becomes equal to an excitation potential, the electrons excite gas atoms in the region  $G_1 G_2$ . In the case of mercury, for example, this occurs when  $V_a$  approaches  $4.9V$ . The mercury atoms excited by inelastic collisions with the electrons of energy  $4.9 \text{ eV}$ , return to the ground state with the emission of spectral radiation of wavelength  $2536 \text{ \AA}$ . This radiation falling on  $G_2$  and  $P$  releases photoelectrons. Let the reversing key be such that  $G_2$  is negative with respect to  $P$ . The photoelectrons released from  $P$  cannot move towards  $G_2$ , while the photoelectrons released from  $G_2$  move towards  $P$ . Therefore, the galvanometer indicates a current. This is the state corresponding to the point  $s$  on the curve.

When  $V_a$  is increased beyond  $4.9V$ , due to deeper penetration of the accelerated electrons into the collision region, more and more atoms in this region get excited. Consequently, the photoelectric current increases. If the potential between  $G_2$  and  $P$  is reversed and the experiment repeated, photoelectrons can move from  $P$  to  $G_2$  and not from  $G_2$  to  $P$ . This results in a current in the galvanometer in a direction opposite to that previously obtained. When  $V_a$  reaches 6.7 volts, there is a rapid increase in the photoelectric current. This is due to the emission of the spectral radiation of wavelength 1849 Å by the mercury atoms excited by inelastic collisions with electrons. As before, this photoelectric current is reversed on reversing  $V$ .

In Fig. 4.15 the curve  $sqr$  above the  $X$ -axis is obtained when  $G_2$  is positive, while the curve  $sq'r'$  below the  $X$ -axis when  $G_2$  is negative with respect to  $P$ . The next abrupt change in the galvanometer current occurs when  $V_a$  approaches 10.4 V. The current however, now does not get reversed as before when  $V$  is reversed. This can be explained if we take 10.4 V as an ionisation potential of mercury. Inelastic collisions of the electrons of energy 10.4 eV with the mercury atoms in the collision region produce positive ions. These positive ions will be accelerated by the large P.D. =  $V_r$ , and will reach the plate, whatever be direction of the comparatively small P.D. =  $V$ . The current recorded by  $G$  is now due to positive ions. This ionisation current cannot be reversed by reversing the direction of  $V$ . Hence 4.9 and 6.7 volts are excitation potentials, while 10.4 volts is the ionisation potential of mercury. This method thus enables one to distinguish between excitation and ionisation potentials.

This method is subject to a systematic error on account of the initial velocity of emission of the electrons from the hot filament. The actual energy of the electrons will therefore be greater than that corresponding to the accelerating potential. The correction for this error can be determined by measuring known critical potentials with the apparatus used.

#### 4.10 DRAWBACKS OF BOHR ATOM MODEL

- (1) In Bohr theory, we have two rival theories, viz., classical and quantum. The equilibrium is governed by the classical laws, while the emission of radiation is explained by quantum rules.
- (2) The theory does not give any information about the relative intensities of different lines.
- (3) Bohr's theory could not explain the *fine structure of the spectral lines*.
- (4) With the help of Bohr theory, it is difficult to treat dynamical problems of atoms containing more than one valence electron.
- (5) Bohr model could not explain how the orbital electrons in an atom are distributed around the nucleus.
- (6) Bohr's theory could not explain Zeeman Effect and Stark Effect in which the spectral lines could be split up under the influence of magnetic and electric fields.

#### 4.11 SOMMERFELD'S RELATIVISTIC ATOM MODEL

**Introduction.** According to Bohr, the lines in the hydrogen spectrum should each have a well-defined wavelength. Spectrographs of high resolving power showed that the  $H_\alpha$ ,  $H_\beta$ , and  $H_\gamma$  lines in the hydrogen spectrum are not single. Each spectral line actually consisted of several very close

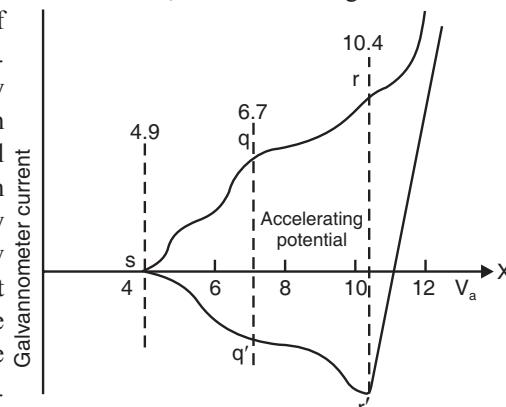


Fig. 4.15

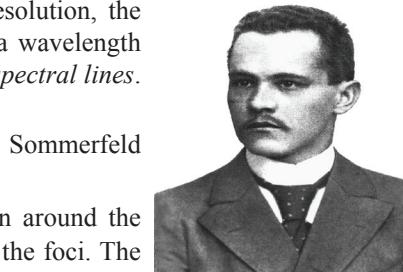
lines packed together. Michelson found that under high resolution, the  $H_{\alpha}$  line can be resolved into two close components, with a wavelength separation of 0.13 Å. This is called the *fine structure of the spectral lines*. Bohr's theory could not explain this fine structure.

To explain the observed fine structure of spectral lines, Sommerfeld introduced two main modifications in Bohr's theory.

(1) According to Sommerfeld, the path of an electron around the nucleus, in general, is an ellipse with the nucleus at one of the foci. The circular orbits of Bohr are a special case of this.

(2) The velocity of the electron moving in an elliptical orbit varies considerably at different parts of the orbit. This causes relativistic variation in the mass of the moving electron. Therefore he took into account the relativistic variation of the mass of the electron with velocity. Hence this is known as the relativistic atom-model. Let us first analyse mathematically Sommerfeld's theory.

**Elliptic orbits for hydrogen.** Consider the electron moving in an elliptical orbit round the nucleus ( $N$ ). Its position at any instant can be fixed in terms of polar coordinates  $r$  and  $\phi$  (Fig. 4.16). In the case of circular orbits, there is only one coordinate that varies periodically, namely, the angle  $\phi$  that the radius vector makes with the  $X$ -axis. In the case of elliptic motion, not only does the angle  $\phi$  vary but the length of the radius vector  $r$  also varies periodically, as shown in Fig. 4.16. We have now to quantise the momenta associated with both these coordinates ( $\phi$  and  $r$ ) in accordance with Bohr's quantum condition. The two quantisation conditions are



*Arnold Sommerfeld  
(1868 - 1951)*

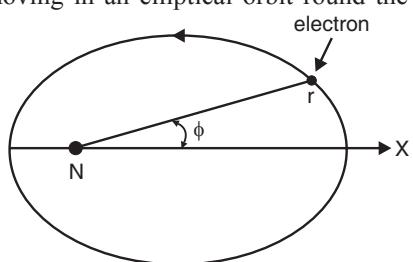


Fig. 4.16

$$\int p_\phi d\phi = n_\phi h \quad \dots(1)$$

$$\text{and} \quad \int p_r dr = n_r h \quad \dots(2)$$

$n_\phi$  is called the *angular* or *azimuthal* quantum number and  $n_r$  is called the *radial* quantum number.  $n_\phi$  and  $n_r$  take only positive integral values and  $n_\phi + n_r = n$ , where  $n$  is the principal quantum number. It can take the integral values 1, 2, 3, ... etc.

To determine the allowed elliptical orbits, we have to evaluate the integrals in equation (1) and (2). The momentum  $p_\phi$  corresponding to the coordinate  $\phi$  is merely the angular momentum  $p$  of the electron in the elliptic orbit and this, from Kepler's law, is a constant, i.e.,  $p_\phi = p = \text{constant}$ . Integrating equation (1) from 0 to  $2\pi$ ,

$$\int_0^{2\pi} p_\phi d\phi = n_\phi h$$

or

$$p_\phi = \frac{n_\phi h}{2\pi} \quad \dots(3)$$

**Evaluation of  $\oint p_r dr = n_r h$**

Now, momentum along the radius  $= p_r = m \frac{dr}{dt}$

$$\text{Now } p_r dr = m \frac{dr}{dt} dr = m \left( \frac{dr}{d\phi} \frac{d\phi}{dt} \right) \frac{dr}{d\phi} d\phi = m \left( \frac{dr}{d\phi} \right)^2 \frac{d\phi}{dt} d\phi$$

But

$$p_\phi = mr^2 \frac{d\phi}{dt}$$

$$\therefore p_r dr = \left( \frac{1}{r} \frac{dr}{d\phi} \right)^2 p_\phi d\phi \quad \dots(4)$$

The equation of an ellipse in polar coordinates is

$$\frac{1}{r} = \frac{1+\varepsilon \cos \phi}{a(1-\varepsilon^2)} \quad \dots(5)$$

where  $a$  is the semimajor axis and  $\varepsilon$  is the eccentricity.

Taking logarithm of Eqn. (5) and differentiating,

$$\frac{1}{r} \frac{dr}{d\phi} = \frac{\varepsilon \sin \phi}{1+\varepsilon \cos \phi} \quad \dots(6)$$

$$\therefore p_r dr = \frac{\varepsilon^2 \sin^2 \phi}{(1+\varepsilon \cos \phi)^2} p_\phi d\phi \quad \dots(7)$$

$$\text{Eqn. (2) now becomes } p_\phi \int_0^{2\pi} \frac{\varepsilon^2 \sin^2 \phi}{(1+\varepsilon \cos \phi)^2} d\phi = n_r h \quad \dots(8)$$

The integral  $I = \int_0^{2\pi} \frac{\varepsilon^2 \sin^2 \phi}{(1+\varepsilon \cos \phi)^2} d\phi$  can be integrated by parts.

$$\int u dv = uv - \int v du .$$

Let  $u = \varepsilon \sin \phi$ ,  $du = \varepsilon \cos \phi d\phi$  and

$$dv = \frac{\varepsilon \sin \phi}{(1+\varepsilon \cos \phi)^2} d\phi \quad \text{so that } v = \frac{1}{1+\varepsilon \cos \phi}$$

$$\text{Then } I = \left[ \frac{\varepsilon \sin \phi}{1+\varepsilon \cos \phi} \right]_0^{2\pi} - \int_0^{2\pi} \frac{\varepsilon \cos \phi}{1+\varepsilon \cos \phi} d\phi .$$

On substitution of the limits of integration, the first term on the R.H.S. becomes zero. Then

$$\begin{aligned} I &= - \int_0^{2\pi} \frac{\varepsilon \cos \phi}{1+\varepsilon \cos \phi} d\phi \\ &= \int_0^{2\pi} \left( \frac{1}{1+\varepsilon \cos \phi} - 1 \right) d\phi \end{aligned}$$

Now  $\int_0^{2\pi} \frac{d\phi}{1+\varepsilon \cos \phi}$  is a standard integral whose value works out to be  $2\pi / (1-\varepsilon^2)^{1/2}$ .

$$\therefore I = \frac{2\pi}{(1-\varepsilon^2)^{1/2}} - 2\pi \quad \dots(9)$$

Now, Eqn. (8) can be written as

$$\frac{2\pi p_\phi}{(1-\varepsilon^2)^{1/2}} - 2\pi p_\phi = n_r h$$

$$\text{or } \frac{n_\phi h}{(1-\varepsilon^2)^{1/2}} - n_\phi h = n_r h \left( \text{since } p_\phi = n_\phi \frac{h}{2\pi} \right). \quad \dots(10)$$

$$\therefore n_r = \frac{n_\phi}{(1-\varepsilon^2)^{1/2}} - n_\phi \quad \text{or} \quad n_r + n_\phi = \frac{n_\phi}{(1-\varepsilon^2)^{1/2}}$$

But,  $n_r + n_\phi = n$ , the principal quantum number.

$$\text{Hence, } n = \frac{n_\phi}{(1-\varepsilon^2)^{1/2}} \quad \text{or} \quad 1-\varepsilon^2 = \frac{n_\phi^2}{n^2} \quad \dots(11)$$

For an ellipse  $1-\varepsilon^2 = b^2/a^2$  where  $a$  and  $b$  are the semi-major and semi-minor axes respectively.

$$\begin{aligned} \text{Hence, } \frac{b^2}{a^2} &= \frac{n_\phi^2}{n^2} \\ \text{or } \frac{b}{a} &= \frac{n_\phi}{n} \end{aligned} \quad \dots(12)$$

Equation (11) is the condition that determines the allowed elliptical orbits. The allowed elliptical orbits are those for which the ratio of major and minor axes is that of two integers. When  $n_\phi = n$ ,  $b = a$ ,  $\varepsilon = 0$  and the orbit becomes circular.  $n_\phi$  cannot be zero, since the ellipse would then degenerate into a straight line passing through the nucleus. Also  $n_\phi$  cannot be greater than  $n$ , since  $b$  is always less than  $a$ . Hence for a given value of  $n$ ,  $n_\phi$  can assume only  $n$  different values, which means there can be only  $n$  elliptical orbits of different eccentricities.

**Example.** For the first orbit,  $n = 1$ . Since  $n_r + n_\phi = 1$  and  $n_\phi \neq 0$ ,  $n_\phi = 1$ . With  $n = n_\phi = 1$ , the first orbit is a circle. With  $n = 2$ ,  $n_\phi$  may have the values 1 or 2, so that there are two possible orbits for  $n = 2$ , a circle and an ellipse. Similarly, there are three possible orbits for  $n = 3$ , a circle and two ellipses (Fig. 4.17).

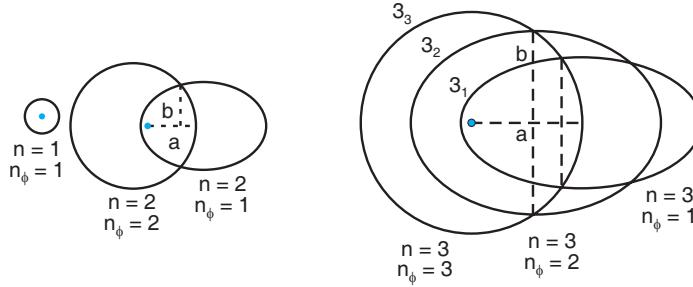


Fig. 4.17

Usually the allowed orbits are described by giving values of  $n$  and  $n_\phi$ . The three orbits for  $n = 3$  are represented by symbols  $3_3$ ,  $3_2$  and  $3_1$ , the subscript being the azimuthal quantum number ( $n_\phi$ ).

**Another notation.** In this notation, the value of azimuthal quantum number  $n_\phi$  is described by the letters,  $s, p, d, f$ , etc. The value of  $n_\phi$  corresponding to these letters is 1, 2, 3, 4 etc., respectively. In this notation, the orbit determined by  $n = 3$  and  $n_\phi = 1$  is represented by  $3s$ . Similarly  $4d$  will represent the orbit  $n = 4$  and  $n_\phi = 3$ .

**Total energy.** The total energy  $E_n$  of a single electron =  $E_n = P.E. + K.E.$

$$\text{The P.E. is } \frac{-Ze^2}{4\pi\varepsilon_0 r}.$$

$$\text{The K.E. can be written as } \frac{1}{2}m \left[ \left( \frac{dr}{dt} \right)^2 + \left( r \frac{d\phi}{dt} \right)^2 \right]$$

$\frac{dr}{dt}$  = radial component of the velocity and  $r \left( \frac{d\phi}{dt} \right)$  = transverse component of the velocity.

$$\therefore E_n = \frac{-Ze^2}{4\pi\varepsilon_0 r} + \frac{1}{2}m \left[ \left( \frac{dr}{dt} \right)^2 + \left( r \frac{d\phi}{dt} \right)^2 \right] \quad \dots(13)$$

It can be shown that

$$E_n = - \left( \frac{me^4 Z^2}{8\epsilon_0^2 h^2} \right) \frac{1}{n^2} \quad \dots(14)$$

We find that the expression for the total energy is the same as that obtained by Bohr. This means that the theory of elliptical orbits introduces no new energy levels, other than those given by Bohr's theory of circular orbits. No new spectral lines, which would explain the fine structure, are to be expected because of this *multiplicity* of orbits. Hence, Sommerfeld proceeded further to find a solution to the problem of fine structure of spectral lines, on the basis of the variation of the mass of the electron with velocity.

**Sommerfeld's relativistic theory.** The velocity of the electron in a circular orbit is constant. But the velocity of the electron in an elliptical orbit varies, being a maximum at the perihelion and a minimum at the aphelion. Furthermore this velocity is quite large ( $\approx \frac{1}{137}c$ ). Therefore, Sommerfeld modified his theory, taking into account variation of the mass of the electron with velocity. He showed that the relativistic equation describing the path of an electron is

$$\frac{1}{r} = \frac{1 + \epsilon \cos \psi \phi}{a(1 - \epsilon^2)} \quad \dots(15)$$

where

$$\psi^2 = 1 - \frac{Z^2 e^4}{16\pi^2 \epsilon_0^2 p^2 c^2}.$$

The path of the electron given by equation (15) is a *rosette* — a *precessing ellipse*, i.e., an ellipse whose major axis precesses slowly in the plane of the ellipse about an axis through the nucleus (Fig. 4.18).

It can be shown that the total energy  $E_n$  in the relativistic theory is,

$$E_n = \frac{me^4 Z^2}{8\epsilon_0^2 h^2 n^2} - \frac{me^4 Z^4 \alpha^2}{8\epsilon_0^2 h^2} \left( \frac{n}{n_\phi} - \frac{3}{4} \right) \frac{1}{n^4} \quad \dots(16)$$

$$\text{where } \alpha = \frac{e^2}{2\epsilon_0 c h} \approx \frac{1}{137}.$$

$\alpha$  is a dimensionless quantity and is called the *fine structure constant*. The first term in equation (16) is the energy of the electron in the orbit with the principal quantum number  $n$ . The second term is the *Sommerfeld relativity correction*. This term shows that the energy does depend on the azimuthal quantum number  $n_\phi$ . This results in a splitting of the energy levels of the atom, for a given value of  $n$ , into  $n$  components, corresponding to the  $n$  permitted values of  $n_\phi$ .

#### Fine structure of the $H_\alpha$ line.

$H_\alpha$  line is due to the transition from  $n = 3$  state to  $n = 2$  state of the hydrogen atom.

For  $n = 3$ , there are three possible energy levels corresponding to the three values of  $n_\phi$ , 1, 2, and 3. Similarly there are two possible energy levels for  $n = 2$ . Theoretically, six transitions are possible :  $3_3 \rightarrow 2_2$ ;  $3_3 \rightarrow 2_1$ ;  $3_2 \rightarrow 2_2$ ;  $3_2 \rightarrow 2_1$ ;  $3_1 \rightarrow 2_2$ ;  $3_1 \rightarrow 2_1$ , and they are shown in Fig. 4.19.

Actually the  $H_\alpha$  line has only *three* components. To make experiment and theory agree, some of the transitions have to be ruled out by some *selection rule*. The selection rule is that  $n_\phi$  can change only by +1 or -1, i.e.,  $\Delta n_\phi = \pm 1$ .

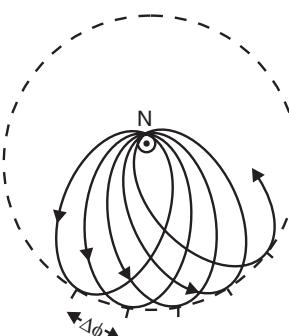


Fig. 4.18

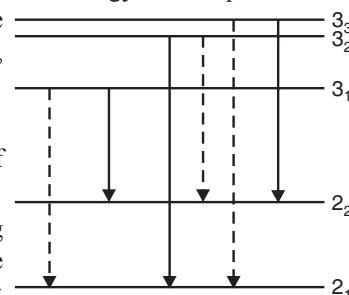


Fig. 4.19

There is no such restriction on  $\Delta n$ . On the basis of this selection rule, there are only three allowed transitions. In Fig. 4.19, the allowed transitions are shown by continuous lines and the forbidden lines by dotted lines.

### 4.12 THE VECTOR ATOM MODEL

**Introduction.** (1) Bohr's theory was able to explain only the series spectra of the simplest hydrogen atom. It could not explain the multiple structure of spectral lines in the simplest hydrogen atom. Sommerfeld's theory was able to give an explanation of the fine structure of the spectral lines of hydrogen. However, Sommerfeld's theory could not predict the correct number of the fine structure lines. Moreover, it gave no information about the relative intensities of the lines. Sommerfeld's theory could not explain the complex spectra of alkali metals like sodium.

(2) These older theories were inadequate to explain new discoveries like Zeeman Effect and Stark Effect in which the spectral lines could be split up under the influence of magnetic and electric fields.

(3) Another drawback of the Bohr model was that it could not explain how the orbital electrons in an atom were distributed around the nucleus.

Therefore, in order to explain the complex spectra of atoms and their relation to atomic structure, the vector atom model was introduced. The two distinct features of the vector atom model are:

1. *the conception of spatial quantisation*, and
2. *the spinning electron hypothesis*.

(1) **Spatial quantisation.** According to Bohr's theory, the orbits are quantised as regards their *magnitude* (*i.e.*, their size and form) only. But according to quantum theory, the *direction* or orientation of the orbits in space also should be quantised. To specify the orientation of the electron orbit in space, we need a fixed reference axis. This reference line is chosen as the direction of an external magnetic field that is applied to the atom. The different permitted orientations of an electron orbit are determined by the fact that *the projections of the quantised orbits on the field direction must themselves be quantised*. The idea of space quantisation leads to an explanation of Zeeman effect. The Stern-Garlach experiment provided an excellent proof of the space quantisation of atom.

(2) **Spinning electron.** To account for the observed fine structure of spectral lines and to explain the anomalous Zeeman effect, the concept of *spinning electron* was introduced by Uhlenbeck and Goudsmit in 1926. According to their hypothesis, the electron spins about an axis of its own, while it also moves round the nucleus of the atom in its orbit (Fig. 4.20). So the electron is endowed with a spin motion over and above the orbital motion. According to the quantum theory, the spin of the electron also should be quantised. Hence a new quantum number called the *spin quantum number* (*s*) is introduced. Since the orbital and spin motions are both quantised in magnitude and direction according to the idea of spatial quantisation, they are considered as *quantised vectors*. Hence the atom model based on these *quantised vectors* is called the "vector atom model", to which vector laws apply.

According to the older theories, the electron was supposed to have only orbital motion round the nucleus. Hence, only the orbital angular momentum and orbital magnetic moment were considered. The spin endows the electron with a spin angular momentum  $s\hbar$  and a spin magnetic moment. Hence the total angular momentum of an atom should be the vector sum of the orbital angular momentum and spin angular momentum. Similarly, the total magnetic moment of an atom should be the vector sum of the orbital and spin magnetic moments.

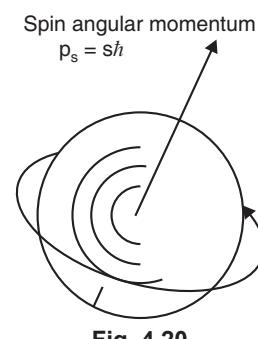


Fig. 4.20

### 4.13 QUANTUM NUMBERS ASSOCIATED WITH THE VECTOR ATOM MODEL

(1) **The principal quantum number ( $n$ ).** This is identical with the one used in Bohr-Sommerfeld theory. The serial number of the shells starting from the innermost is designated as its principal quantum number ( $n$ ). It can take only integral values excluding zero. i.e.,  $n = 1, 2, 3, 4, \dots$

(2) **The orbital quantum number ( $l$ ).** This may take any integral value 0, 1, 2, 3, 4, ...,  $(n - 1)$ . Thus, if  $n = 4$ ,  $l$  can take the four values 0, 1, 2, 3. By convention, an electron for which  $l = 0$ , is called  $s$  electron; if  $l = 1$ ,  $p$  electron;  $l = 2$ ,  $d$  electron;  $l = 3$ ,  $f$  electron etc. The orbital angular momentum  $p_l$  of the electron is given by  $p_l = l\hbar$ .

According to the wave mechanics,  $p_l = [l(l+1)]^{1/2}\hbar$

(3) **The spin quantum number ( $s$ ).** This has only one value,  $s = \frac{1}{2}$ . The spin angular momentum  $p_s = s\hbar$  where  $s = \frac{1}{2}$ . According to wave mechanics,  $p_s = [s(s+1)]^{1/2}\hbar$

(4) **Total angular momentum quantum number ( $j$ ).** It represents the total angular momentum of the electron which is the sum of the orbital angular momentum and spin angular momentum. The vector  $\vec{j}$  is defined by the equation  $\vec{j} = \vec{l} + \vec{s}$  with the restriction that  $\vec{j}$  is positive. The spin angular momentum  $s = \pm \frac{1}{2}$ .  $\therefore j = l \pm s$ , plus sign when  $s$  is parallel to  $l$  and minus sign when  $s$  is antiparallel. Thus for  $l = 2$  and  $s = \frac{1}{2}$ ,  $j$  can have the values  $\frac{5}{2}$  and  $\frac{3}{2}$  (Fig. 4.21).

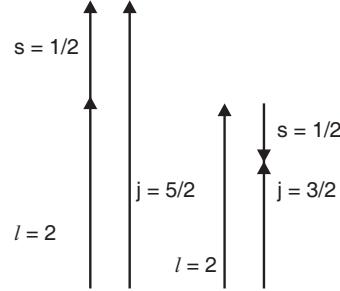


Fig. 4.21

The total angular momentum of the electron =  $p_j = j\hbar$ .

According to wave mechanics,  $p_j = \sqrt{j(j+1)}\hbar$ .

To explain the splitting of spectral lines in a magnetic field, three more magnetic quantum numbers are introduced.

(5) **Magnetic orbital quantum number ( $m_l$ ).** The projection of the orbital quantum number  $l$  on the magnetic field direction is called the magnetic orbital quantum number  $m_l$ . The possible values of  $m_l$  are  $l, l-1, l-2, \dots, 0, -1, -2, \dots, -l$ . i.e., there are  $(2l+1)$  possible values of  $m_l$ . This is illustrated in Fig. 4.22 for  $l = 3$ . The angle  $\theta$  between  $l$  and  $B$  is given by  $\cos \theta = \frac{m_l}{l}$ . Conversely, the permitted orientations of the  $l$  vector relative to the field direction  $B$  is also  $(2l+1)$ . For example, if  $l = 3$ , the permitted orientations of  $l$  are 7, for which  $m_l = 3, 2, 1, 0, -1, -2, -3$ . Hence, the  $l$  vector can take only the 7 directions shown in Fig. 4.22.  $l$  cannot be inclined to  $B$  at any other angle. This is known as 'spatial quantisation'.

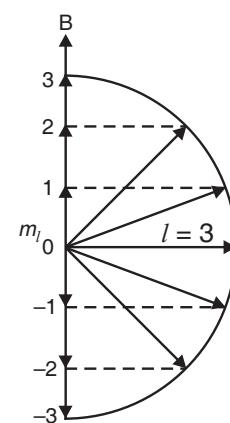


Fig. 4.22

(6) **Magnetic spin quantum number ( $m_s$ )**. This is the projection of the spin vector  $s$  along the direction of the magnetic field. The spin angular momentum ( $s$ ) can assume only two possible positions with respect to the magnetic field: it may be parallel to it or

antiparallel.  $m_s$  can have only two values  $+\frac{1}{2}$  or  $-\frac{1}{2}$ , as illustrated in Fig. 4.23.

(7) **Magnetic total angular momentum quantum number ( $m_j$ )**. This is the projection of total angular momentum vector  $j$  on the direction of the magnetic field. Since we are dealing with a single electron,  $j$  can have only odd half-integral values ( $\therefore j = l \pm \frac{1}{2}$ ). Hence,  $m_j$  must have only odd half-integral values.  $m_j$  can have only  $(2j + 1)$  values, from  $+j$  to  $-j$  zero excluded. Fig. 4.24 shows the possible values of  $m_j$  for  $j = 3/2$ .

#### Notes

- (1) The state of an electron in an atom is completely specified by the four quantum numbers,  $n$ ,  $l$ ,  $m_l$  and  $m_s$ .
- (2) In spectroscopic notation, small letters,  $l$ ,  $s$ ,  $j$  and  $s$ ,  $p$ ,  $d$ ,  $f$ ,  $g$  etc., are used to describe the state of the electron and capital letters,  $L$ ,  $S$ ,  $J$ , and  $S$ ,  $P$ ,  $D$ ,  $F$ ,  $G$ , etc., are used to describe the state of the atom as a whole.

### 4.14 COUPLING SCHEMES

In an atom having two or more electrons, the orbital and spin angular momenta of all its electrons can be added together in two ways. The method of combination depends on the interaction or *coupling* between the orbital and spin angular momenta. Two types of schemes have been developed. They are: (1)  $L-S$  coupling or the Russel-Saunders coupling and (2) the  $j-j$  coupling.

(1)  **$L-S$  coupling**. The type of coupling which occurs most frequently is the  $L-S$  coupling. In this type, all the orbital angular momentum vectors of the various electrons combine to form a resultant  $L$  and independently, all their spin angular momentum vectors combine to form a resultant  $S$ . These resultants  $L$  and  $S$  then combine to form the total angular momentum  $J$  of the atom (Fig. 4.25). This scheme may be summarised as follows:  $L = \sum l_i$ ;  $S = \sum s_i$ ;  $J = L + S$ .  $L$  is always an integer including zero.  $S$  is an integer for an even number of electrons, and odd multiple of  $\frac{1}{2}$  for an odd number of electrons

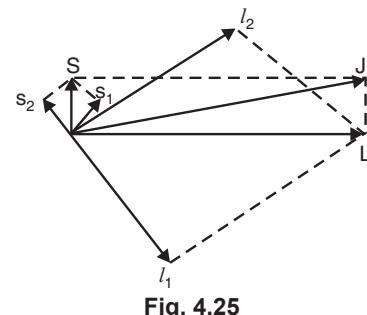


Fig. 4.25

(Fig. 4.26). Hence  $J$  must be an integer, if  $S$  is an integer and  $J$  must be an odd multiple of  $\frac{1}{2}$  if  $S$  is an odd multiple of  $\frac{1}{2}$ . It can be shown that, when  $L > S$ ,  $J$  can have  $(2S + 1)$  values and when  $L < S$ ,  $J$  can have  $(2L + 1)$  values. In particular, if  $L = 0$ ,  $J$  can have only one value namely  $J = S$ .

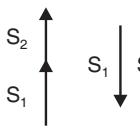
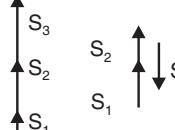
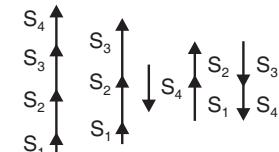
2 Electrons	3 Electrons	4 Electrons
		
$S = 1, 0$	$3/2, 1/2$	$2, 1, 0$

Fig. 4.26

(2) **The j-j coupling.** This method is employed when the interaction between the spin and orbital vectors of each electron is stronger than the interaction between either the spin vectors or the orbital vectors of the different electrons. The orbital and spin angular momenta of each electron in the atom are added to obtain the resultant angular momentum of the electron (Fig. 4.27).

$$\text{Thus } \vec{j}_i = \vec{l}_i + \vec{s}_i.$$

The vector sum of all the individual  $j$  vectors gives the total angular momentum  $J$  of the atom. Thus  $J = \sum j_i$ . This type of coupling exists mainly in heavy atoms.

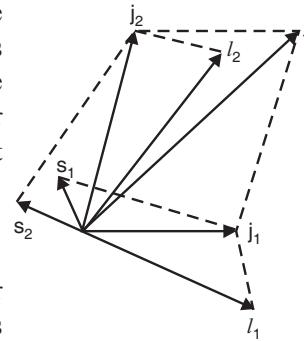


Fig. 4.27

**Application of spatial quantisation.** The resultant vectors  $\mathbf{L}$ ,  $\mathbf{S}$  and  $\mathbf{J}$ , representing the atom, can be obtained by the above coupling schemes. According to quantum theory  $\mathbf{L}$ ,  $\mathbf{S}$  and  $\mathbf{J}$  are quantised in magnitude and direction. Hence the number of permitted orientations of  $\mathbf{L}$ ,  $\mathbf{S}$  and  $\mathbf{J}$  with respect to a given field direction are  $(2L+1)$ ,  $(2S+1)$  and  $(2J+1)$  respectively. The corresponding magnetic quantum numbers  $m_L = \sum m_l$ ,  $m_S = \sum m_s$  and  $m_J = \sum m_j$ , can have only  $(2L+1)$ ,  $(2S+1)$  and  $(2J+1)$  values respectively.

#### Note

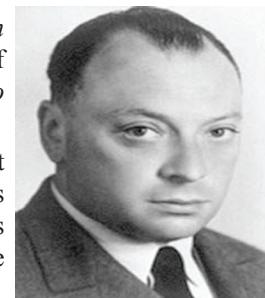
For a one-electron atom with only a single effective electron, the state of the atom as a whole is identical with the state of the electron, i.e.,  $L = l$ ,  $S = s$  and  $J = j$ .

### 4.15 THE PAULI EXCLUSION PRINCIPLE

**Statement.** *No two electrons in an atom exist in the same quantum state.* The four quantum numbers  $n$ ,  $l$ ,  $m_l$  and  $m_s$  determine the state of an electron completely. Hence the principle may be stated as “*No two electrons in an isolated atom may have the same four quantum numbers.*”

**Explanation.** The principle implies that each electron in an atom must have a different set of quantum numbers  $n$ ,  $l$ ,  $m_l$  and  $m_s$ . If two electrons have all their quantum numbers identical, then one of those two electrons would be excluded from entering into the constitution of the atom. Hence the name “*exclusion principle*”.

**Application.** This principle enables us to calculate the number of electrons that can occupy a given subshell.



**Wolfgang Pauli**  
(1900 – 1958)

- Consider the  $K$ -shell with  $n = 1$ ,  $l = 0$  and hence  $m_l = 0$ . Since  $s = \frac{1}{2}$ ,  $m_s$  can be either  $+\frac{1}{2}$

or  $-\frac{1}{2}$ . Hence, the  $K$ -shell can have two electrons: *electron 1 with quantum numbers  $n = 1, l = 0, m_l = 0, m_s = \frac{1}{2}$*  and *electron 2 with quantum numbers  $n = 1, l = 0, m_l = 0, m_s = -\frac{1}{2}$* .

$m_l = 0, m_s = \frac{1}{2}$ ; and electron 2 with quantum numbers  $n = 1, l = 0, m_l = 0, m_s = -\frac{1}{2}$ . If there were

a third electron, its quantum numbers will be identical with those of the first or second electron, which is against Pauli's exclusion principle. The  $K$ -shell is therefore completed or closed with two electrons.

(2) For the  $L$ -shell,  $n = 2$  and  $l = 0$  or  $1$ . For the subshell  $n = 2, l = 0, m_l$  must be zero and  $m_s = +\frac{1}{2}$  or  $-\frac{1}{2}$ . Hence there can be only 2 electrons in this subshell. For the subshell  $n = 2, l = 1, m_l$  can have three values  $-1, 0, +1$ . For each of these three values of  $m_l, m_s$  may be either  $+\frac{1}{2}$  or  $-\frac{1}{2}$ . Hence there will be six possible set of values for the quantum numbers characterizing the

electrons. Therefore, the maximum number of electrons in this subshell is 6. The  $L$ -shell with two subshells [( $n = 2, l = 0$ ) and ( $n = 2, l = 1$ )] is, therefore, completed when it contains  $2 + 6 = 8$  electrons.

(3) For the  $M$ -shell with  $n = 3$ , there can be three subshells with  $l = 0, 1, 2$ . The first and second subshells are completed by 2 and 6 electrons as explained above. The third subshell is completed with  $2(2l+1)$ , i.e., 10 electrons since  $l = 2$ . Hence the total number of electrons required to complete the  $M$ -shell is 18.

(4) Similarly, the  $N$ -shell can have a maximum of 32 electrons.

In general, we get the two following conclusions:

(i) In the  $n^{\text{th}}$  shell there are  $n$  sub-shells corresponding to the values  $0, 1, 2, 3 \dots (n-1)$  of  $l$ .

The maximum number of electrons in a sub-shell with a given value of  $l$  is  $2(2l+1)$ .

Orbital quantum number ( $l$ )	0	1	2	3	4	.....
Number of possible electron states	2	6	10	14	18	.....
Subshell symbol	$s$	$p$	$d$	$f$	$g$	.....

(ii) The number of electrons that can be accommodated in a shell with principal quantum number  $n$  = Sum of the electrons in the constituent  $n$  subshells.

$$\begin{aligned} &= \sum_{l=0}^{l=n-1} 2(2l+1) = 2 \sum_{l=0}^{l=n-1} (2l+1) \\ &= 2 [1 + 3 + 5 + 7 + \dots + \{2(n-1) + 1\}] = 2n^2 \end{aligned}$$

The following table shows the distribution of electrons according to this scheme.

Shell symbol	$K$	$L$	$M$	$N$	$O$
Quantum number ( $n$ )	1	2	3	4	5
Number of electrons ( $2n^2$ )	2	8	18	32	50

The distribution of electrons in the various states (shells and sub-shells) according to the exclusion principle is given in the following table (Table 4.1).

TABLE 4.1

$n$	$l$	$m_l$	$m_s$	Number of electrons in sub-shell with spectroscopic notation	Total number of electrons in shell = $2n^2$
1	0	0	$+\frac{1}{2}, -\frac{1}{2}$	2 $1s^2$	2
2	0	0	$+\frac{1}{2}, -\frac{1}{2}$	2 $2s^2$	8
2	1	-1, 0, +1	$+\frac{1}{2}, -\frac{1}{2}$	6 $2p^6$	
3	0	0	$+\frac{1}{2}, -\frac{1}{2}$	2 $3s^2$	18
3	1	-1, 0, +1	$+\frac{1}{2}, -\frac{1}{2}$	6 $3p^6$	
3	2	-2, -1, 0, +1, +2	$+\frac{1}{2}, -\frac{1}{2}$	10 $3d^{10}$	
4	0	0	$+\frac{1}{2}, -\frac{1}{2}$	2 $4s^2$	32
4	1	-1, 0, +1	$+\frac{1}{2}, -\frac{1}{2}$	6 $4p^6$	
4	2	-2, -1, 0, +1, +2	$+\frac{1}{2}, -\frac{1}{2}$	10 $4d^{10}$	
4	3	-3, -2, -1, 0, +1, +2, +3	$+\frac{1}{2}, -\frac{1}{2}$	14 $4f^{14}$	

## 4.16 THE PERIODIC CLASSIFICATION OF ELEMENTS

**The periodic table.** The periodic table is an arrangement of different elements that exist in nature, based on their chemical properties and atomic numbers. Table 4.2 shows the simplest form of the periodic table.

Elements with similar properties form the *groups* shown as vertical columns in the table. Thus group I consists of hydrogen plus the alkali metals, all of which are extremely active chemically and all of which have valence of +1. Group VII consists of the halogens that have valence of -1. Group VIII consists of the inert gases which are chemically inactive.

The horizontal rows are called *periods*. As we go from left to right in the same period, the chemical and physical properties of the elements vary gradually as the atomic number increases. Since the atomic number gives also the number of electrons in the atom, it follows that the atoms of successive elements in the periodic table are formed by the addition of one more electron at each step.

We have already seen the arrangement of electrons in an atom by applying Pauli's exclusion principle. The notion of electron shells and subshells fits perfectly into the pattern of the periodic table. The total orbital and spin angular momenta of the electrons in a closed subshell are zero. The electrons in a closed shell are all very tightly bound, since the positive nuclear charge is large relative to the negative charge of the inner shielding electrons. Since an atom containing only closed shells has no dipole moment, it does not attract other electrons, and its electrons, cannot be readily

TABLE 4.2

Period	Group		Group		Group		Group		Group		Group		
	I	II	III	IV	V	VI	VII	VIII					
1	1 <b>H</b> 1.00												
2	3 <b>Li</b> 6.94	4 <b>Be</b> 9.01											
3	11 <b>Na</b> 22.99	12 <b>Mg</b> 24.31											
4	19 <b>K</b> 39.10	20 <b>Ca</b> 40.08	21 <b>Sc</b> 44.96	22 <b>Ti</b> 47.90	23 <b>V</b> 50.94	24 <b>Cr</b> 52.00	25 <b>Mn</b> 54.94	26 <b>Fe</b> 55.85	27 <b>Co</b> 58.93	28 <b>Ni</b> 58.71	29 <b>Cu</b> 63.54	30 <b>Zn</b> 65.37	
5	37 <b>Rb</b> 85.47	38 <b>Sr</b> 87.66	39 <b>Y</b> 88.91	40 <b>Zr</b> 91.22	41 <b>Nb</b> 92.91	42 <b>Mo</b> 95.94	43 <b>Tc</b> (99)	44 <b>Ru</b> 101.1	45 <b>Pd</b> 102.91	46 <b>Ag</b> 106.4	47 <b>Cd</b> 107.87	48 <b>In</b> 112.82	49 <b>Sn</b> 114.82
6	55 <b>Cs</b> 132.91	56 <b>Ba</b> 137.34	57-71 * <b>Ba</b> 137.34	72 <b>Hf</b> 178.49	73 <b>Ta</b> 180.95	74 <b>Re</b> 183.85	75 <b>W</b> 186.2	76 <b>Ir</b> 190.0	77 <b>Pt</b> 192.2	78 <b>Au</b> 195.09	79 <b>Hg</b> 197.0	80 <b>Tl</b> 200.59	81 <b>Pb</b> 204.37
7	87 <b>Fr</b> (223)	88 <b>Ra</b> 226.05		89-103 ***									
* Rare earths		57 <b>La</b> 138.91	58 <b>Ce</b> 140.12	59 <b>Pr</b> 140.91	60 <b>Nd</b> 144.24	61 <b>Pm</b> (145)	62 <b>Sm</b> 150.35	63 <b>Eu</b> 152.0	64 <b>Gd</b> 157.25	65 <b>Tb</b> 158.92	66 <b>Dy</b> 162.50	67 <b>Ho</b> 164.92	68 <b>Er</b> 167.26
** Actinides		89 <b>Ac</b> (227)	90 <b>Th</b> 232.04	91 <b>Pa</b> (231)	92 <b>U</b> (231)	93 <b>Np</b> (237)	94 <b>Pu</b> (242)	95 <b>Am</b> (243)	96 <b>Cm</b> (247)	97 <b>Bk</b> (249)	98 <b>Cf</b> (251)	99 <b>Es</b> (254)	100 <b>Fm</b> (253)
													101 <b>Md</b> (256)
													102 <b>No</b> (254)
													103 <b>Lw</b> (257)
													104 <b>Lu</b> 174.97
													105 <b>Yb</b> 173.04

detached. We expect such atoms to be passive chemically, like the inert gases and the inert gases all turn out to have closed shell electron configurations.

Those atoms which have a single electron in their outermost shell, tend to lose this electron. Hydrogen and the alkali metals are in this category and accordingly have valences of +1. Atoms whose outer shells lack a single electron for being closed, tend to acquire such an electron, which accounts for the chemical behaviour of the halogens. It is clear that the chemical and physical properties of an atom are determined by the number and arrangement of the electrons in the *outermost-shell* and not by the total number of electrons in the atom. In this manner the similarities of the members of the various groups of the periodic table may be accounted for.

#### 4.17 SOME EXAMPLES OF ELECTRON CONFIGURATIONS WITH THEIR MODERN SYMBOLIC REPRESENTATIONS

The *electronic configuration* of an atom is the distribution of electrons in various subshells around the nucleus of the atom. In describing the electron configuration, small letters are used to represent the values of  $l$  as follows:

$$\begin{array}{cccccc} l = 0, & 1, & 2, & 3, & 4, & 5, \dots \\ s, & p, & d, & f, & g, & h, \dots \end{array}$$

i.e., if an electron is in a shell for which  $l = 0$ , it is called an  $s$  electron; for  $l = 1$ , a  $p$  electron and so on. The value of the principal quantum number  $n$  is written as a prefix to the letter representing its  $l$  value. For example, a state in which  $n = 2$ ,  $l = 0$  is a  $2s$ -state;  $n = 4$ ,  $l = 2$  is a  $4d$ -state and so on. The number of electrons having the same  $n$  and  $l$  values is indicated by an index written at the upper right of the letter representing their  $l$  value. Thus the 11 electrons of sodium in the normal state are designated as follows:  $1s^2 2s^2 2p^6 3s$ . i.e., there are two  $1s$  electrons, two  $2s$  electrons, six  $2p$  electrons, and one  $3s$  electron. We shall now consider electron configurations of a few elements.

(1) **Hydrogen ( $Z = 1$ )**. The normal state of an atom is one in which all the electrons are in the lowest possible energy levels. In hydrogen, the normal state is characterised by the quantum numbers  $n = 1$ ,  $l = 0$ ,  $m_l = 0$  and  $m_s$  may be either  $+\frac{1}{2}$  or  $-\frac{1}{2}$ . The symbolic representation is  $1s$ . The  $K$ -shell

requires one more electron to be completed. Hence atomic hydrogen is very active chemically.

(2) **Helium ( $Z = 2$ )**. It has both its electrons in the shell  $n = 1$ ,  $l = 0$ .  $m_s = \frac{1}{2}$  for one electron and  $-\frac{1}{2}$  for the second electron. The symbolic representation is  $[1s^2]$ . This shell is now completed

or closed. The rectangular enclosure indicates that the electrons are interlocked in a closed shell. Therefore, helium may be expected to have a very stable configuration. This should also be true of all the other inert gases.

(3) **Lithium ( $Z = 3$ )**. It has three electrons. Two electrons can be put in the shell  $n = 1$ ,  $l = 0$ . The third electron must be put into a new shell  $n = 2$ ,  $l = 0$ . So the neutral lithium atom is represented by  $1s^2 2s$ . Lithium is one of the alkali elements and has a valence of unity. This means that the valence  $2s$  electron can be detached easily from the atom to form the lithium ion  $\text{Li}^+$ . This is indicated by the fact that its ionisation potential is only 5.39 volts, whereas for  $\text{He}$  it is 24.58 volts. Lithium is chemically quite active and is monovalent. Similarly, all alkali metals ( $\text{Na}$ ,  $\text{K}$ ,  $\text{Rb}$ ,  $\text{Cs}$ ) have one electron in their outermost shell and hence are monovalent having similar chemical properties.

(4) **Beryllium ( $Z = 4$ )**. It has two electrons in the completed  $K$ -shell ( $n = 1$ ). It has two additional electrons in the  $(n = 2, l = 0)$  subshell. It is represented by  $1s^2 2s^2$ . Beryllium is one of

the alkaline earth elements with a valence of 2. Its optical spectrum is that of a two-electron atom. The atoms of the other elements of the group (*Mg, Ca, Sr, Ba, Ra*), should have similar structures. They have two electrons outside an inert gas or closed shell configuration. They have very similar chemical and physical properties.

Similarly the electronic configurations from boron ( $Z = 5$ ) to neon ( $Z = 10$ ) are:

- (i) Boron :  $1s^2 2s^2 2p$
- (ii) Carbon ( $Z = 6$ ) :  $1s^2 2s^2 2p^2$
- (iii) Nitrogen ( $Z = 7$ ) :  $1s^2 2s^2 2p^3$
- (iv) Oxygen ( $Z = 8$ ) :  $1s^2 2s^2 2p^4$
- (v) Flourine ( $Z = 9$ ) :  $1s^2 2s^2 2p^5$
- (vi) Neon ( $Z = 10$ ) :  $1s^2 2s^2 2p^6$

In neon, both the  $n = 1$  and  $n = 2$  shells are completed. Neon is one of the inert gases and has a very stable configuration.

The next eight elements from sodium ( $Z = 11$ ) to argon ( $Z = 18$ ) are formed by adding the additional electrons to the  $M$  shell for which  $n = 3$ .

- (i) Sodium ( $Z = 11$ ) :  $1s^2 2s^2 2p^6$   $3s$ . Sodium has an electron ( $3s$  electron) outside a closed shell. This single electron, like that in lithium, is easily ionised; the valence is 1; the spectrum is that of one-electron atom.
- (ii) Magnesium ( $Z = 12$ ) :  $1s^2 2s^2 2p^6$   $3s^2$ . The two electrons in the outermost incomplete  $M$  shell ( $n = 3$ ) are the valence electrons making *Mg divalent*.
- (iii) Aluminium ( $Z = 13$ ) :  $1s^2 2s^2 2p^6$   $3s^2 3p$ . *Al* is trivalent.

#### 4.18 MAGNETIC DIPOLE MOMENT DUE TO ORBITAL MOTION OF THE ELECTRON

Consider an electron moving in an elliptical orbit of area  $A$  with a period  $T$  (Fig. 4.28). The electron crosses any point in the orbit  $1/T$  times in unit time. This is equivalent to a current  $i = e/T$  in a loop of area  $A$ , where  $e$  is charge of the electron. Applying Ampere's theorem, this current gives rise to a magnetic dipole moment  $\mu_l$  given by

$$\mu_l = iA = eA/T \quad \dots(1)$$

where  $A$  is the area of the orbit.

Since the areal velocity in a central orbit is  $\frac{1}{2} r^2 \frac{d\phi}{dt}$ , the area

$$A = \int_0^T \frac{1}{2} r^2 \left( \frac{d\phi}{dt} \right) dt.$$

Now, the angular momentum of the electron

$$p_l = mr^2 \frac{d\phi}{dt} = \text{constant}$$

$$\text{or } \frac{1}{2} r^2 \frac{d\phi}{dt} = \frac{p_l}{2m} = \text{constant}$$

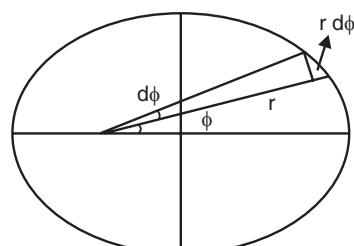


Fig. 4.28

$$\therefore A = \int_0^T \left( \frac{p_l}{2m} \right) dt = \frac{p_l T}{2m} \quad \dots(2)$$

$$\text{Substituting this value of } A \text{ in (1), } \mu_l = \frac{e}{2m} p_l \quad \dots(3)$$

Since angular momentum is quantised, we have  $p_l = l\hbar$

$$\therefore \mu_l = \frac{e}{2m} \times l\hbar = \frac{e\hbar}{2m} l \quad \dots(4)$$

$\mu_l$  is directly proportional to  $l$ .

$e\hbar/(2m)$  is the smallest unit of magnetic dipole moment and is called the *Bohr electron magneton* ( $\mu_B$ ).

$$\mu_B = \frac{e\hbar}{2m} = \frac{(1.602 \times 10^{-19}) \times (1.055 \times 10^{-34})}{2 \times (9.109 \times 10^{-31})} = 9.274 \times 10^{-24} \text{ JT}^{-1}$$

Since the electronic charge is negative, the magnetic dipole moment vector is directed opposite to that of the angular momentum (Fig. 4.29). The ratio (magnetic dipole moment/angular momentum) is called the *gyromagnetic ratio*.

$$\text{i.e., } \frac{\mu_l}{p_l} = \frac{e}{2m} = 8.8 \times 10^9 \text{ C/kg.} \quad \dots(5)$$

Eq. (5) may now be written as

$$\frac{\mu_l}{p_l} = -g_l \frac{e}{2m} \text{ with } g_l = 1 \quad \dots(5a)$$

$$\text{Similarly, } \frac{\mu_s}{S} = -g_s \frac{e}{2m} \text{ with } g_s = 2 \quad \dots(5b)$$

$$\frac{\mu_j}{J} = -g_j \frac{e}{2m} \quad \dots(5c)$$

Here,  $g$  is called *Lande's splitting factor*.

### Note

When a magnetic dipole  $\mu$  is placed in an external magnetic field  $B_{ext}$ , it experiences a torque (Fig. 4.30).

$$\vec{\tau} = \vec{\mu} \times \vec{B}_{ext}.$$

The torque tends to align the dipole moment vector along that of the field. Maximum torque occurs when the angle  $\theta$  between  $\mu$  and  $B_{ext}$  is  $90^\circ$ . The P.E. of a magnetic dipole, at any angle  $\theta$  relative to an external magnetic field is

$$\Delta U_m = -\vec{\mu} \cdot \vec{B}_{ext} = -\mu B_{ext} \cos\theta.$$

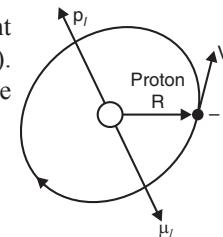


Fig. 4.29

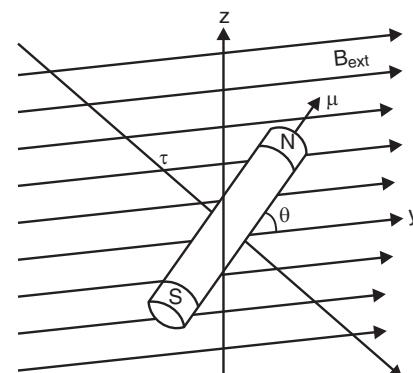


Fig. 4.30

## 4.19 MAGNETIC DIPOLE MOMENT DUE TO SPIN

An electron spinning about its axis should also behave as a tiny magnet and possess a magnetic dipole moment due to this spin. However, nothing is known about the shape of an electron or the

manner in which its charge is distributed. Hence it is impossible to calculate its spin magnetic dipole moment in a manner analogous to that used for the orbital motion. In order to obtain agreement with experimental results, the spin magnetic dipole moment ( $\mu_s$ ) is assigned the value,

$$\begin{aligned}\mu_s &= 2 \times \frac{e}{2m} p_s \text{ where } p_s = s\hbar. \\ \therefore \mu_s &= 2 \frac{e}{2m} \times s\hbar = 2s \frac{e\hbar}{2m} = \frac{e\hbar}{2m} \left( \because s = \frac{1}{2} \right)\end{aligned}$$

#### 4.20 THE STERN AND GERLACH EXPERIMENT

We have seen that the orbital and spin motions of the electrons in atoms endow the atoms with magnetic moments. Direct evidence for the existence of magnetic moments of atoms and their space quantisation is provided by the experiments of Stern and Gerlach.

**Principle and Theory.** The experiment is based on the behaviour of a magnetic dipole (atomic magnet) in a *non-uniform magnetic field* ( $B$ ). In a uniform magnetic field ( $B$ ), the dipole experiences a torque that tends to align the dipole parallel to the field. If the dipole moves in such a field in a direction normal to the field, it will trace a straight line path without any deviation. In an inhomogeneous magnetic field, the dipole experiences, in addition, a translatory force. If the atomic magnet flies across such an inhomogeneous magnetic field normal to the field direction, it will be deviated away from its rectilinear path. An expression for the deviation produced may be obtained as follows.

Let the magnetic field vary along the  $X$ -direction, so that the field gradient is  $dB/dx$  and is positive (Fig. 4.31).  $CD$  is the atomic magnet (of pole strength  $p$ , length  $l$ , dipole moment  $M$ ) with its axis inclined at an angle  $\theta$  to the field direction. If the field strength at the pole  $C$  is  $B$ , then the field strength at the other pole  $D$  will be  $B + \frac{dB}{dx}l \cos \theta$ . Hence the forces on the two poles are  $pB$  and  $p \left( B + \frac{dB}{dx}l \cos \theta \right)$ . Hence the atomic magnet experiences not only a torque ( $= plB = \mu_s B$ ) but also a translatory force

$$\begin{aligned}F_x &= \frac{dB}{dx} pl \cos \theta. \\ \therefore F_x &= \frac{dB}{dx} \mu_s \cos \theta \quad \dots(1)\end{aligned}$$

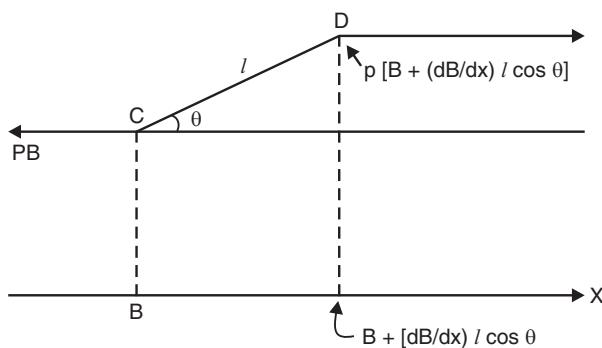


Fig. 4.31

Let

$V$  = velocity of the atomic magnet of mass  $m$  as it enters the field,

$L$  = length of the path of the atom in the field and

$t$  = the time of travel of the atom through the field =  $L/V$ .

The acceleration given to the atom along the field direction, by the translatory force  $F_x$  is  $F_x/m$ .

The displacement of the atom along the field direction, on emerging out of the field is

$$\begin{aligned}\alpha &= \frac{1}{2} \left( \frac{F_x}{m} \right) t^2 = \frac{1}{2} \frac{F_x}{m} \frac{L^2}{V^2} \\ &= \frac{1}{2} \frac{dB}{dx} \frac{\mu_s \cos \theta}{m} \frac{L^2}{V^2} \quad \dots(2)\end{aligned}$$

If  $\mu$  is resolved component of the magnetic moment in the field direction,  $\mu = \mu_s \cos \theta$

$$\therefore \alpha = \frac{1}{2} \frac{dB}{dx} \frac{\mu}{m} \frac{L^2}{V^2} \quad \dots(3)$$

**Experimental Arrangement.** Silver is boiled in an oven [Fig. 4.32 (a)]. Atoms of silver stream out from an opening in the oven. By the use of slits  $S_1$  and  $S_2$ , a sharp linear beam of atoms is obtained. These atoms then pass through a very inhomogeneous magnetic field between the shaped poles of a magnet  $M M$ . A high degree of non-uniformity in the magnetic field is produced by making one of the pole-pieces of a powerful electromagnet a knife-edge shape and the other flat

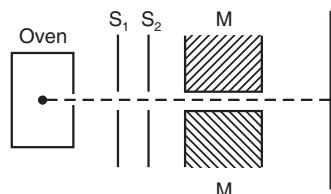


Fig. 4.32 (a)

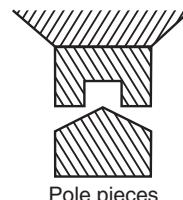


Fig. 4.32 (b)

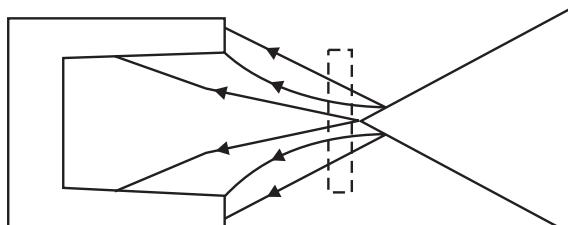


Fig. 4.32 (c)

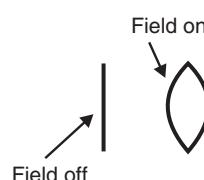


Fig. 4.32 (d)

with a groove cut in it opposite the knife edge [Fig. 4.32 (b)]. The lines of force are close together at the knife edge and the field there is much stronger than that at the other pole piece [Fig. 4.32 (c)]. The magnetic field is at right angles to the direction of movement of the atoms. Finally the atoms fall on a photographic plate  $P$ . The whole arrangement is enclosed in an evacuated chamber.

**Results.** With no field, the beam produces a narrow continuous line on the plate [Fig. 4.32 (d)]. In terms of the vector atom model, those atoms, with electron spins directed parallel to the magnetic field, will experience a force in one direction, whereas those with oppositely directed spins will experience a force in the opposite direction. According to this, the beam of atoms should split into two beams in its passage through the inhomogeneous magnetic field. This splitting of the beam into two parts of approximately equal intensity was actually observed in these experiments. On applying the inhomogeneous magnetic field, it was found that the stream of silver atoms splits into two separate lines [Fig. 4.32 (d)]. Knowing  $dB/dx$ ,  $L$ ,  $V$  and  $\alpha$ ,  $\mu$  was calculated. It was found that each silver atom had a magnetic moment of one *Bohr magneton* in the direction of the field.

## 4.21 OPTICAL SPECTRA

The outer (valence) electrons of an atom determine the chemical and *optical* properties of their atoms. The electrons in the closed inner orbits do not take part in the emission of spectral lines.

**Spectral terms.** Atoms are divided into two main categories, *viz.*, one electron system and many-electron system.

(i) The alkali metals (such as lithium, sodium and potassium) have a single valence electron outside completely filled subshells. The valence electron in the alkali metals therefore behaves much like the orbiting electron in the hydrogen atom. Hence the alkali metals have a hydrogen-like spectra. Their spectrum is also referred to as the *one-electron spectra*.

(ii) In the many-electron system, the atoms have more than one valence or *optical electron* outside completely filled subshells. Hence they become effective in fixing the spectral properties. For example, the alkaline earths belonging to the two-electron system have spectra which are similar among themselves.

**Spectral notation.** The states of the atom, in which the values of its  $L$  vector are 0, 1, 2, 3, 4, 5 are represented by the capital letters,  $S, P, D, F, G, H$ , etc., respectively. The value of the total angular momentum of the atom  $J$  is written as a subscript at the lower right of the letter representing the particular  $L$  value of the atomic state. The multiplicity of the total spin ( $S$ ) is written as a superscript at the upper left of letter. If  $S$  is the total spin, the multiplicity is equal to  $2S + 1$ .

**Examples:** (i) A state with  $L = 1$ ,  $S = \frac{1}{2}$  and  $J = 3/2$  would be written  $^2P_{3/2}$  and read “doublet

$P$  three halves”. (Since  $S = \frac{1}{2}$ , the multiplicity of the state  $= 2 \times \frac{1}{2} + 1 = 2$ ).

(ii) A state with  $L = 2$ ,  $S = 1$  and  $J = 2$  would be written  $^3D_2$  and read “triplet  $D$  two”.

(iii) The spectral terms corresponding to  $L = 1$  and  $S = \frac{1}{2}$  are written as  $^2P_{1/2}$  and  $^2P_{3/2}$ . (Since  $L = 1$ , the capital letter which represents the state is  $P$ ; the multiplicity of the state  $= 2 \times \frac{1}{2} + 1 = 2$ ;

this is put as a superscript;  $J = L \pm S$ ;  $J = 1 + \frac{1}{2} = \frac{3}{2}$  or  $1 - \frac{1}{2} = \frac{1}{2}$ .  $\frac{1}{2}$  or  $\frac{3}{2}$  are put as subscripts).

**The selection rules.** An electron cannot jump from one energy level to all other energy levels. A transition of an electron between two levels is possible only if certain rules called *selection rules* are satisfied. For the vector atom model, three selection rules have been devised.

(i) **The selection rule for  $L$**  is  $\Delta L = \pm 1$  *i.e.*, only those lines are observed for which the value of  $L$  changes by  $\pm 1$ . For example,  $L$  can change from 0 to 1 ( $\Delta L = 1$ ) or from 1 to 0 ( $\Delta L = -1$ ).  $L$  cannot change from 0 to 2 ( $\Delta L = +2$ ) or from 2 to 0 ( $\Delta L = -2$ ). In symbols, a transition is possible between  $S$  and  $P$  levels ( $\Delta L = \pm 1$ ), but not possible between  $S$  and  $D$  levels ( $\Delta L = \pm 2$ ).

(ii) **The selection rule for  $J$**  is  $\Delta J = \pm 1$  or 0. But  $0 \rightarrow 0$  is excluded.

(iii) **The selection rule for  $S$**  is given by  $\Delta S = 0$ .

**Intensity Rules.** Whether an allowed transition is weak or strong is determined by what are known as *intensity rules*. They are:

(i) Transitions for which  $L$  and  $J$  change in the same way (*i.e.*,  $\Delta J = \Delta L$ ) are strong. For other changes in  $L$  and  $J$  ( $\Delta L \neq \Delta J$ ) we get weak transitions.

(ii) Transitions for which  $L$  and  $J$  increase (*i.e.*,  $L \rightarrow L + 1$  and  $J \rightarrow J + 1$ ) are less intense than those for which  $L$  and  $J$  decrease (*i.e.*,  $L \rightarrow L - 1$  and  $J \rightarrow J - 1$ ).

(iii) Transitions for which changes in  $L$  and  $J$  are opposite (*i.e.*,  $\Delta L = -\Delta J$ ) are forbidden. Symbolically, these intensity rules are written as

$\Delta L = -1,$	$\Delta J = -1$	strongest
$\Delta L = -1,$	$\Delta J = 0$	less intense
$\Delta L = +1,$	$\Delta J = +1$	weak
$\Delta L = +1,$	$\Delta J = 0$	very weak
$\Delta L = -1,$	$\Delta J = +1$	Forbidden.
$\Delta L = +1,$	$\Delta J = -1$	

**The Interval Rule.** Lande discovered a rule regarding the interval in frequency between the different levels constituting a multiplet. It states that *the frequency interval between two levels with total angular momenta ( $J+1$ ) and  $J$  respectively is proportional to ( $J+1$ )*.

#### Fine structure of the sodium D-line

Ten out of the eleven electrons of the normal sodium atom are interlocked in closed shells. They contribute nothing to the angular momentum of the atom. We have to consider only the states of eleventh optical electron in discussing the spectrum of neutral sodium.

The  $D$ -line belongs to principal series. Lines of the principal series are due to transitions from a  $P$  state to the  $S$  state. For the upper  $P$  state,  $L = 1$ ,  $J = L \pm S = 3/2$  or  $1/2$ . Hence the two

possible terms are:  $^2P_{3/2}$  and  $^2P_{1/2}$ . For the lower  $S$  state,  $L = 0$ ,

$J = \frac{1}{2}$ , so that only one term  $^2S_{1/2}$  is possible.

Fig. 4.33 shows the two possible transitions between the two terms of the  $P$  state and the single term of the  $S$  state. They are :

(i)  $^2P_{1/2} \rightarrow ^2S_{1/2}$  gives the  $D_1$  line of wavelength  $5896 \text{ \AA}$ .

(ii)  $^2P_{3/2} \rightarrow ^2S_{1/2}$  gives the  $D_2$  line of wavelength  $5890 \text{ \AA}$ .

Now applying the selection rules  $\Delta L = \pm 1$  and  $\Delta J = \pm 1$  or 0 (excluding  $0 \rightarrow 0$ ), both the transitions are allowed. This explains the doublet fine structure of the sodium  $D$  line.

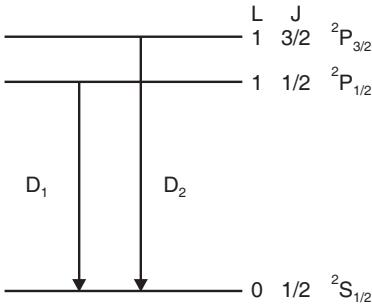


Fig. 4.33

#### 4.22 FINE STRUCTURE OF $H_\alpha$ LINE

According to Bohr's theory,  $H_\alpha$  line, the first of Balmer series, arises due to transition from the third quantum state ( $n = 3$ ) to the second quantum state ( $n = 2$ ).

Fig. 4.34 shows the different energy states for  $n = 2$  and  $n = 3$  in case of Hydrogen atom.

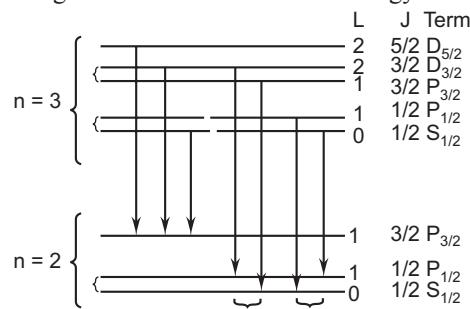


Fig. 4.34

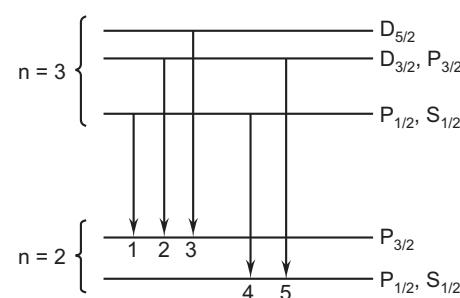


Fig. 4.35

**For the upper state  $n = 3$** 

$L$ values	0	1	2
$J = L \pm S$	$\frac{1}{2}$	$\frac{3}{2}$ and $\frac{1}{2}$	$\frac{5}{2}$ and $\frac{3}{2}$ ( $\because S = \frac{1}{2}$ )
Spectral terms	$^2S_{1/2}$	$^2P_{3/2}$ and $^2P_{1/2}$	$^2D_{5/2}$ and $^2D_{3/2}$

**For the lower state  $n = 2$** 

$L$ values	0	1
$J = L \pm S$	$\frac{1}{2}$	$\frac{3}{2}$ and $\frac{1}{2}$
Spectral terms	$^2S_{1/2}$	$^2P_{3/2}$ and $^2P_{1/2}$

Theoretically fifteen transitions are possible between the five terms of the upper state ( $n = 3$ ) and three terms of the lower state ( $n = 2$ ). But by applying the selection rules namely,  $\Delta L = \pm 1$  and  $\Delta J = 0$  or  $\pm 1$ , these transitions reduce to the following seven transitions:

$$\begin{aligned} 3^2D_{5/2} &\rightarrow 2^2P_{3/2}, \quad 3^2D_{3/2} \rightarrow 2^2P_{3/2}, \quad 3^2D_{3/2} \rightarrow 2^2P_{1/2}, \quad 3^2P_{3/2} \rightarrow 2^2S_{1/2}, \\ 3^2P_{1/2} &\rightarrow 2^2S_{1/2}, \quad 3^2S_{1/2} \rightarrow 2^2P_{3/2}, \quad 3^2S_{1/2} \rightarrow 2^2P_{1/2}. \end{aligned}$$

Out of these seven allowed transitions, two pairs in separate cases are identical, i.e.,

$$\begin{aligned} [3^2D_{3/2} \rightarrow 2^2P_{1/2} \text{ and } 3^2P_{3/2} \rightarrow 2^2S_{1/2}]; \\ [3^2P_{1/2} \rightarrow 2^2S_{1/2} \text{ and } 3^2S_{1/2} \rightarrow 2^2P_{1/2}] \end{aligned}$$

since they represent transitions between coincident levels i.e., levels whose  $L$  values differ by unity, but  $J$  values are the same.

*Thus the fine structure of  $H_\alpha$  line should have five components.*

Now applying the intensity rules, the transitions  $D_{5/2} \rightarrow P_{3/2}$  and  $[D_{3/2} \rightarrow P_{1/2} \text{ and } P_{3/2} \rightarrow S_{1/2}]$  have the maximum intensity because they are of type  $\Delta L = -1$  and  $\Delta J = -1$ . The two components  $D_{3/2} \rightarrow P_{3/2}$  and  $P_{1/2} \rightarrow S_{1/2}$  are less intense since they are of the type  $\Delta L = -1$  and  $\Delta J = 0$ . The components  $S_{1/2} \rightarrow P_{3/2}$  is still weaker since here  $\Delta L = +1$  and  $\Delta J = +1$ . The weakest transition is  $S_{1/2} \rightarrow P_{1/2}$  because here  $\Delta L = +1$  and  $\Delta J = 0$ .

**Energy level diagram.** The various energy levels are drawn in as the horizontal levels as shown in Fig. 4.34. The coincident levels are drawn close to each other. There are two such levels in upper state and one in the lower state. All the seven allowed transitions are also shown in Fig. 4.34. Leaving the transitions between coincident levels, the five transitions are separately shown in Fig. 4.35.

### 4.23 ZEEMAN EFFECT

Zeeman effect is a magneto-optical phenomenon.

If a source of light producing line spectrum is placed in a magnetic field, the spectral lines are split up into components. When the splitting occurs into two or three lines, it is called *normal Zeeman effect* and can be explained quantitatively by classical theory. The splitting of a spectral line into more than three components in ordinary *weak magnetic fields* is called *anomalous Zeeman effect*. This cannot be explained by classical theory.

**Experimental arrangement for the normal Zeeman effect.** The arrangement is shown in Fig. 4.36.  $MM$  is an electromagnet capable of producing a very strong magnetic field. Its conical pole-pieces  $PP$  have longitudinal holes drilled through them. A source of light ( $L$ ) emitting line

spectrum (say, a sodium vapour lamp) is placed between the pole-pieces. The spectral lines are observed with the help of a spectrograph (*S*) of high resolving power. The Zeeman effect may be observed in two ways.

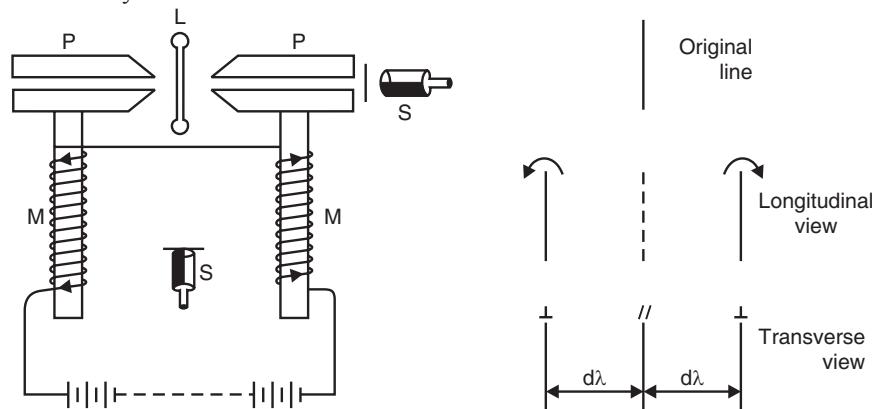


Fig. 4.36

(i) The position of the spectral line is noted without applying the magnetic field. The magnetic field is now switched on and the spectral line is viewed *longitudinally* through the hole drilled in the pole-pieces and hence parallel to the direction of the field. It is found that the spectral line is split into two components, one slightly shorter in wavelength and the other slightly longer in wavelength than the original line. The original line is not present. The two components are found to be symmetrically situated about the position of the parent line. Analysing the two lines with a Nicol prism, both the lines are found to be circularly polarised in opposite directions. This is called *normal longitudinal Zeeman effect*.

(ii) The spectral line is then viewed *transversely* i.e., perpendicular to the direction of the magnetic field. In this case, the single spectral line is *split* up into three components when the magnetic field is applied. The central line has the same wavelength as the original line and is plane polarised with vibrations parallel to the field. The outer lines are symmetrically situated on either side of the central line. The displacement of either outer line from the central line is known as the *Zeeman shift*. The two outer lines are also plane polarised having vibrations in a direction perpendicular to the field. This is called *normal transverse Zeeman effect*.

**Lorentz classical theory of normal Zeeman effect.** The emission of light by a glowing gas is due to the vibratory motion of the electrons. The electrons in the atom execute simple harmonic vibration about the centre of the atom. The frequency of the spectral line is given by the frequency of vibration of the electron. Any linear motion of an electron can be resolved into three components—a linear motion along the magnetic field and two opposite circular motions perpendicular to the field. It is known that no mechanical force acts upon a charge when it moves in the direction of the field. So the linear motion along the field is not affected by the field and hence the frequency of this component is unchanged by the magnetic field.

The other two circular components are affected by the magnetic field, one being retarded and the other being accelerated. Hence the frequency of one of these circular components is increased and that of the other is decreased by the same amount.

In the longitudinal view, the unaltered linear component along the field gives no light waves along the direction of observation due to the transverse nature of light vibrations. Hence the original line is not observed. The two altered circular components at right angles to the field produce circularly polarised light waves. This explains the longitudinal Zeeman pattern.

In the transverse view, the unaffected linear component sends out vibrations parallel to the field and hence perpendicular to the direction of observation. This gives the central line of the triplet

occupying the same position as the original line and plane polarised with vibrations parallel to the field. The two circular vibrations which are altered in frequency send out vibrations perpendicular to the field and when viewed sideways appear as plane polarised vibrations in opposite senses.

**Expression for the Zeeman shift.** Consider an electron in the atom moving in a circular orbit of radius  $r$  with a linear velocity  $v$  and angular velocity  $\omega$ . Let  $e$  be the charge of the electron and  $m$  its mass.

The centripetal force on the electron towards the centre in the absence of the magnetic field is

$$F = \frac{mv^2}{r} = m\omega^2 r \quad \dots(1)$$

Now let an external magnetic field of flux density  $B$  be applied in a direction perpendicular to the plane of the orbits of the two circular components and directed towards the reader. Then an additional radial force of magnitude  $Bev$  acts on the electron. The direction of this force will be outwards from the centre for clockwise motion but inwards towards the centre for anticlockwise motion (Fig. 4.37). The resulting complex motion of the electron subjected to an additional radial force is called *Larmor precession*. This produces a change in the angular velocity without any change in the form of the orbit.

Let  $\delta\omega$  be the change in angular velocity caused by the field. For the circular motion in the *clockwise* direction, the additional radial force is directed away from the centre.

$$\therefore F - Bev = m(\omega + \delta\omega)^2 r \quad \dots(2)$$

$$m\omega^2 r - m(\omega + \delta\omega)^2 r = Be\omega r$$

$$\text{or } -2mr\omega \delta\omega = Be\omega r \quad \text{neglecting } (\delta\omega)^2$$

$$\text{or } \delta\omega = -Be/2mr \quad \dots(3)$$

For the circular motion in the *anticlockwise* direction, the additional radial force is directed towards the centre.

$$\therefore F + Bev = m(\omega + \delta\omega)^2 r$$

$$\text{or } \delta\omega = +Be/2mr \quad \dots(4)$$

The two cases can be combined into the equation

$$\delta\omega = \pm \frac{Be}{2mr} \quad \dots(5)$$

If  $v$  = frequency of vibration of the electron,

$$\omega = 2\pi v; \delta\omega = 2\pi \delta v \quad \text{or} \quad \delta v = \frac{\delta\omega}{2\pi}$$

$$\left. \begin{array}{l} \text{Change in frequency} \\ \text{of the spectral line} \end{array} \right\} = \delta v = \pm \frac{Be}{4\pi mr} \quad \dots(6)$$

If  $v$  and  $\lambda$  are the frequency and wavelength of the original line,

$$v = \frac{c}{\lambda} \quad \text{or} \quad \delta v = \frac{-c}{\lambda^2} \delta\lambda.$$

$$\therefore \text{The Zeeman shift} = \delta\lambda = \pm \frac{Be\lambda^2}{4\pi mc} \quad \dots(7)$$

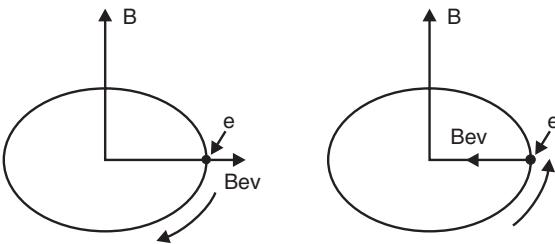


Fig. 4.37

**EXAMPLE 1.** Calculate the wavelength separation between the unmodified line of wavelength  $6000 \text{ \AA}$  and the modified lines when a magnetic induction of  $1 \text{ Wbm}^{-2}$  is applied, in normal Zeeman effect.

**SOL.** We have

$$\begin{aligned} d\lambda &= \frac{Be\lambda^2}{4\pi mc} = \frac{1 \times (1.6 \times 10^{-19}) (6000 \times 10^{-10})^2}{4\pi (9.1 \times 10^{-31}) (3 \times 10^8)} \\ &= 0.168 \times 10^{-10} \text{ m} = 0.168 \text{ \AA}. \end{aligned}$$

**EXAMPLE 2.** Calculate the wavelength separation between the two component lines which are observed in the normal Zeeman effect. The magnetic field used is  $0.4 \text{ weber/m}^2$ ; the specific charge  $= 1.76 \times 10^{11} \text{ C kg}^{-1}$  and  $\lambda = 6000 \text{ \AA}$ .

**SOL.** The wavelength separation

$$\lambda_1 - \lambda_2 = 2d\lambda = \frac{Be\lambda^2}{2\pi mc}$$

Here,

$$B = 0.4 \text{ weber/m}^2; e/m = 1.76 \times 10^{11} \text{ C kg}^{-1}$$

$$\lambda = 6000 \times 10^{-10} \text{ m}; c = 3 \times 10^8 \text{ ms}^{-1};$$

$$\begin{aligned} \therefore \lambda_1 - \lambda_2 &= \frac{0.4 \times (1.76 \times 10^{11}) (6000 \times 10^{-10})^2}{2\pi (3 \times 10^8)} \text{ m} \\ &= 1.335 \times 10^{-11} \text{ m} = 0.1335 \text{ \AA} \end{aligned}$$

**EXAMPLE 3.** The red line of cadmium splits into three components separated by  $120 \text{ MHz}$  when the source is placed in a magnetic field of flux density  $8.6 \text{ mT}$ , the light being examined in a direction perpendicular to the magnetic field. Calculate the ratio of charge to mass ( $e/m$ ) of the electron.

**SOL.** We have,

$$\delta v = Be/4\pi m.$$

$$\therefore \frac{e}{m} = \frac{4\pi \delta v}{B} = \frac{4\pi \times (1.2 \times 10^8)}{8.6 \times 10^{-3}} = 1.76 \times 10^{11} \text{ C kg}^{-1}$$

#### 4.24 LARMOR'S THEOREM

**Statement.** The effect of a magnetic field on an electron moving in an orbit is to superimpose on the orbital motion a precessional motion of the entire orbit about the direction of the magnetic field with angular velocity  $\omega$  given by  $\omega = Be/2m$ .

**Explanation.** Fig. 4.38 shows two positions of the vector  $\vec{l}$  as it precesses about the magnetic field at constant inclination and the corresponding positions of the electronic orbit. Just as a mechanical top precesses in a gravitational field, an electron in an orbit precesses in a magnetic field. This is called Larmor precession.

**Change in K.E. due to Larmor precession.** In the magnetic field, the angular velocity of the atomic system changes by an amount  $Be/2m$ . Let  $\omega_0$  be the original angular velocity. Then change in K.E.

$$\begin{aligned} &= \frac{1}{2} mr^2 \left( \omega_0 + \frac{Be}{2m} \right)^2 - \frac{1}{2} mr^2 \omega_0^2 \\ &= \frac{1}{2} mr^2 \omega_0^2 \left( 1 + \frac{Be}{2m\omega_0} \right)^2 - \frac{1}{2} mr^2 \omega_0^2 \end{aligned}$$

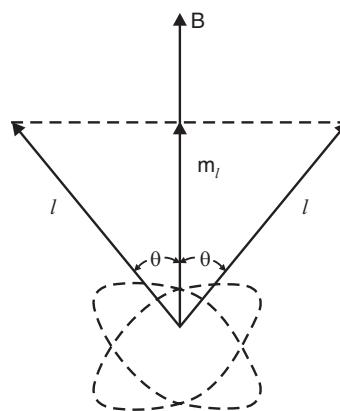


Fig. 4.38

$$\begin{aligned}
 &= \frac{1}{2} mr^2 \omega_0^2 \left( 1 + \frac{Be}{m\omega_0} \right) - \frac{1}{2} mr^2 \omega_0^2 \\
 &= mr^2 \omega_0^2 \frac{Be}{2m\omega_0} = L \frac{Be}{2m}
 \end{aligned}$$

Here,  $\mathbf{L}$  = original angular momentum =  $mr^2 \omega_0$ .

#### 4.25 QUANTUM MECHANICAL EXPLANATION OF THE NORMAL ZEEMAN EFFECT

Debye explained the normal Zeeman effect without taking into account the concept of electron spin. We neglect the spin motion of the electron,

$$\text{Orbital angular momentum of the electron} = \mathbf{L} = l\hbar \quad \dots(1)$$

$$\text{Magnetic moment of the electron} = \mu_l = l\hbar \frac{e}{2m} = \frac{e}{2m} \mathbf{L} \quad \dots(2)$$

In the presence of an external magnetic field of flux density  $\mathbf{B}$ , the vector  $\mathbf{l}$  precesses around the direction of the magnetic field as axis. The precession is known as Larmor precession.

$$\text{The frequency of Larmor precession} = \omega = Be/2m \quad \dots(3)$$

Fig. 4.38 shows two positions of the vector  $\mathbf{l}$ , as it precesses about the magnetic field at constant inclination and the corresponding positions of the electronic orbit. The additional energy of the electron due to this precessional motion

$$\begin{aligned}
 \Delta E &= \mu_l B \cos\theta = \left( \frac{e}{2m} l\hbar \right) B \cos\theta \quad \left( \because \mu_l = \frac{e}{2m} l\hbar \right) \\
 &= \frac{Be}{2m} \hbar l \cos\theta
 \end{aligned}$$

But  $Be/2m = \omega$  and  $l \cos \theta = \text{projection of } \mathbf{l} \text{ on } \mathbf{B} = m_l$

$$\therefore \Delta E = m_l \frac{e\hbar}{2m} B = m_l \omega \hbar \quad \dots(4)$$

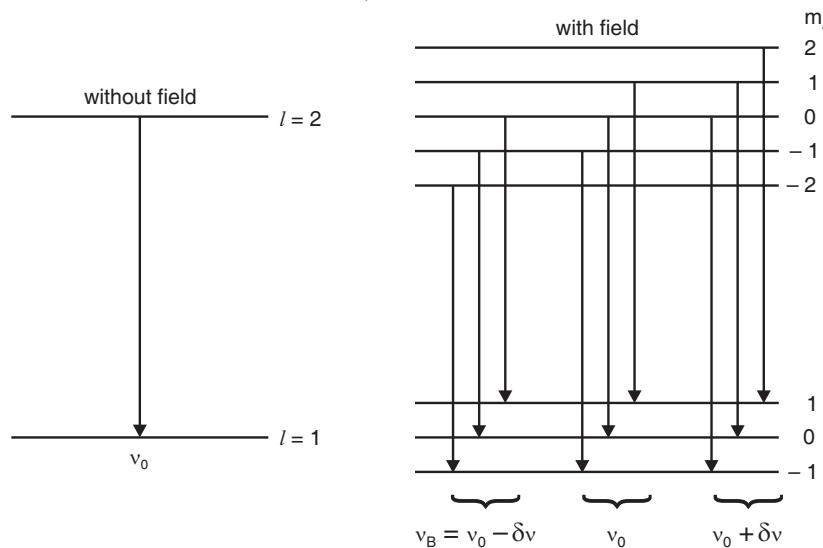


Fig. 4.39

Now,  $m_l$  can have  $(2l + 1)$  values from  $+l$  to  $-l$ . Therefore, an external magnetic field will split a single energy level into  $(2l + 1)$  levels. The  $d$ -state ( $l = 2$ ) is split into 5 sub-levels and the  $p$ -state ( $l = 1$ ) is split into 3 sublevels (Fig. 4.39).

Let  $E'_o$  represent the energy of the level  $l = 1$  in the absence of the magnetic field and  $E'_B$  represent the energy of this level in the presence of magnetic field. Then,

$$E'_B = E'_o + \Delta E' = E'_o + m'_l \frac{e\hbar}{2m} B \quad \dots(5)$$

Similarly, if  $E''_o$  and  $E''_B$  represent the energies of the level  $l = 2$  without and with the magnetic field respectively, then,

$$E''_B = E''_o + \Delta E'' = E''_o + m''_l \frac{e\hbar}{2m} B \quad \dots(6)$$

The quantity of energy radiated in the presence of magnetic field is

$$\begin{aligned} E''_B - E'_B &= (E''_o - E'_o) + (m''_l - m'_l) \frac{e\hbar}{2m} B \\ E''_B - E'_B &= (E''_o - E'_o) + (m''_l - m'_l) \frac{e\hbar}{4\pi m} B \\ h\nu &= h\nu_o + \Delta m_l \frac{e\hbar}{4\pi m} B \\ \text{or } \nu &= \nu_o + \Delta m_l \frac{eB}{4\pi m} \end{aligned} \quad \dots(7)$$

where  $\nu$  = frequency of the radiation emitted with the magnetic field and  $\nu_o$  = frequency of the radiation in the absence of the magnetic field. The selection rule for  $m_l$  is  $\Delta m_l = 0$  or  $\pm 1$ .

Hence we have three possible lines,

$$\nu_1 = \nu_o \text{ for } \Delta m_l = 0 \quad \dots(8)$$

$$\nu_2 = \nu_o + \frac{eB}{4\pi m} \text{ for } \Delta m_l = +1 \quad \dots(9)$$

$$\text{and } \nu_3 = \nu_o - \frac{eB}{4\pi m} \text{ for } \Delta m_l = -1 \quad \dots(10)$$

Fig. 4.39 represents the normal Zeeman effect. Although there are nine possible transitions, they are grouped into only three different frequency components as indicated by equations (8), (9) and (10). For three transitions in a bracket, change in the value of  $\Delta m_l$  is the same and hence they represent same change of energy and a single line.

## 4.26 ANOMALOUS ZEEMAN EFFECT

This can be explained only by using the idea of the spin of the electron. With the introduction of spin, we have two angular momentum vectors  $\mathbf{l}$  and  $\mathbf{s}$  associated with each electron. The total angular momentum vector  $\mathbf{j} = \mathbf{l} + \mathbf{s}$  ...(1)

The magnetic moment due to orbital motion

$$\mu_l = \mathbf{l} \frac{e\hbar}{2m} \quad \dots(2)$$

$\mu_l$  is directed oppositely to  $\mathbf{l}$  because of the negative charge of the electron.

Similarly, the magnetic moment due to the spin of the electron,

$$\mu_s = \frac{e\hbar}{2m} \quad \dots(3)$$

$\mu_s$  is oppositely directed to  $\mathbf{s}$  because of the negative charge of the electron. The relationships between the magnetic moments and the angular momenta are shown in Fig. 4.40.

The resultant magnetic moment  $\mu$  is not along  $\mathbf{j}$ . Since  $\mathbf{l}$  and  $\mathbf{s}$  precess about  $\mathbf{j}$ ,  $\mu_l$  and  $\mu_s$  must also precess about  $\mathbf{j}$ .

To find out the resultant magnetic moment of the electron, each of these vectors  $\mu_l$  and  $\mu_s$  is resolved into two components, one along  $\mathbf{j}$  and the other perpendicular to it. The value of the perpendicular component of each vector, averaged over a period of the motion will be zero, since it is constantly changing direction. The effective magnetic moment of the electron will be

$\mu_j = \text{component of } \mu_l \text{ along the direction of } \mathbf{j} + \text{component of } \mu_s \text{ along the direction of } \mathbf{j}$

$$\begin{aligned} &= \frac{e\hbar}{2m} \mathbf{l} \cos(\mathbf{l}, \mathbf{j}) + \frac{e\hbar}{2m} \mathbf{s} \cos(\mathbf{s}, \mathbf{j}) \\ &= \frac{e\hbar}{2m} [\mathbf{l} \cos(\mathbf{l}, \mathbf{j}) + 2\mathbf{s} \cos(\mathbf{s}, \mathbf{j})] \end{aligned}$$

But according to cosine law,

$$\cos(\mathbf{l}, \mathbf{j}) = \frac{l^2 + j^2 - s^2}{2lj}$$

and

$$\cos(\mathbf{s}, \mathbf{j}) = \frac{s^2 + j^2 - l^2}{2sj}$$

Hence,

$$\begin{aligned} \mu_j &= \frac{e\hbar}{2m} \left[ \frac{l^2 + j^2 - s^2}{2j} + \frac{s^2 + j^2 - l^2}{j} \right] \\ &= \frac{e\hbar}{2m} \left[ \frac{3j^2 + s^2 - l^2}{2j} \right] \\ &= \frac{e\hbar}{2m} j \left[ 1 + \frac{j^2 + s^2 - l^2}{2j^2} \right] \end{aligned}$$

Writing  $j^2 = j(j+1)$  and so on,

$$\mu_j = \frac{e\hbar}{2m} j \left[ 1 + \frac{j(j+1) + s(s+1) - l(l+1)}{2j(j+1)} \right]$$

The quantity  $1 + \frac{j(j+1) + s(s+1) - l(l+1)}{2j(j+1)} = g$

is called the Lande  $g$  factor.

Hence  $\mu_j = \frac{e\hbar}{2m} gj$ .

If the atom is placed in a weak magnetic field, the total angular momentum vector  $\mathbf{j}$  precesses about the direction of the magnetic field as axis.

The additional energy  $\Delta E$  due to the action of the magnetic field on this atomic magnet is

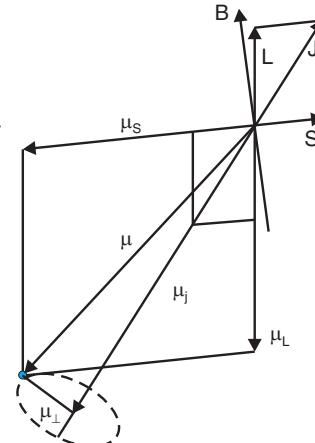


Fig. 4.40

$$\Delta E = \mu_j B \cos(j, B) = \frac{e\hbar}{2m} g j B \cos(j, B).$$

But  $j \cos(j, B)$  = the projection of the vector  $\mathbf{j}$  on the direction of the magnetic field =  $m_j$ .

Hence,

$$\Delta E = \frac{e\hbar}{2m} B g m_j.$$

The quantity  $\frac{e\hbar}{2m} B$  is called a *Lorentz unit*. It is a unit of energy used for expressing the splitting of the energy levels in a magnetic field.

Since  $m_j$  has  $(2j + 1)$  values, a given energy level is split up into  $(2j + 1)$  sublevels with the application of magnetic field. When this  $m_j$  is subjected to the selection rule  $\Delta m_j = 0$  or  $\pm 1$ , we get the transitions shown in Fig. 4.41 for the sodium D-lines.

The ground state  $^2S_{1/2}$  splits into two sublevels. Here,

$$l = 0, s = \frac{1}{2}; \quad j = \frac{1}{2}; \quad \text{hence,}$$

$$g = 1 + \frac{\frac{1}{2} \times \frac{3}{2} + \frac{1}{2} \times \frac{3}{2}}{2 \times \frac{1}{2} \times \frac{3}{2}} = 2.$$

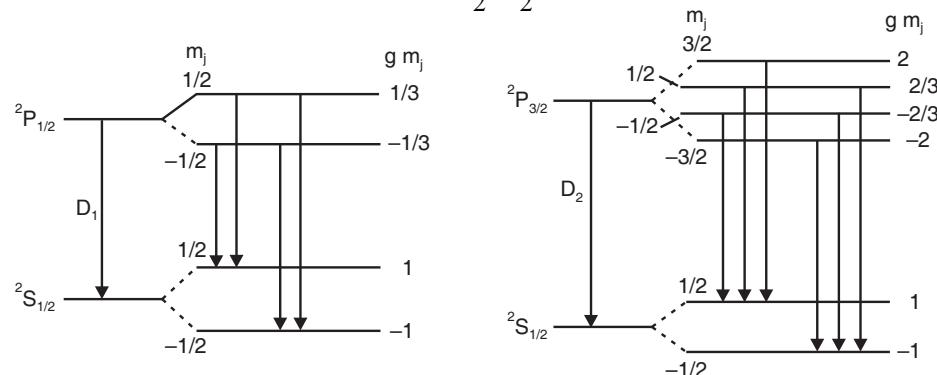


Fig. 4.41

Since  $m_j$  can have the values  $\frac{1}{2}$  and  $-\frac{1}{2}$ ,  $g m_j$  can have the values  $+1$  and  $-1$ . The following

table (Table 4.3) gives the values for the quantum numbers necessary for the determination of the splitting factor  $g m_j$  for each of the energy levels of the sodium D-lines.

TABLE 4.3

State	$l$	$s$	$j$	$g$	$m_j$	$m_j g$
$^2S_{1/2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	2	$\frac{1}{2}, -\frac{1}{2}$	1, -1
$^2P_{1/2}$	1	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{2}, -\frac{1}{2}$	$\frac{1}{3}, -\frac{1}{3}$
$^2P_{3/2}$	1	$\frac{1}{2}$	$\frac{3}{2}$	-	$\frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}$	$2, \frac{2}{3}, -\frac{2}{3}, -2$

The longer wavelength component  $^2P_{1/2} \rightarrow ^2S_{1/2}$  splits into four lines. The shorter wavelength component  $^2P_{3/2} \rightarrow ^2S_{1/2}$  splits into six lines.

Thus the introduction of electron spin has led to complete agreement between the experimental results and the theory of the anomalous Zeeman effect.

### 4.27 PASCHEN-BACK EFFECT

*Paschen and Back found that whatever be the anomalous Zeeman pattern of a given line in a weak magnetic field, the pattern always approximates the normal Zeeman triplet as the field strength is progressively increased. This reduction may occur either through the coalescence of lines or through the disappearance of certain lines. This transition phenomenon is called Paschen-Back effect.*

**Explanation.** In a strong magnetic field, the coupling between  $\mathbf{l}$  and  $\mathbf{s}$  breaks down, and  $\mathbf{j}$  loses its significance.  $\mathbf{l}$  and  $\mathbf{s}$  are quantised separately.  $\mathbf{l}$  and  $\mathbf{s}$  precess separately about the external magnetic field  $\mathbf{B}$  independent of each other (Fig. 4.42). The energy change due to the presence of the field will then be made up of two parts, one arising from the precession of  $\mathbf{l}$  about  $\mathbf{B}$  and the other from the precession of  $\mathbf{s}$  about  $\mathbf{B}$ .

Hence

$$\begin{aligned}\Delta E &= (\Delta E)_l + (\Delta E)_s \\ &= B \frac{eh}{4\pi m} [l \cos(l, B) + 2s \cos(s, B)] \\ &= \frac{eh}{4\pi m} B (m_l + 2m_s).\end{aligned}$$

In terms of frequency change,

$$\Delta\nu = \frac{eB}{4\pi m} \Delta (m_l + 2m_s).$$

The quantity  $(m_l + 2m_s)$  is known as the *strong field quantum number* and is evidently an integer. Now since  $\Delta m_l = 0$  or  $\pm 1$ ,  $\Delta m_s = 0$ ,  $\Delta(m_l + 2m_s) = 0$  or  $\pm 1$ .

Hence in a strong magnetic field, a given spectral line will split into three components only and this is the usual characteristic of the normal Zeeman effect.

### 4.28 STARK EFFECT

The Stark effect is the electrical analogue of the Zeeman effect. The Stark effect is the splitting of spectral lines due to the action of an external electric field on the radiating substance. Even very strong external electric fields are weak compared to the interatomic fields. Hence the action of electric field on the motion of the atomic electrons can be regarded as small perturbations. Consequently, the Stark line splitting is very minute and can be observed only with instruments having a high resolving power. The lines are split into a series of components (satellites) located, in case of hydrogen, symmetrically on both sides of the original line.

**Experimental study.** Here the hydrogen atoms emitting spectral lines are subjected to a powerful electric field. The arrangement used by Stark is shown in Fig. 4.43. The canal rays are produced in an ordinary glass discharge tube provided with a perforated cathode  $C$ . When the pressure in the tube is not very low, discharge takes place between the anode  $A$  and cathode  $C$  maintained at a suitable P.D. The canal rays stream through the perforations in the cathode and form behind the cathode narrow cylindrical bundles of luminous rays. An auxiliary electrode  $F$  is placed parallel and close to  $C$  at a distance of a few millimetres. A very strong electric field of several thousand volts per metre is maintained between  $F$  and  $C$ . The effect produced can be studied both transversely [Fig. 4.43 (a)] and longitudinally [Fig. 4.43 (b)]. Stark observed that the lines in the spectrum emitted by the canal rays of hydrogen were split up into numerous sharp components under the action of the electric field.

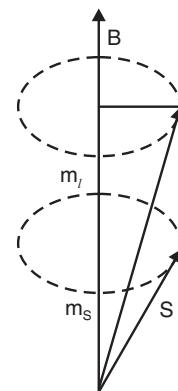


Fig. 4.42

**Results.** The results obtained with the lines of the Balmer series of the hydrogen spectrum are given below:

(i) Every line is split up into a number of sharp components. All hydrogen lines form symmetrical patterns. The pattern depends markedly on the quantum number  $n$  of the term involved. The number of lines and the total width of the pattern increases with  $n$ . Thus, the number of components of  $H_{\beta}$  line is greater than that of the  $H_{\alpha}$  line; similarly, the number of components of  $H_{\gamma}$  is greater than that of  $H_{\beta}$ .

(ii) Observation perpendicular to the direction of the electric field (transverse view) shows that the components are polarised, some parallel to the direction of the field and others perpendicular to it.

(iii) Up to fields of about  $10^7 \text{ V/m}$ , the resolution increases in proportion to the field strength ( $E$ ). In this region, we have *linear or first order Stark effect*. When  $E$  exceeds  $10^7 \text{ V/m}$ , there are shifts in the line patterns which are proportional to  $E^2$  and we speak of the *second-order Stark effect*.

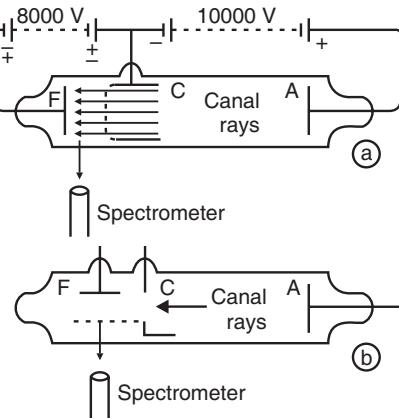


Fig. 4.43

## EXERCISES

### Short Answer Questions

1. State and explain the basic postulates of Bohr atom model. (Bharathiar, April 2012)
2. Write notes on critical potentials. (Bharathiar, April 2013)
3. Explain the various quantum numbers associated with vector atom model. (Bharathiar, April 2011)
4. Write a note on L-S and  $j-j$  coupling. (Bharathiar, April 2012)
5. Explain the periodic classification of elements. (B.U., April 2013)
6. Explain what is magnetic moment of an atom and derive an expression for it. Hence show that the ratio of the orbital magnetic moment to angular momentum of the atom is  $e/2m$ . (G.N.D.U. 2008)
7. What is Bohr magneton? Calculate its value. What is its physical significance? Is magnetic moment quantised? (G.N.D.U. 2008)
8. Explain the fine structure of sodium D-line. (Bharathiar University, April 2013)
9. State and explain normal Zeeman effect. (Bharathiar University, April 2014)
10. Derive an expression for Zeeman shift. (Bharathiar University, April 2011)
11. State and explain Larmor's theorem. (Bharathiar University, April 2011)
12. Explain Paschen-Back effect. (Bharathiar University, April 2012)
13. State and explain Stark effect. (Bharathiar University, April 2014)

### Long Answer Questions

14. Derive Rutherford's scattering formula. How do you estimate the nuclear dimensions from Rutherford scattering? (Meerut 1991, 86)
15. Stating Bohr's postulates, derive an expression for the frequencies of radiation emitted in hydrogen atom. Hence explain the various spectral series of this atom. (Rohilkhand 81)
16. Describe with necessary theory Davis and Goucher's method of determining the critical potentials of a gas. (B.U., April 2014)
17. Describe the Sommerfeld's relativistic atom model. (B.U., April 2013)

18. Describe vector atom model and explain different quantum numbers associated with it.  
*(B.U., April 2014)*
19. Give an account of different quantum numbers required to specify the state of electron in an atom. Describe Pauli's exclusion principle and use it to find distribution of electrons in different shells.  
*(Meerut 91)*
20. State and explain Pauli's exclusion principle. How is it used to explain the classification of elements according to periodic table?  
*(B.U., April 2012)*
21. Calculate the magnetic dipole moment of an electron due to orbital and spin motion.  
*(B.U., April 2012)*
22. Describe Stern-Gerlach experiment for the existence of space quantisation.  
*(B.U., April 2013)*
23. What is Zeeman effect? Discuss the quantum mechanical explanation of the normal Zeeman effect.  
*(B.U., April, 2012)*
24. Derive an expression for Lande's splitting factor and explain the anomalous Zeeman effect of sodium doublet lines D<sub>1</sub> and D<sub>2</sub> with its help.  
*(B.U., April, 2011)*

**Numerical Problems**

25. Calculate the closest distance of approach of an  $\alpha$ -particle of energy 5 MeV shot at a gold nucleus.  
**[Ans.**  $4.54 \times 10^{-14} \text{ m}$ **]**
26. How many revolutions does an electron in the  $n = 2$  state of a hydrogen atom make before dropping to the  $n = 1$  state? (The average lifetime of an excited state is about  $10^{-8} \text{ s}$ ).  
**[Ans.**  $8.16 \times 10^6$  revolutions**]**

**After reading this chapter, you should be able to**

- ◆ Distinguish between the linear and mass absorption coefficients of a material
- ◆ Discuss the Bragg's law of x-ray diffraction and utilize it to solve problems
- ◆ Explain the use of Bragg X-ray spectrometer to study the structure of crystals
- ◆ Distinguish between continuous and characteristic X-ray spectra
- ◆ Solve problems using Moseley's Law
- ◆ Apply classical, and quantum physics to examine and understand the process of scattering of X-rays by light elements (for example carbon)
- ◆ Discuss how the Compton effect established the particle nature of radiation.

## 5.1 PRODUCTION OF X-RAYS

**The Coolidge tube.** X-rays are produced when fast moving electrons are suddenly stopped by a solid target. A Coolidge tube is shown in Fig. 5.1. The tube is exhausted to the best possible vacuum of the order of  $10^{-5}$  mm of mercury. The cathode consists of a tungsten filament (*F*) heated by a low tension battery. Thermionic electrons emitted by the filament are accelerated towards the target (*T*) by a high P.D. maintained between *F* and *T*. The filament is placed inside a metal cup *G* to focus the electrons on to the target. The target must be cooled to remove the heat generated in it by continuous electron-bombardment. The usual method is to mount the target material on a hollow copper tube through which cold water is continuously circulated. The target is made of a metal like tungsten or molybdenum having a high melting point and a high atomic number. Metals with high atomic number give more energetic and intense X-rays when used as targets.

**Control of intensity and quality.** In the Coolidge tube, the intensity and frequency of X-rays can be easily controlled.

- (1) The intensity of X-rays depends on the number of electrons striking the target per second. The number of electrons given out by the filament is proportional to its temperature, which can be adjusted by varying the current in the filament circuit. Therefore, the intensity of X-rays varies with the filament current.
- (2) The frequency of X-rays emitted depends on the voltage between the cathode and the anode (target).

- Let  $V$  be the accelerating potential across the tube.
- $e$  is the charge on the electron.

$$\text{Work done on the electron in moving} \left. \begin{array}{l} \text{from the cathode to the anticathode} \end{array} \right\} = eV.$$

The electron thus acquires K.E. which is converted into X-rays, when the electron strikes the target.

- If  $v_{\max}$  is the maximum frequency of the X-rays produced, then

$$h v_{\max} = eV.$$

$$\therefore \text{The minimum wavelength produced by an X-ray tube} = \lambda_{\min} = \frac{c}{v_{\max}} = \frac{hc}{eV}.$$

- The minimum wavelength of the X-rays for a given voltage  $V$  across an X-ray tube is

$$\lambda_{\min} = \frac{hc}{eV}$$

$$\lambda_{\min} \propto \frac{1}{V}.$$

This is **Duane-Hunt law**.

- X-rays are electromagnetic waves of short wavelengths in the range of 1 nm to 0.05 nm. The longer wavelength end of the spectrum is known as the “soft X-rays” and the shorter wavelength end is known as “hard X-rays.”

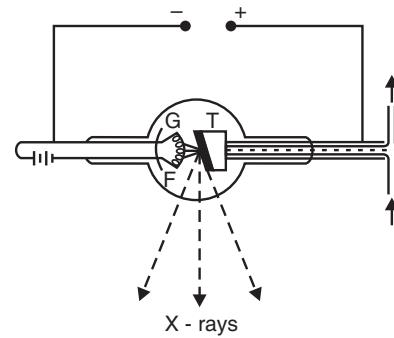


Fig. 5.1

## 5.2 THE ABSORPTION OF X-RAYS

All materials through which X-rays pass absorb them to some extent. If a sheet of any substance is interposed in the path of a homogeneous beam of X-rays, its intensity decreases.

- High frequency X-rays are absorbed less and are called *hard* X-rays.
- Low frequency X-rays are absorbed more and are called *soft* X-rays.

**Intensity of X-rays.** *The intensity of X-ray beam, at any point, is defined as the amount of energy carried per second per unit area, perpendicular to the direction of flow of energy.*

- Let  $I_0$  be the intensity of the incident X-ray beam.
- Let  $I$  be the intensity of the beam after it has traversed a thickness  $dx$  of the absorber.
- The decrease in intensity ( $dI$ ) follows the equation,  $dI = -\mu I dx$ .

Here  $\mu$  is called the linear absorption coefficient which depends on the wavelength of X-ray used and the nature of the absorbing material.  $\mu$  has the dimensions of reciprocal length ( $L^{-1}$ ) and its units are therefore  $m^{-1}$ .

$$\text{Now, } \frac{dI}{I} = -\mu dx$$

$$\therefore I = I_0 e^{-\mu x} \quad \dots(1)$$

$$\text{If } x = \frac{1}{\mu}, I = \frac{I_0}{e}.$$

*The linear absorption co-efficient of the material is, therefore, defined as the reciprocal of the thickness after which the intensity of X-rays falls to  $1/e$  of its original intensity.*

Eq. (1) may be put in the form

$$I = I_0 e^{-(\mu/\rho) \times \rho x} = I_0 e^{-\mu_m m}$$

$\mu/\rho = \mu_m$ , is called the mass absorption coefficient and  $m$  is the mass of unit area of the absorbing sheet. Theory shows that the mass absorption coefficient ( $\mu_m$ ) varies as the cube of the incident wavelength ( $\lambda$ ) and also the atomic number ( $Z$ ) of the absorber. Thus,  $\mu_m = k\lambda^3 Z^3$ . This explains why materials with high atomic numbers are preferred for shielding against X-rays.

**Experimental study.** An ionization chamber is used to study the penetrating ability of X-rays with apparatus shown in Fig. 5.2.

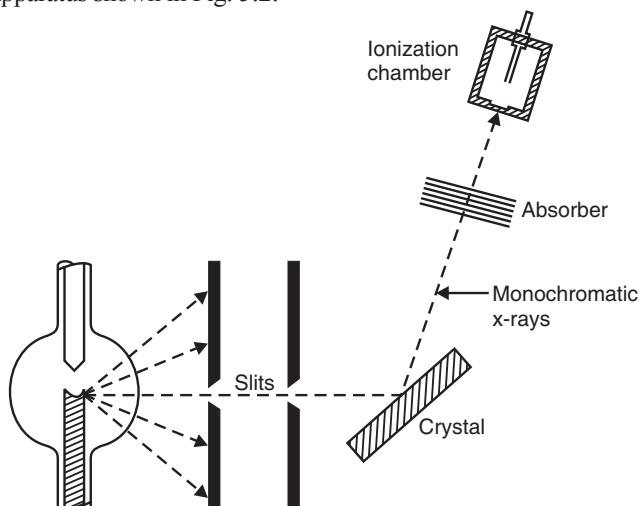


Fig. 5.2

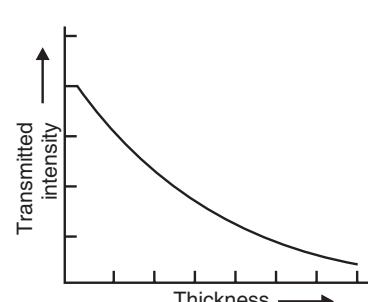


Fig. 5.3

X-rays are collimated by slits, rendered monochromatic by a Bragg reflection, and passed through the material under study. The thickness of the absorbing material is varied and the transmitted intensity is plotted against thickness (Fig. 5.3). This is an exponential decay curve ( $I = I_0 e^{-\mu x}$ ).

### 5.2.1. X-ray Absorption Edges

The value of the linear absorption coefficient  $\mu$  depends upon the X-ray wavelength and on the nature of the absorbing material. Let X-ray photons of sufficiently high energy (short wavelength) be made to fall upon a material. A major part of the incident energy of X-ray photons is spent in ejecting electrons from the  $K$ -shell of the atom. The energy of the X-ray beam is thus absorbed photo electrically, causing the ejection of the  $K$ -electrons. But if the incident radiation has a wavelength slightly longer (or the energy slightly smaller) than that required to eject the  $K$ -electrons, there is no photo-electric absorption. At this stage, there is a sudden fall in the value of the absorption coefficient (Fig. 5.4).

*A wavelength at which there is a sudden change in the absorption coefficient of a given material for the X-ray beam is called **absorption edge**.*

As the wavelength is further increased, the absorption coefficient again increases. The X-ray photons now start ejecting electrons from the  $L$ -shell. There is again a sudden decrease in the value of  $\mu$  at another definite wavelength  $\lambda = \lambda_{L_I}$ . Immediately after this there are two further discontinuities at two closely lying wavelengths  $\lambda_{L_{II}}$  and  $\lambda_{L_{III}}$ . In the case of ionization from the  $L$ -shell there are three absorption edges ( $L_I$ ,  $L_{II}$ ,  $L_{III}$ ), indicating that the electrons in the  $L$ -shell can exist in three energy sub-group states.

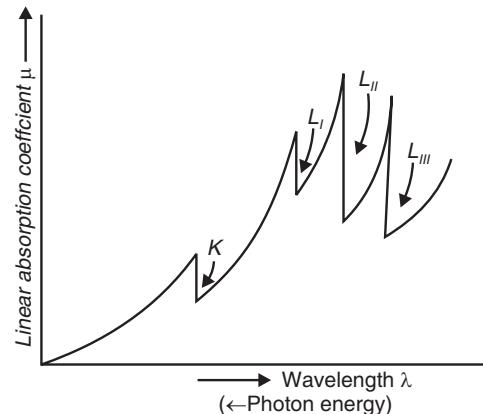


Fig. 5.4

## 5.3 BRAGG'S LAW

When monochromatic X-rays impinge upon the atoms in a crystal lattice, each atom acts as a source of scattering radiation of the same wavelength. The crystal acts as a series of parallel reflecting planes. The intensity of the reflected beam at certain angles will be maximum when the path difference between two reflected waves from two different planes is an integral multiple of  $\lambda$ .

**Derivation of Bragg's law.** Consider a set of parallel planes of atom points at a spacing  $d$  between two successive planes. Let a narrow monochromatic X-ray beam of wavelength  $\lambda$  be incident on the first plane at a glancing angle  $\theta$  (Fig. 5.5). Consider the ray  $PQ$  incident on the first plane. The corresponding reflected ray  $QR$  must also be inclined at the same angle  $\theta$  to the plane. Since X-rays are much more penetrating than ordinary light, there is only partial reflection at each plane. The complete absorption takes place only after penetrating several layers.

- Consider two parallel rays  $PQR$  and  $P'Q'R'$  in the beam, which are reflected by two atoms  $Q$  and  $Q'$ .  $Q'$  is vertically below  $Q$ . The ray  $P'Q'R'$  has a

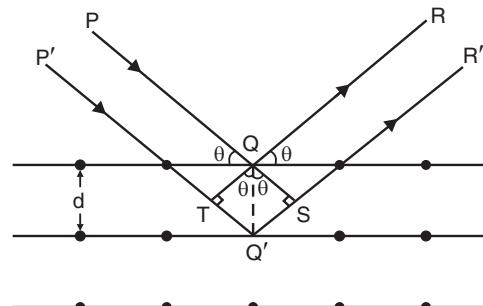


Fig. 5.5

longer path than the ray  $PQR$ . To compute the path-difference between the two rays, from  $Q$  draw normals  $QT$  and  $QS$  on  $P'Q'$  and  $Q'R'$  respectively.

$$\text{Path-difference} = TQ' + Q'S = d \sin \theta + d \sin \theta = 2d \sin \theta.$$

Hence the two rays will reinforce each other and produce maximum intensity, if

$$2d \sin \theta = n\lambda.$$

Here,  $n = 1, 2, 3, \dots$ . The integer  $n$  gives the order of the scattered beam,  $\lambda$  is the wavelength of the X-rays used. This equation is called *Bragg's law*.

## 5.4 THE BRAGG X-RAY SPECTROMETER

**Experimental arrangement.** Fig. 5.6 shows the essential parts of a Bragg spectrometer.

It consists of three parts. (1) a source of X-rays (2) a crystal held on a circular table which is graduated and provided with vernier and (3) a detector (ionisation chamber).

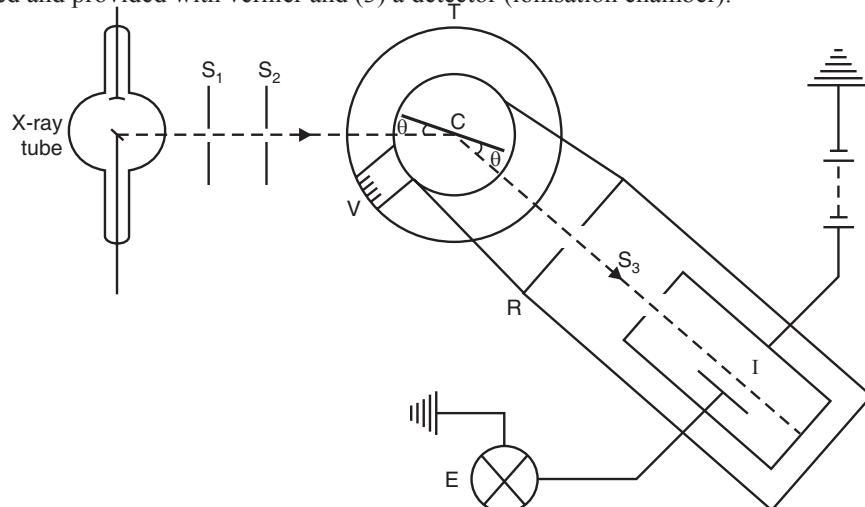


Fig. 5.6

- X-rays from an X-ray tube, limited by two narrow lead slits  $S_1$  and  $S_2$ , are allowed to fall upon the crystal  $C$ .
- The crystal is mounted on the circular table  $T$ , which can rotate about a vertical axis and its position can be determined by the vernier  $V$ .
- The table is provided with a radial arm ( $R$ ) which carries an ionisation chamber ( $I$ ). This arm also can be rotated about the same vertical axis as the crystal. The position of this arm can be determined by a second vernier (not shown in the figure). The ionisation chamber is connected to an electrometer ( $E$ ) to measure the ionisation current. Hence we can measure the intensity of the diffracted beam of X-rays, diffracted in the direction of the ionisation chamber.  $S_3$  is a lead slit, to limit the width of the diffracted beam. In practice, the crystal table is geared to the ionisation chamber so that the chamber turns through  $2\theta$  when the crystal is turned through  $\theta$ .

**Working.** To begin with, the glancing angle  $\theta$  for the incident beam is kept very small. The ionisation chamber is adjusted to receive the reflected beam till the rate of deflection is maximum. The glancing angle ( $\theta$ ) and the intensity of the diffracted beam ( $I$ ) are measured. The glancing angle is next increased in equal steps, by rotating the crystal table. The ionisation current is noted for

different glancing angles. The graph of ionisation current against glancing angle is drawn. The graph obtained is as in Fig. 5.7 and is called an **X-ray spectrum**.

- The prominent peaks  $A_1, A_2, A_3$  refer to X-rays of wavelength  $\lambda$ . The glancing angles  $\theta_1, \theta_2, \theta_3$  corresponding to the peaks  $A_1, A_2, A_3$  are obtained from the graph. It is found that

$$\sin \theta_1 : \sin \theta_2 : \sin \theta_3 = 1 : 2 : 3.$$

- This shows that  $A_1, A_2, A_3$ , refer to the first, second and third order reflections of the same wavelength.  $B_1, B_2, B_3$  are such peaks for the first, second and third order for another wavelength ( $\lambda_2$ ). Thus Bragg experimentally verified the relation

$$2d \sin \theta = n\lambda.$$

- The wavelength of X-rays is determined by using the equation  $2d \sin \theta = n\lambda$ . The glancing angle  $\theta$  is experimentally determined as explained already for a known order. If  $d$  is known,  $\lambda$  can be calculated.

**Calculation of  $d$ .** Rocksalt ( $NaCl$ ) possesses a cubic structure with sodium ions and chlorine ions situated alternately at corners of a cube. If  $d$  is the distance between two neighbouring ions and  $\rho$  is the density of the crystal, then mass of the unit cube =  $\rho d^3$ . Now each corner ion is shared by 8 neighbour cubes. This is illustrated by the ion indicated by an asterisk in Fig. 5.8. Hence each ion contributes only  $\frac{1}{8}$  th of its mass to the cube. The unit cube is made up of 4 sodium ions and 4 chlorine ions. Hence, the mass of each cube = the mass of  $\frac{1}{2} NaCl$  molecule. If  $M$  is the molecular weight of  $NaCl$  and  $N_A$  the Avogadro's No.,

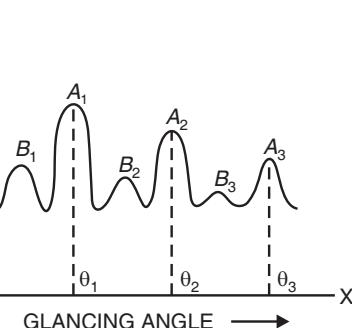


Fig. 5.7

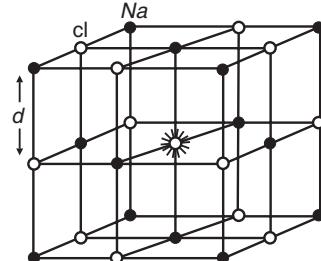


Fig. 5.8

$$\text{Mass of } \frac{1}{2} NaCl \text{ molecule} = M/2N_A$$

$$\therefore \rho d^3 = \frac{M}{2N_A} \text{ or } d = 3 \sqrt[3]{\frac{M}{2N_A \rho}}$$

Knowing  $M, N_A$  and  $\rho$ ,  $d$  can be found.

For  $NaCl$ ,  $M = 58.45 \text{ kg}$ ;  $\rho = 2170 \text{ kg m}^{-3}$  and  $N_A = 6.06 \times 10^{26}$ .

$$\therefore d = 3 \sqrt[3]{\frac{58.45}{2 \times (6.06 \times 10^{26}) \times 2170}} = 2.81 \times 10^{-10} \text{ m.}$$

**EXAMPLE 1.** The spacing between principal planes of  $NaCl$  crystal is  $2.82 \text{ \AA}$ . It is found that first order Bragg reflection occurs at an angle of  $10^\circ$ . What is the wavelength of X-rays?

**SOL.** By Bragg equation,  $2d \sin \theta = n\lambda$ .

Here,  $d = 2.82 \times 10^{-10} \text{ m}$ ;  $n = 1$  and  $\theta = 10^\circ$ .  $\lambda = ?$

$$\lambda = \frac{2d \sin \theta}{n} = \frac{2 \times (2.82 \times 10^{-10}) \sin 10^\circ}{1} = 0.98 \times 10^{-10} \text{ m.}$$

## 5.5 X-RAY SPECTRA

Urey and his co-workers analysed the X-ray beam emitted from an X-ray tube using different potential differences and same target. Using tungsten as the target and different potential differences, the intensities of the rays produced are plotted against wavelength. Fig. 5.9 represents these graphs. For applied P.D.s 30 kV, 40 kV and 50 kV, the spectrum is white. But, for the applied P.D. 70 kV, two sharp peaks are seen. The sharp peaks show the *line* or *characteristic radiation*. The line spectra are absent, till the P.D. is greater than a particular value. The smoothly varying curves represent the *continuous spectrum*. The superimposed lines on the continuous background constitute the *characteristic spectrum*.

### Main features of continuous X-ray spectrum

(1) For each anode potential, there is a minimum wavelength ( $\lambda_{min}$ ) below which no radiation is emitted. Above this critical value, the intensity of the radiation increases rapidly with increasing wavelengths and after reaching a maximum, decreases gradually. The intensity never reaches zero showing that the radiation contains all possible wavelengths above the minimum limit.

(2) When the voltage across the X-ray tube is increased,  $\lambda_{min}$  is shifted towards smaller values. Duane and Hunt showed that  $\lambda_{min}$  is inversely proportional to the applied voltage  $V$  or  $v_{max}$  is directly proportional to  $V$ . If the limiting frequencies ( $v_{max}$ ) are plotted against the applied voltages ( $V$ ), a straight line graph passing through the origin is obtained (Fig. 5.10). This empirical law of Duane and Hunt is expressed analytically as

$$eV = hv_{max} = \frac{hc}{\lambda_{min}}.$$

**Duane-Hunt Law: Statement.** The short wavelength limit ( $\lambda_{min}$ ) of the X-ray spectrum is inversely proportional to potential ( $V$ ) applied across the tube terminals.

$$\lambda_{min} \propto \frac{1}{V}.$$

**Explanation.**  $eV$  is the K.E. of the bombarding electron. If the entire K.E. of the electron striking the target is converted into the energy of the X-ray photon, then,  $eV = hv_{max}$  according to Einstein's theory.

But

$$v_{max} = \frac{c}{\lambda_{min}}$$

∴

$$eV = \frac{hc}{\lambda_{min}} \text{ or } \lambda_{min} = \frac{hc}{eV} \text{ or } \lambda_{min} \propto \frac{1}{V}.$$

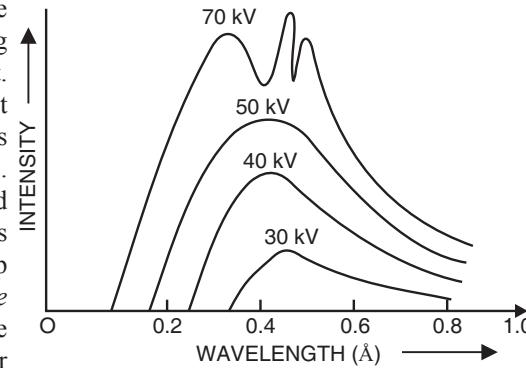


Fig. 5.9

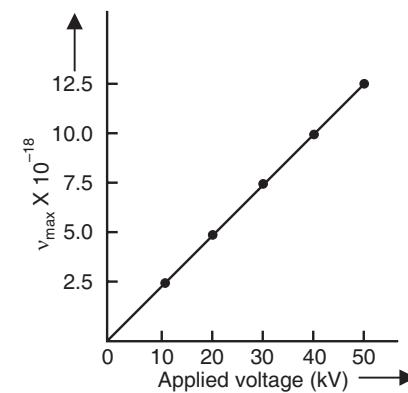


Fig. 5.10

Most of the electrons that generate X-ray photons give up only a part of their energy in this way. Therefore, most of the X-radiation is of longer wavelength than  $\lambda_{min}$ . Thus, the continuous spectrum is the result of the inverse photoelectric effect, with electron kinetic energy ( $eV$ ) being transformed into photon energy ( $h\nu$ ).

## 5.6 CHARACTERISTIC X-RAY SPECTRUM

There are two methods of producing characteristic X-rays.

(1) The characteristic X-rays of an element can be excited by using the element as the target in the X-ray tube and thus subjecting it to direct bombardment by electrons. For each target there is a minimum potential below which the line spectra do not appear. This critical P.D. below which the line spectra do not appear, is different for different targets. Molybdenum shows up the line spectra only if the P.D. is above 35 kV.

(2) Characteristic X-rays of an element can also be excited by allowing primary X-rays from a hard X-ray tube to fall on the element. The primary X-rays must be harder than the characteristic X-rays to be produced.

The peaks obtained in the X-ray spectrum (Fig. 5.12) give us the line spectrum which is characteristic of the element used in the target. The group of lines of shortest wavelength is called the *K*-series. Usually two lines of this series are detected. These lines are termed as  $K_\alpha$  and  $K_\beta$  lines in the order of decreasing wavelengths. The next group is called the *L*-series of longer wavelengths ( $L_\alpha$ ,  $L_\beta$ ,  $L_\gamma$ , etc.). For heavier elements a third series, called the *M*-series has been detected.

**Origin of characteristic X-rays.** This can be understood in terms of Bohr's theory. Suppose an atom in the target of an X-ray tube is bombarded by a high-speed electron and a *K*-electron is removed. A vacancy is created in the *K*-shell. This vacancy can be filled up by an electron from either of *L*, *M* or *N* shells or a free electron. These possible transitions can result in the  $K_\alpha$ ,  $K_\beta$  lines and the limiting line. Similarly, the longer wavelength *L*-series originates when an *L* electron is knocked out of the atom, the *M*-series when an *M* electron is knocked out and so on (Fig. 5.11).

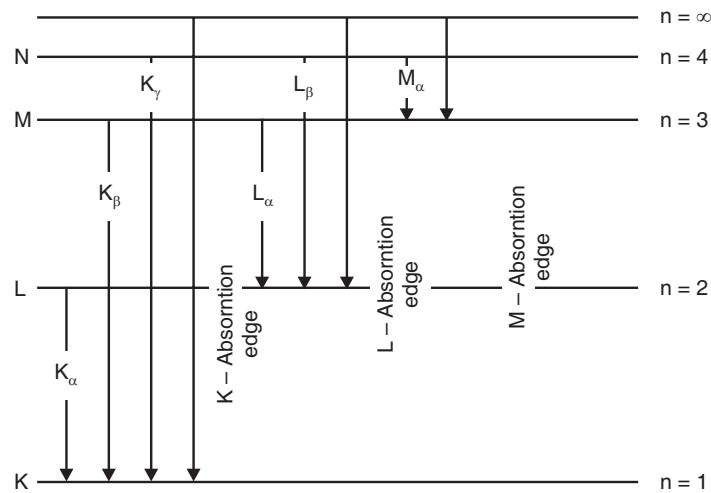
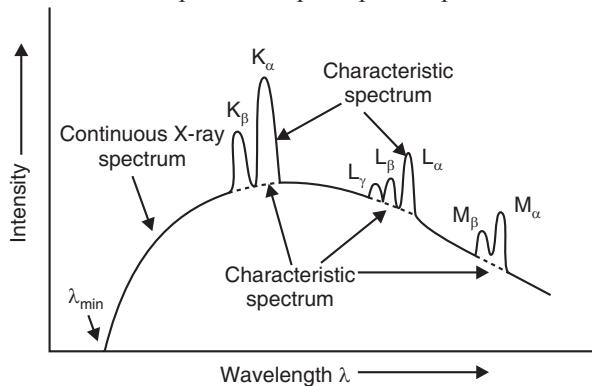


Fig. 5.11

It is clear that continuous spectra and line spectra both emitted by the same target are of different origin. The continuous spectrum is the result of the inverse photoelectric effect, with electron K.E.

being transformed into photon energy  $h\nu$ . The line spectrum has its origin in electronic transitions within atoms that have been disturbed by the incident electrons.

Fig. 5.12 shows a characteristic spectrum superimposed upon continuous spectrum.



**Fig. 5.12**

### 5.7 AUGER EFFECT

We have seen that when a  $K$ -electron is knocked out from an atom either by high-energy electron impact or through absorption of an X-ray photon, the vacancy thus created in the  $K$ -shell is filled up by transition from an outer-shell (say,  $L$ ) electron. The energy of the emitted characteristic x-ray photon is  $h\nu_{KL} = E_K - E_L$ .

But Auger found that there are certain radiationless transitions.

The excess energy in such transitions may be directly absorbed by another  $L$ -electron which gets emitted. There are now two vacancies created in  $L$ -shell and the energy level of the atom is different from  $L$ -level. If we designate this as  $LL$ -level of energy  $E_{LL}$ , then the kinetic energy of the emitted second electron is given by

$$\frac{1}{2}mv^2 = E_K - E_{LL}$$

Such a transition in which there is no emission of electromagnetic radiation, but emission of two electrons from the same atom is called **Auger transition** or **Auger effect**. It is also called *radiationless transition*.

Thus, the de-excitation of the atom may be accompanied either by the emission of a photon (characteristic radiation) or an electron (Auger electron).

### 5.8 SATELLITES

- Moseley and his contemporaries observed the X-ray spectral lines which were intense and easily resolvable. These lines are due to transitions between states of *single ionization* and can be easily fitted on an energy level diagram. Therefore, they are called *diagram lines*.
- Later on, with the improved resolving power of the instruments, many more lines were discovered. Most of them were rather faint and were usually found close to and on the short wavelength side of the more intense lines. Hence they were called “satellites”. They could not be fitted in the energy level diagram. Hence the new lines are called *nondiagram lines*.

### Characteristics of Satellite lines

1. Satellite lines have low intensity.
2. The excitation potential of certain satellites is greater than that of the corresponding "parent" line.
3. Siegbahn has shown that five satellites can be excited by the side of  $L_{\alpha_1}$  line of Mo by raising the excitation potential. Their appearance corresponds to singly, doubly, trebly etc., ionized atoms, since the intensity of ionisation can be increased by increasing the excitation potential.
4. The intensity of  $K$  satellite lines decreases in a continuous manner with increasing atomic number of radiating material. The intensity of  $L_{\alpha}$  satellites decreases abruptly as the atomic number increases from 47 to 50. The intensity again increases abruptly at 75. The  $L_{\alpha}$  satellites are absent between the atomic numbers 50 and 75.
5. If somehow or the other it is possible to avoid multiple ionisation, the satellites are found to be absent.

### Explanation of Satellite lines

The explanation of satellites was put forward by Wentzel and Druyvesteyn on the principle of *multiple ionisation of the inner electrons*. It is found in the case of  $K$  satellites that the energy of excitation is equal to the energy required to eject a  $K$  electron and in addition an  $L$  electron from the atom. Hence the initial state for the emission of  $K$  satellite is a state of double ionisation. In this case, the atom has an electronic vacancy both in  $K$  shell and  $L$  shell. Such a state of an atom is designated as  $KL$  atomic state. In a similar fashion, other states of double ionisation are designated as  $KK$ ,  $KM$ ,  $LL$ ,  $LM$  etc.

An atom in  $KL$  atomic state may undergo a radiative transition as  $KL \rightarrow KM$  (an electron jumping from  $M$  shell to  $L$  shell) or  $KL \rightarrow LL$  (an electron dropping from  $L$  shell to  $K$  shell). The estimates of atomic energy indicate that the change in energy in the transition  $KL \rightarrow LL$  is greater than that which takes place in the transition  $K \rightarrow L$  (the transition which gives rise to  $K_{\alpha}$  line). Hence the former transition gives rise to satellites on the short wavelength side of  $K_{\alpha}$  line. Similarly, the transition  $KL \rightarrow LM$  gives rise to satellites on short wavelength side of  $K_{\alpha}$  line. If now we assume that a cathode ray electron ejects two electrons at once from the atom, the probability of this ejection decreases with increasing atomic number of radiating material. This is found to be the case for  $K$  series.

## 5.9 MOSELEY'S LAW

Moseley plotted the square root of the frequencies ( $\sqrt{v}$ ) of a given line (say  $K_{\alpha}$ ) against the atomic numbers ( $Z$ ) of the elements emitting that line. Moseley obtained a straight line as shown in Fig. 5.13. The same linear relation was found to hold good for any line in any series. He concluded, therefore, that atomic number (and not atomic weight) is the fundamental property of elements.

**Moseley's law: statement.** *The frequency of a spectral line in X-ray spectrum, varies as the square of the atomic number of the element emitting it, or  $v \propto Z^2$ .* Moseley's law may be written as  $\sqrt{v} = a(Z - b)$ . Here,  $Z$  is the atomic number of the element and  $a$  and  $b$  are constants depending upon the particular line.

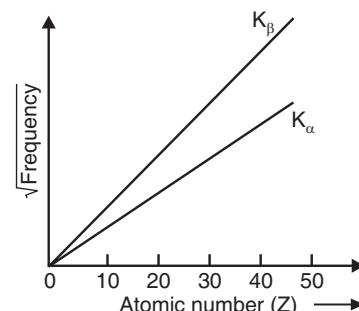


Fig. 5.13

**Explanation, according to Bohr's theory.** Bohr's theory of hydrogen spectrum gives the frequency of a spectral line as

$$\nu = Z^2 R c \left( \frac{1}{n_1^2} - \frac{1}{n_2^2} \right)$$

Here,  $R$  is Rydberg's constant and  $c$  the velocity of light.

$K_\alpha$  line originates from the transition of electron from second to first orbit.

Now,  $n_1 = 1$  and  $n_2 = 2$ .

$$\therefore \text{frequency of } K_\alpha \text{ line} = \nu = Z^2 R c \left( \frac{1}{1^2} - \frac{1}{2^2} \right) = \frac{3}{4} c R Z^2$$

This approximately corresponds to Moseley's law.

**Importance of Moseley's law.** (1) According to this law, it is the atomic number and not atomic weight of an element which determines its characteristic properties, both physical and chemical. Therefore, the atoms must be arranged in the periodic table according to their atomic numbers and not according to their atomic weights. This would remove some discrepancies in the order of certain elements from the point of view of their atomic weights. For example, argon  ${}_{18}\text{Ar}^{40}$  comes before potassium ( ${}_{19}\text{K}^{39}$ ), cobalt ( ${}_{27}\text{Co}^{58.9}$ ) comes before nickel ( ${}_{28}\text{Ni}^{58.7}$ ), etc. So the arrangement is correct in the order of atomic number.

(2) Moseley's work has also helped to perfect the periodic table by (i) the discovery of new elements, e.g., hafnium (72), illinium (61), masurium (43), rhenium (75), etc., and (ii) the determination of the atomic numbers of rare-earths and fixing their positions in the periodic table.

## 5.10 POLARISATION OF X-RAYS

**Experimental arrangement.** Fig. 5.14 shows the experimental arrangement for showing the polarisation of X-rays. A beam of X-rays travelling along  $AZ$  is allowed to strike at a scattering sheet  $Z$ . This scatters X-rays in all directions. With the help of slits  $S_3$  and  $S_4$ , a beam in the direction  $ZO$  which is perpendicular to the direction  $OX$  is isolated. The beam  $ZO$  strikes another scatterer at  $O$ . An ionization chamber  $I.C.$  is placed to receive X-rays in the directions  $OX$  and  $OY$ . It is observed that the intensity of X-rays scattered in the direction  $OX$  is maximum and is nearly zero in the  $OY$  direction. The beam  $ZO$  is said to be polarized.

Thus X-rays from the X-ray tube become completely polarized after having been scattered in the direction at right angles to the initial direction. The scatterer at  $Z$  behaves like a polarizer and that at  $O$  as an analyser.

## 5.11 SCATTERING OF X-RAYS (THOMSON'S FORMULA)

When an electromagnetic wave interacts with a single free electron, the electron performs simple harmonic motion and in so doing reradiates. The radiations thus emitted by the vibrating electron under the influence of the incident X-ray beam are called *scattered X-rays*.

At any point  $O$  within the material let  $E$  and  $B$  be the electric and magnetic vectors of the incident X-ray beam (Fig. 5.15). Both are perpendicular to the incident direction of the beam. The electron at the point  $O$  will experience an

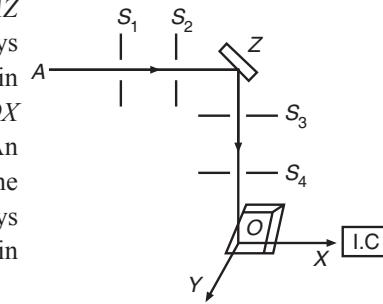


Fig. 5.14

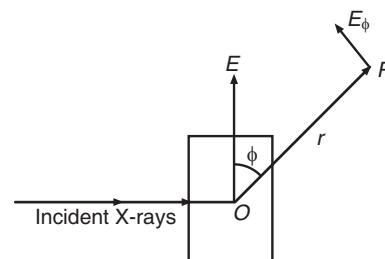


Fig. 5.15

acceleration ( $a$ ) in the direction of the electric field  $E$ .

$$a = eE/m \quad \dots (1)$$

According to classical electromagnetic theory, an accelerated charge radiates electromagnetic waves. The electric vector of the scattered X-rays at a point  $P$  with polar coordinates  $r$  and  $\phi$  is given by

$$E_\phi = \frac{e a \sin \phi}{4\pi\epsilon_0 c^2 r} = \frac{E e^2 \sin \phi}{4\pi\epsilon_0 m c^2 r} \quad \dots (2)$$

$$\therefore \frac{E_\phi}{E} = \frac{e^2 \sin \phi}{4\pi\epsilon_0 m c^2 r} \quad \dots (3)$$

Intensity is proportional to the square of the electric vector.

So the ratio of the scattered intensity at  $P$  to the incident intensity is

$$\frac{I_\phi}{I} = \left( \frac{E_\phi}{E} \right)^2 = \frac{e^4 \sin^2 \phi}{(4\pi\epsilon_0)^2 m^2 c^4 r^2} \quad \dots (4)$$

Consider a case where the incident beam is along the  $X$ -axis and its electric vector  $E$  is in the  $YZ$  plane (Fig. 5.16).

Resolve  $E$  into two perpendicular components along the  $Y$  and  $Z$  axes.

$$\begin{aligned} E^2 &= E_y^2 + E_z^2 \\ \therefore I &= I_y + I_z \end{aligned}$$

Here  $I_y$  and  $I_z$  are the components of the wave along the  $Y$  and  $Z$  axes.

The incident beam is unpolarized.

$$\therefore I_y = I_z = \frac{I}{2} \quad \dots (5)$$

Let  $I_1$  and  $I_2$  be the intensities of the scattered beam at the point  $P$  due to the  $y$  and  $z$  components.

$$\therefore I_1 = I_y \left[ \frac{e^4 \sin^2 \phi_1}{(4\pi\epsilon_0)^2 m^2 c^4 r^2} \right]$$

$$I_y = \frac{I}{2} \text{ and } \sin^2 \phi_1 = \cos^2 \theta$$

Here  $\theta$  is the angle between  $r$  and the  $x$ -axis.

$$\therefore I_1 = \frac{I}{2} \left[ \frac{e^4 \cos^2 \theta}{(4\pi\epsilon_0)^2 m^2 c^4 r^2} \right] \quad \dots (6)$$

$$\text{Similarly } I_2 = I_z \left[ \frac{e^4 \sin^2 \phi_2}{(4\pi\epsilon_0)^2 m^2 c^4 r^2} \right]$$

$$\text{But } I_z = \frac{I}{2} \text{ and } \sin^2 \phi_2 = 1.$$

Here the angle  $\phi_2$  between  $OP$  and  $OZ$  is always  $\pi/2$ .

$$\therefore I_2 = \left( \frac{I}{2} \right) \left[ \frac{e^4}{(4\pi\epsilon_0)^2 m^2 c^4 r^2} \right] \quad \dots (7)$$

The total intensity of the scattered wave at  $P$  is given by

$$I_e = I_1 + I_2$$

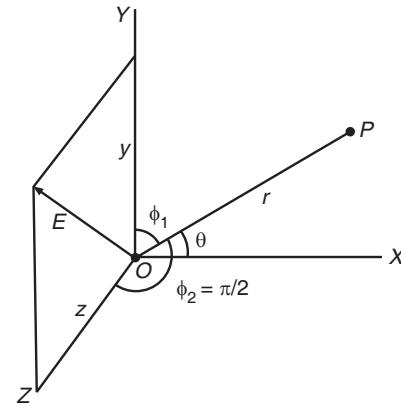


Fig. 5.16

$$I_e = \left(\frac{I}{2}\right) \left[ \frac{e^4 [1 + \cos^2 \theta]}{(4\pi\epsilon_0)^2 m^2 c^4 r^2} \right] \quad \dots (8)$$

Let  $n$  = number of electrons per unit volume of the scattering medium.

We assume that each electron is equally effective in producing scattered X-rays, irrespective of the presence of the others.

The intensity of the scattered X-rays at the point  $P$ , per unit volume of the scatterer is

$$\begin{aligned} I_s &= n I_e \\ I_s &= \frac{I}{2} \left[ \frac{n e^4 (1 + \cos^2 \theta)}{(4\pi\epsilon_0)^2 m^2 c^4 r^2} \right] \end{aligned} \quad \dots (9)$$

Equation (9) represents Thomson's scattering formula.

## 5.12 DETERMINATION OF THE NUMBER OF ELECTRONS PER ATOM

Consider a sphere of radius  $r$  with centre  $O$  (Fig. 5.17). The scatterer is at  $O$ .

Take two points  $A$  and  $B$  on the surface. The element  $AB$  subtends an angle  $d\theta$  at  $O$ .

Surface area of the element

$$\begin{aligned} &= 2\pi (AC)(AB) \\ &= 2\pi (r \sin \theta) r d\theta \\ &= 2\pi r^2 \sin \theta d\theta \end{aligned}$$

The total amount of energy scattered per unit volume of the scatterer in unit time is

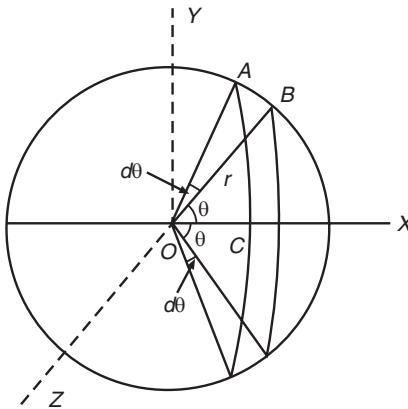


Fig. 5.17

$$U_s = \int_0^\pi I_s (2\pi r^2 \sin \theta) d\theta = \left(\frac{I}{2}\right) \left[ \frac{2\pi n e^4 r^2}{(4\pi\epsilon_0)^2 m^2 c^4 r^2} \right] \int_0^\pi (1 + \cos^2 \theta) \sin \theta d\theta \quad \dots (1)$$

$$U_s = \left(\frac{8\pi}{3}\right) \left[ \frac{I n e^4}{(4\pi\epsilon_0)^2 m^2 c^4} \right] \quad \dots (2)$$

The linear scattering coefficient  $\sigma$  of the material is

$$\sigma = \frac{U_s}{I} = \frac{8\pi}{3} \left[ \frac{n e^4}{(4\pi\epsilon_0)^2 m^2 c^4} \right] \quad \dots (3)$$

$$\therefore n = \left[ \frac{3}{8\pi} \right] \left[ \frac{(4\pi\epsilon_0)^2 m^2 c^4}{e^4} \right] \sigma \quad \dots (4)$$

By measuring  $\sigma$ , the value of  $n$  can be determined.

$n / \rho$  represents the number of electrons per unit mass.

$$\therefore \frac{n}{\rho} = \left[ \frac{3}{8\pi} \right] \left[ \frac{(4\pi\epsilon_0)^2 m^2 c^4}{e^4} \right] \left( \frac{\sigma}{\rho} \right) \quad \dots (5)$$

The ratio of  $\sigma$  to the density  $\rho$  of the scatterer is the mass scattering coefficient  $\sigma_m$ . Barkla and his coworkers found that the value of  $\sigma_m$  is  $0.02 \text{ m}^2/\text{kg}$  for carbon.

From Eq. (5), the number of electrons per kilogram is

$$\frac{n}{\rho} = 3 \times 10^{26} \text{ electrons/kg}$$

The number of carbon atoms per kilogram is

$$\frac{N_A}{A} = \frac{6.02 \times 10^{26} \text{ atoms/k mole}}{12 \text{ kg/k mole}} = 5 \times 10^{25} \text{ atoms/kg}$$

$$\left. \begin{array}{l} \text{The number of electrons} \\ \text{per carbon atom} \end{array} \right\} = \frac{3 \times 10^{26}}{5 \times 10^{25}} = 6.$$

The ratio of the power scattered to the primary intensity is called the *scattering cross section* (or *scattering coefficient*) of the free electron, designated by  $\sigma_e$ .

$$\sigma_e = \left( \frac{8\pi}{3} \right) \left( \frac{e^2}{4\pi\epsilon_0 mc^2} \right)^2$$

### 5.13 COMPTON SCATTERING

**The Compton Effect.** Compton discovered that when X-rays of a sharply defined frequency were incident on a material of low atomic number like carbon, they suffered a change of frequency on scattering. The scattered beam contains two wavelengths. In addition to the expected incident wavelength, there exists a line of longer wavelength. The change of wavelength is due to loss of energy of the incident X-rays. This elastic interaction is known as Compton effect.

In the case of incoherent scattering, a scattered beam undergoes not only deviation in its direction but also change of wavelength occurs. In Compton effect, there is a change in wavelength of the scattered beam along with the change in its direction. Hence Compton effect is an incoherent scattering.

This effect was explained by Compton on the basis of quantum theory of radiation. The whole process is treated as a particle collision event between X-ray photon and a loosely bound electron of the scatterer. In this process, both momentum and energy are conserved. In the photon-electron collision, a portion of the energy of the photon is transferred to the electron. As a result, the X-ray proceeds with less than the original energy (and therefore has a lower frequency or a higher wavelength).

The incident photon with an energy  $hv$  and momentum  $hv/c$  strikes an electron at rest. The initial momentum of the electron is zero and its initial energy is only the rest mass energy,  $m_0 c^2$ . The scattered photon of energy  $hv'$  and momentum  $hv'/c$  moves off in a direction inclined at an angle  $\theta$  to the original direction. The electron acquires a momentum  $mv$  and moves at an angle  $\phi$  to the original direction. The energy of the recoil electron is  $mc^2$  (Fig. 5.18).

According to the principle of conservation of energy,

$$hv + m_0 c^2 = hv' + mc^2 \quad \dots(1)$$

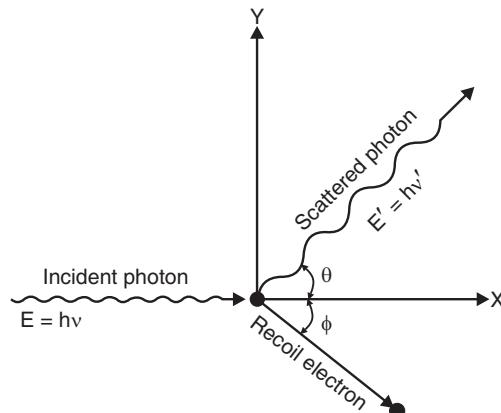


Fig. 5.18

Considering the  $x$  and  $y$  components of the momentum and applying the principle of conservation of momentum,

$$\frac{hv}{c} = \frac{hv'}{c} \cos \theta + mv \cos \phi \quad \dots(2)$$

$$\text{and} \quad 0 = \frac{hv'}{c} \sin \theta - mv \sin \phi \quad \dots(3)$$

$$\text{From (2), } mvc \cos \phi = h(v - v' \cos \theta) \quad \dots(4)$$

$$\text{From (3), } mvc \sin \phi = hv' \sin \theta \quad \dots(5)$$

Squaring and adding (4) and (5),

$$\begin{aligned} m^2 c^2 &= h^2 (v^2 - 2vv' \cos \theta + v'^2 \cos^2 \theta) + h^2 v'^2 \sin^2 \theta \\ &= h^2 (v^2 - 2vv' \cos \theta) + h^2 v'^2 = h^2 (v^2 - 2vv' \cos \theta + v'^2) \end{aligned} \quad \dots(6)$$

From (1),

$$\begin{aligned} mc^2 &= h(v - v') + m_0 c^2 \\ \therefore m^2 c^4 &= h^2 (v^2 - 2vv' + v'^2) + 2h(v - v') m_0 c^2 + m_0^2 c^4 \end{aligned} \quad \dots(7)$$

Subtracting (6) from (7),

$$m^2 c^2 (c^2 - v^2) = -2h^2 vv'(1 - \cos \theta) + 2h(v - v') m_0 c^2 + m_0^2 c^4 \quad \dots(8)$$

The value of  $m^2 c^2 (c^2 - v^2)$  can be obtained from the relativistic formula

$$\begin{aligned} m &= \frac{m_0}{\sqrt{(1 - v^2/c^2)}}. \text{ Squaring,} \\ m^2 &= \frac{m_0^2}{1 - v^2/c^2} = \frac{m_0^2 c^2}{c^2 - v^2} \\ \therefore m^2 c^2 (c^2 - v^2) &= m_0^2 c^4 \end{aligned} \quad \dots(9)$$

From (8) and (9),

$$\begin{aligned} m_0^2 c^4 &= -2h^2 vv'(1 - \cos \theta) + 2h(v - v') m_0 c^2 + m_0^2 c^4 \\ \therefore 2h(v - v') m_0 c^2 &= 2h^2 vv'(1 - \cos \theta) \\ \text{or} \quad \frac{v - v'}{vv'} &= \frac{h}{m_0 c^2} (1 - \cos \theta) \text{ or } \frac{1}{v'} - \frac{1}{v} = \frac{h}{m_0 c^2} (1 - \cos \theta) \\ \text{or} \quad \frac{c}{v'} - \frac{c}{v} &= \frac{h}{m_0 c} (1 - \cos \theta) \\ \text{or} \quad \lambda' - \lambda &= \frac{h}{m_0 c} (1 - \cos \theta) \end{aligned} \quad \dots(10)$$

$$\therefore \text{The change in wavelength} = d\lambda = \frac{h}{m_0 c} (1 - \cos \theta)$$

This relation shows that  $d\lambda$  is independent of the wavelength of the incident radiations as well as the nature of the scattering substance.  $d\lambda$  depends upon the angle of scattering only.

**Case 1.** When  $\theta = 0$ ,  $\cos \theta = 1$  and hence  $d\lambda = 0$

**Case 2.** When  $\theta = 90^\circ$ ,  $\cos \theta = 0$  and hence

$$d\lambda = \frac{h}{m_0 c} = \frac{6.63 \times 10^{-34}}{(9.11 \times 10^{-31}) \times (3 \times 10^8)} m = 0.0243 \text{ \AA}$$

This is known as *Compton wavelength*.

**Case 3.** When  $\theta = 180^\circ$ ,  $\cos \theta = -1$  and hence  $d\lambda = 2h/m_0c = 0.0485 \text{ \AA}$ .  $d\lambda$  has the maximum value at  $\theta = 180^\circ$ .

**Experimental verification.** Monochromatic X-rays of wavelength  $\lambda$  are allowed to fall on a scattering material like a small block of carbon (Fig. 5.19). The scattered X-rays are received by a Bragg spectrometer and their wavelength is determined. The spectrometer can freely swing in an arc about the scatterer. The wavelength of the scattered X-rays is measured for different values of the scattering angle. The experimental results obtained by Compton are shown in Fig. 5.20. In the scattered radiation in addition to the incident wavelength ( $\lambda$ ), there exists a line of longer wavelength ( $\lambda'$ ). The “*Compton shift*”  $d\lambda$  is found to vary with the angle at which the scattered rays are observed.

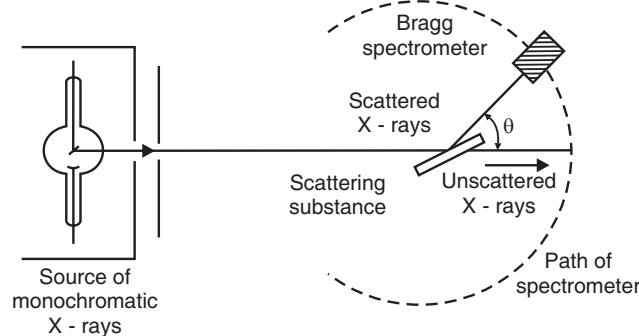


Fig. 5.19

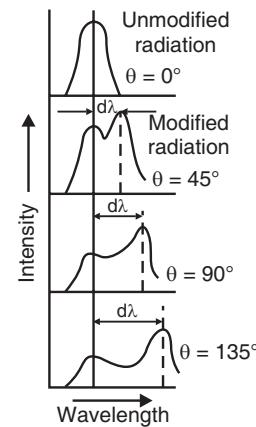


Fig. 5.20

**Direction of Recoil electron.** Dividing Eq. (5) by Eq. (4), we get

$$\tan \phi = \frac{hv' \sin \theta}{h(v - v' \cos \theta)} = \frac{v' \sin \theta}{(v - v' \cos \theta)}. \quad \dots(12)$$

Using Eq. (10), we get

$$\frac{1}{v'} = \frac{1}{v} + \frac{h}{m_0 c^2} (1 - \cos \theta) = \frac{1}{v} + \frac{h}{m_0 c^2} \cdot 2 \sin^2 \frac{\theta}{2}$$

or  $v' = \frac{v}{1 + \left( \frac{hv}{m_0 c^2} \right) 2 \sin^2 \frac{\theta}{2}} = \frac{v}{1 + 2\beta \sin^2 \left( \frac{\theta}{2} \right)}$  where  $\beta = \frac{hv}{m_0 c^2} \quad \dots(13)$

Substituting this value of  $v'$  in Eq. (12), we get

$$\begin{aligned} \tan \phi &= \frac{v \sin \theta / \left[ 1 + 2\beta \sin^2 \left( \frac{\theta}{2} \right) \right]}{\left[ v - \left\{ v \cos \theta / \left( 1 + 2\beta \sin^2 \frac{\theta}{2} \right) \right\} \right]} = \frac{\cot \left( \frac{\theta}{2} \right)}{(1 + \beta)} \\ \therefore \tan \phi &= \frac{\cot \left( \frac{\theta}{2} \right)}{1 + \left( \frac{hv}{m_0 c^2} \right)} \quad \dots(14) \end{aligned}$$

**Kinetic Energy of Recoil electron.** The K.E. of recoil electron is the difference between the energies of incident and scattered photons, *i.e.*,

$$\text{K.E.} = h\nu - h\nu'$$

$$\text{K.E.} = h\nu - h \left[ \frac{\nu}{1+2\beta \sin^2(\theta/2)} \right] = h\nu \left[ \frac{2\beta \sin^2(\theta/2)}{1+2\beta \sin^2(\theta/2)} \right] \quad \dots (15)$$

$$\text{where } \beta = h\nu / m_0 c^2$$

**EXAMPLE 1.** X-rays of wavelength  $0.7080 \text{ \AA}$  are scattered from a carbon block through an angle of  $90^\circ$  and are analysed with a calcite crystal, the interplanar distance of whose reflecting planes is  $3.13 \text{ \AA}$ . Determine the angular separation, in the first order, between the modified and the unmodified rays.

$$\begin{aligned} \text{SOL.} \quad \text{Wavelength of the modified rays} &= \lambda' = \lambda + \frac{h}{m_0 c} (1 - \cos \theta) \\ &= 0.7080 \times 10^{-10} \text{ m} + \left( \frac{6.63 \times 10^{-34}}{9.11 \times 10^{-31} \times 3 \times 10^8} \right) \text{ m} = 0.7323 \text{ \AA} \end{aligned}$$

Let  $\theta$  and  $\theta'$  be the angles of Bragg reflections corresponding to the wavelengths  $\lambda$  and  $\lambda'$ . Then, for  $n = 1$  (first order),

$$2d \sin \theta = n\lambda = 0.7080 \times 10^{-10} \text{ m}$$

and

$$2d \sin \theta' = n\lambda' = 0.7323 \times 10^{-10} \text{ m}.$$

Here,

$$d = 3.13 \times 10^{-10} \text{ m}; \therefore \theta = 6^\circ 30' \text{ and } \theta' = 6^\circ 43'.$$

$\therefore$  The angular separation, in the first order,  $\left. \right\} = \theta' - \theta = 13'$   
between the modified and unmodified rays

## EXERCISE

1. The minimum wavelength of continuous X-ray is given by ..... *(B.U., April 2012)*

[Ans.  $\lambda_{\min} = \frac{hc}{eV}$ ]

2. ..... is the equation of Bragg's law. *(B.U., April 2012)* [Ans.  $2d \sin \theta = n\lambda$ ]

3. According to Moseley's law, the frequency of a spectral line in X-ray spectrum is directly proportional to ..... *(B.U., April 2014)* [Ans. (b)]

(a)  $Z$

(b)  $Z^2$

(c)  $Z^3$

(d)  $Z^4$

4. X-rays are more penetrative than visible light. Why? *(B.U., April 2011)*

[Ans. Wavelength of X-rays is very small compared to the wavelength of visible light *i.e.*, energy of X-rays is very high compared to the energy of visible light. Hence X-rays are more penetrative than visible light.]

5. Explain the presence of unmodified line in Compton scattering. *(H.P.U. 1996)*

[Hint. When a photon collides with a bound electron, its wavelength does not change.]

6. Explain why Compton effect is experimentally not observed for visible light rays. *(H.P.U. 2001)*

[Ans.  $\Delta\lambda = \frac{h}{m_0 c} (1 - \cos \theta) = 0.024 (1 - \cos \theta)$ .  $\Delta\lambda$  depends only on the scattering angle  $\theta$ . Maximum value of  $\cos \theta = -1$  when  $\theta = 180^\circ$ . Hence the maximum value of  $\Delta\lambda = 0.048 \text{ \AA}$ . This means that Compton effect can be detected only for those radiations whose wavelength is not greater than a few \text{\AA}. For visible light ( $\lambda \approx 5000 \text{ \AA}$ ),  $(\Delta\lambda)_{\max}$  is only about 0.001% of the initial wavelength which cannot be detected.]

- 
7. Explain and derive Bragg's law of X-ray diffraction from a crystalline solid. (Meerut U. 2001)
  8. Describe and explain the Bragg's X-ray spectrometer method of determining wavelength of X-rays. (B.U., April 2012)
  9. Explain the continuous and characteristic spectra of X-rays. (B.U., April 2011)
  10. State and explain Mosley's law. What is its importance? (B.U., April 2013)
  11. Discuss the polarisation of X-rays. (B.U., April 2012)
  12. Derive the Thomson's formula for the scattering of X-rays. (B.U., April 2012)
  13. State and explain Compton effect. (B.U., April 2014)
  14. What is Compton effect? Derive an expression for the change of wavelength. (B.U., April 2012)
  15. What is Compton effect? Obtain an expression for the shift in wavelength of X-ray beam. Give its experimental verification. (Meerut 90, Agra 85, 86, 87, Garhwal 92)
  16. The P.D. across an X-ray tube is  $10^5$  volt. What is the maximum frequency of X-rays emitted? What is the corresponding wavelength ? [Ans.  $2.4 \times 10^{19}$  Hz;  $1.25 \times 10^{-11}$  m]
  17. *NaCl* has its principal planes spaced at  $2.820\text{ \AA}$ . The first order of Bragg reflection is located at  $10^\circ$ . Calculate (a) the wavelength of the X-rays and (b) the angle for the second order Bragg reflection. [Ans.  $0.9790\text{ \AA}$ ;  $20^\circ 19'$ ]
  18. The  $K_\alpha$  line from molybdenum has a wavelength of  $0.7078\text{ \AA}$ . Calculate the wavelength of the  $K_\alpha$  line of copper. Atomic number of molybdenum = 42. Atomic number of copper = 29.  
[Hint. From Moseley's law,  $(Z-1)^2 \lambda = \text{constant}$ . Let  $\lambda'$  be the wavelength of  $K_\alpha$  line of copper. Then,  $(42-1)^2 \times 0.7078 = (29-1)^2 \lambda' \therefore \lambda' = 1.517\text{ \AA}$ .]
  19. Monochromatic X-radiation of wavelength  $0.124\text{ \AA}$ , undergoes Compton effect from a carbon block. Calculate the wavelength that is scattered through  $180^\circ$ . [Ans.  $0.1724\text{ \AA}$ .]
  20. Photon of energy  $1.02\text{ MeV}$  undergoes Compton scattering through  $180^\circ$ . Calculate the energy of the scattered photon.  
[Hint. Energy of the photon =  $E = 1.02\text{ MeV} = 1.02 \times (1.6 \times 10^{-13}) J$ .  
$$\therefore \lambda = \frac{hc}{E} = 0.012\text{ \AA}; d\lambda = 0.048\text{ \AA}$$
  
$$\therefore \text{Wavelength of scattered photon} = \lambda' = 0.060\text{ \AA}$$
  
Hence the energy of the scattered photon =  $\frac{hc}{\lambda'} = 0.204\text{ MeV}$ ]

# PHOTOELECTRIC EFFECT AND PLANCK'S QUANTUM THEORY

**After reading this chapter, you should be able to**

- ◆ Demonstrate a working, quantitative understanding of the photoelectric effect
- ◆ Describe a typical photoelectric-effect experiment
- ◆ Formulate Einstein's light quantum hypothesis and explain the laws of photoelectric emission
- ◆ List some photoelectric devices and explain their performance
- ◆ Discuss how Planck explained the blackbody radiation by postulating the quantization of radiant energy
- ◆ Explain Max Planck's contribution to the development of quantum mechanics.

## 6.1 INTRODUCTION

Whenever light or electromagnetic radiations (such as X-rays, Ultraviolet rays) fall on a metal surface, it emits electrons. *This process of emission of electrons from a metal plate, when illuminated by light of suitable wavelength, is called the photoelectric effect.* The electrons emitted are called the *photoelectrons*. In the case of alkali metals, photoelectric emission occurs even under the action of visible light. Zinc, cadmium etc., are sensitive to only ultraviolet light.

### The Nature of Photo-particles

**Experimental arrangement.** The apparatus consists of two plates *A* and *C* placed in an evacuated quartz bulb (Fig. 6.1).

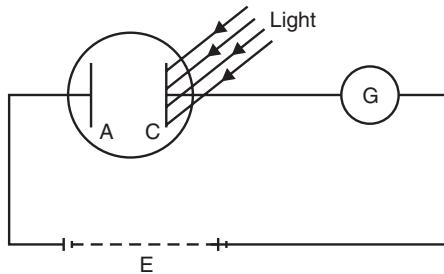


Fig. 6.1

The galvanometer (*G*) and battery (*E*) are connected as shown. When ultraviolet light is incident on the negative plate *C*, a current flows in the circuit as indicated by the galvanometer. But when light falls on the positive plate *A*, there is no current in the circuit. These observations show that photo-particle must be negatively charged.

## 6.2 LENARD'S METHOD TO DETERMINE E/M FOR PHOTOELECTRONS

**Apparatus.** The apparatus used is shown in Fig. 6.2. It consists of a glass tube *G* which can be evacuated through the side tube *T*. Ultraviolet light passes through a quartz window *W* and falls on an aluminium cathode *C* enclosed in *G*. An earthed metal screen *A* with a small central hole forms the anode. The cathode (*C*) can be maintained at a desired potential, positive or negative relative to the anode *A*. *P*<sub>1</sub> and *P*<sub>2</sub> are small metal electrodes connected to electrometers *E*<sub>1</sub> and *E*<sub>2</sub>.

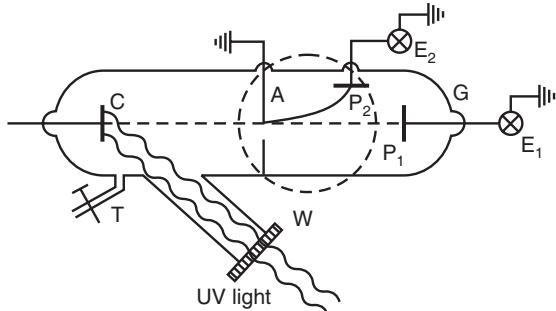


Fig. 6.2

**Working.** When *C* is raised to a negative potential and illuminated, negatively charged particles are produced and accelerated towards the anode. A few particles pass through the hole in *A* and proceed with uniform velocity to *P*<sub>1</sub>. Their arrival at *P*<sub>1</sub> is indicated by *E*<sub>1</sub>. By applying a uniform magnetic field *B* (represented by the dotted circle) perpendicular to the plane of the figure and directed towards the reader, the photoelectrons can be deflected towards *P*<sub>2</sub>. Their arrival at *P*<sub>2</sub> is indicated by the deflection they produce in *E*<sub>2</sub>.

### Variation of photoelectric current with cathode potential

Lenard first studied the relation between current and the potential applied to  $C$ . When the cathode potential was several volts positive, the current was zero. When  $V$  was +2 volts, there was a feeble current showing that a few particles possessed enough velocity to overcome the retarding potential of 2 volts. When the potential was further decreased, the current increased and reached a saturation value for -20 volts. Fig. 6.3 shows the variation of photoelectric current with cathode potential.

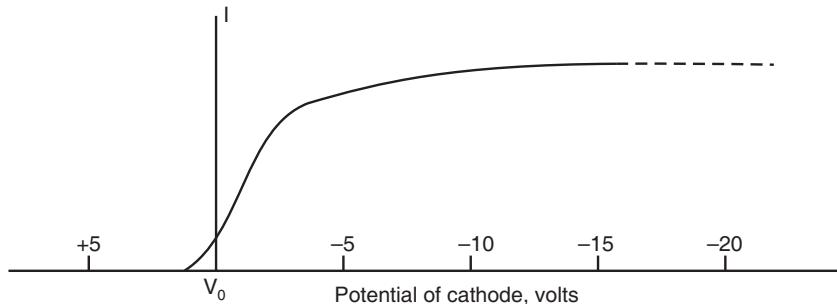


Fig. 6.3

**Determination of  $e/m$ .** After this preliminary investigation, Lenard applied to  $C$  a negative potential  $V$ , very large compared to the potential of 2 volts. The velocity imparted by the accelerating potential is so large that the velocity of the particles in the act of emission is negligible in comparison to it. Let  $V$  be the accelerating potential and  $v$  the velocity acquired by the photoelectrons. Then,

$$\frac{1}{2}mv^2 = eV \quad \dots(1)$$

where  $e$  is the charge and  $m$  the mass of the photoelectron.

Let  $R$  be the radius of the circular path described by the photoelectrons in the region of uniform magnetic field of strength  $B$ .

$$\begin{aligned} \text{Then } \frac{mv^2}{R} &= Bev. \\ \therefore v &= \frac{BeR}{m} \end{aligned} \quad \dots(2)$$

Substituting this value of  $v$  in Eq. (1),  $\frac{1}{2}m(BeR/m)^2 = eV$

$$\therefore \frac{e}{m} = \frac{2V}{B^2 R^2} \quad \dots(3)$$

Knowing  $V$ ,  $B$  and  $R$ ,  $e/m$  is calculated. Lenard found the value of  $e/m$  to be the same as that for electrons. *This clearly shows that the photoparticles are nothing but electrons.*

### 6.3 RICHARDSON AND COMPTON EXPERIMENT

**Apparatus:** The apparatus used by them is shown in Fig. 6.4. The emitter of photoelectrons ( $C$ ) is a small strip of the metal under study and is kept at the centre of a spherical glass bulb  $B$ .  $B$  is silvered on the inner side and can be evacuated through the tube  $T$ . The silver coating on the inside of the bulb serves as the anode  $A$  and is connected to an electrometer  $E$  which measures the photoelectric current. Monochromatic light  $L$  is made to pass through a quartz window  $W$  and fall on  $C$ .  $C$  can be maintained at any desired potential  $V$  relative to  $A$  and this potential can be read with a voltmeter.

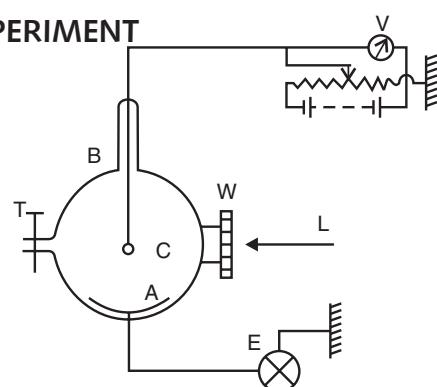


Fig. 6.4

(i) **Relation between photoelectric current and retarding potential.** They first studied the relation between the photoelectric current  $I$  and the retarding potential  $V$ . Irradiating the cathode  $C$  with monochromatic light of a given intensity, the cathode potential ( $V$ ) was varied from a few volts positive upto zero and negative values. The photoelectric current  $I$  was measured for different values of  $V$ . The observations were repeated with the intensity of illumination ( $X$ ) doubled, trebled, etc.

**Graph.** Fig. 6.5 shows the relation between  $I$  and  $V$ .

- For a given intensity of illumination ( $X_1$ ), there is no photoelectric current when the positive potential on the cathode is greater than a critical value  $+V_0$ . At potentials just less than  $+V_0$ , a small current is produced. As the potential decreases to zero, the current rapidly increases and reaches a maximum value when  $V = 0$ . No further increases in current are observed when  $V$  becomes negative. Hence  $I_1$  is the maximum current due to illumination of intensity  $X_1$ .
- When the intensity is doubled ( $X_2$ ) as in (2), the maximum current  $I_2$  is double  $I_1$ : But the critical potential  $V_0$  is the same as before.
- When the intensity of illumination is increased three fold ( $X_3$ ), the corresponding maximum current increases proportionately, but the critical potential remains unchanged.

**Result.** The maximum current  $I_m$  is proportional to the intensity of illumination  $X$ , i.e.,  $I_m \propto X$ . The critical potential  $V_0$  is independent of the intensity.

(ii) **Relation between velocity of photoelectrons and the frequency of light.** Several monochromatic radiations of wavelengths  $\lambda_1, \lambda_2, \lambda_3$ , etc., are allowed to fall on the emitter. The intensity of illumination for each wavelength is adjusted to give the same value of  $I_m$  in each case. In each case, the photoelectric current  $I$  for different values of  $V$  is determined.

**Graph.** Fig. 6.6 shows the nature of the curves.

- If  $\lambda_1 > \lambda_2 > \lambda_3$ , the critical potentials are  $V_1 < V_2 < V_3$ . Hence, as the wavelength of light increases, the critical retarding potential decreases. This means that the maximum K.E. of the photoelectrons (given by  $\frac{1}{2}mv_m^2 = eV_0$ ) increases with increasing frequency of light.

- Since  $V_0$  is independent of the intensity of illumination, we conclude that the velocity and K.E. of photoelectrons are independent of the intensity of illumination, but dependent on the frequency of the incident light.

A linear relation is found to exist between the maximum energy of emission ( $eV_0$ ) and the frequency ( $v$ ) of the light. If  $eV_0$  is plotted against  $v$ , we get, for any emitting surface a straight line (Fig. 6.7), whose intercept on the frequency axis gives the threshold frequency  $v_0 = c/\lambda_0$  for this surface. The slope of the straight line is

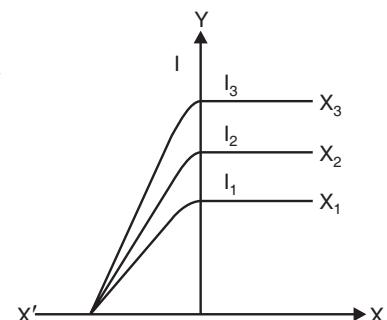


Fig. 6.5

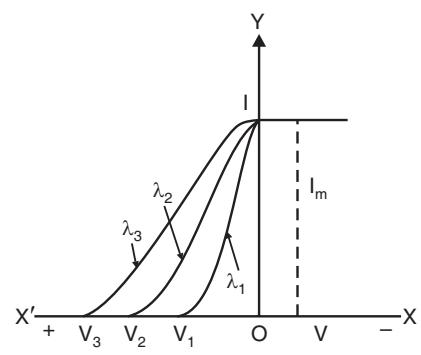


Fig. 6.6

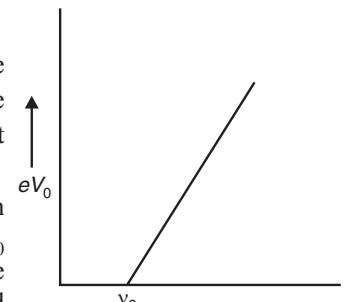


Fig. 6.7

the same for all emitting surfaces and is found to be equal to  $h$ . The value of  $v_0$  however, depends on the emitting surface.  $v_0$  is called the threshold frequency, as it represents the beginning of the photoelectric activity of the emitter.

**Definition.** Threshold frequency is defined as the minimum value of frequency of incident light below which the photo-electric emission stops completely, howsoever high the intensity of light may be.

- At threshold frequency, the K.E. of emitted photo-electrons is just zero.

## 6.4 EXPERIMENTAL INVESTIGATIONS ON THE PHOTOELECTRIC EFFECT

**Apparatus:** Photoelectric effect can be studied in detail with the help of the apparatus shown in Fig. 6.8.

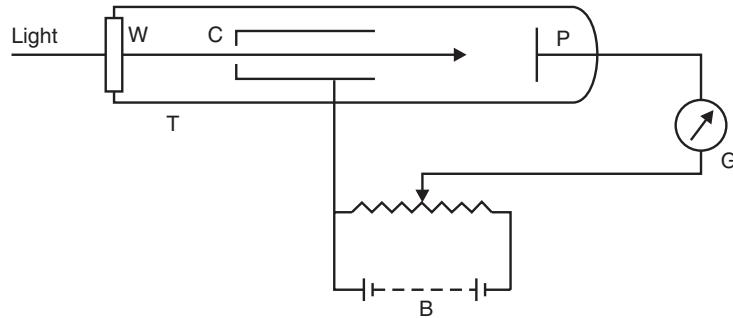


Fig. 6.8

It consists of an evacuated glass tube  $T$  with a quartz window  $W$ .  $P$  is a photoelectrically sensitive plate.  $C$  is a hollow cylinder and it has a small hole that permits the incident light to fall on the plate  $P$ .  $P$  is connected to the negative end.  $C$  is connected to the positive terminal of a battery  $B$ .

**Working.** When light from some source falls on the plate  $P$ , the photoelectrons are ejected out of the plate  $P$ . These photoelectrons are attracted by the positively charged cylinder  $C$ . Hence a photoelectric current flows from  $P$  to  $C$  in the bulb and from  $C$  to  $P$  outside the bulb. This current can be measured from the deflection produced in the galvanometer  $G$ . It is found that the strength of the photoelectric current increases as the potential of  $C$  is more and more positive with respect to  $P$ . The deflection in  $G$  decreases when the potential of  $C$  is negative with respect to  $P$ . The results obtained are summarised into four statements. They are called the laws of photoelectric emission.

**Laws of photoelectric emission.** (i) For every metal, there is a particular minimum frequency of the incident light, below which there is no photoelectric emission, whatever be the intensity of the radiation. This minimum frequency, which can cause photoelectric emission, is called the threshold frequency.

(ii) The strength of the photoelectric current is directly proportional to the intensity of the incident light, provided the frequency is greater than the threshold frequency.

(iii) The velocity and hence the energy of the emitted photoelectrons is independent of the intensity of light and depends only on the frequency of the incident light and the nature of the metal.

(iv) Photoelectric emission is an instantaneous process. The time lag, if any, between incidence of radiation and emission of the electrons, is never more than  $3 \times 10^{-9}$  sec.

**Failure of the electromagnetic theory.** The above experimental facts could not be explained on the basis of the electromagnetic theory of light.

(1) Calculations showed that it would require about 500 days to dislodge a photoelectron from sodium by exposure to violet light of wavelength 4000 Å. Experimentally, however, we observe that electron ejection commences without delay.

(2) According to the classical theory, light of greater intensity should impart greater K.E. to the liberated electrons. But, this does not happen. Also, the velocity of the emitted electron should not depend on the frequency of the incident light. But it does.

The phenomenon was adequately explained by Einstein on the basis of Planck's Quantum theory of radiation.

**Quantum theory.** According to Planck, the energy of a monochromatic wave with frequency  $\nu$  can only assume those values which are integral multiples of energy  $h\nu$ . i.e.,  $E_n = nh\nu$ , where  $n$  is an integer referring to the number of "Photons". Thus the energy of a single *PHOTON* of frequency  $\nu$  is  $E = h\nu$ .

## 6.5 EINSTEIN'S PHOTOELECTRIC EQUATION

According to Einstein, light of frequency  $\nu$  consists of a shower of corpuscles or photons each of energy  $h\nu$ . When a photon of light of frequency  $\nu$  is incident on a metal, the energy is completely transferred to a free electron in the metal. A part of the energy acquired by the electron is used to pull out the electron from the surface of the metal and the rest of it is utilised in imparting K.E. to the emitted electron. Let  $\phi$  be the energy spent in extracting the electron from the emitter to which it is bound (photoelectric work function) and  $\frac{1}{2}mv^2$  the K.E. of the photoelectron.

$$\text{Then } h\nu = \phi + \frac{1}{2}mv^2 \quad \dots(1)$$

This relation is called the *Einstein's Photoelectric equation*. If  $\nu_0$  is the threshold frequency which just ejects an electron from the metal without any velocity then,  $\phi = h\nu_0$ .

$$\therefore h\nu = h\nu_0 + \frac{1}{2}mv_{\max}^2 \quad \dots(2)$$

Here,  $v_{\max}$  is the maximum velocity acquired by the electron.

$$\text{or } \frac{1}{2}mv_{\max}^2 = h(\nu - \nu_0) \quad \dots(3)$$

- The work function of a metal is defined as the energy which is just sufficient to liberate electrons from the metal surface with zero velocity.
- Equation (3) suggests that the energy of the emitted photoelectrons is independent of the intensity of the incident radiation but increases with the frequency.

### Experimental verification of Einstein's Photoelectric Equation—Millikan's Experiment.

**Theory.** Millikan's experiment is based on the "stopping potential". The stopping potential is the necessary retarding potential difference required in order to just halt the most energetic photoelectron emitted.

$$\left. \begin{array}{l} \text{The K.E. of a photoelectron leaving} \\ \text{the surface of a metal irradiated with} \\ \text{light of frequency } \nu \end{array} \right\} = \frac{1}{2}mv_{\max}^2 = h\nu - \phi.$$

Let  $V$  be the P.D. which is applied between the emitter and a collecting electrode in order to prevent the photoelectron from just leaving the emitter, the emitter being maintained at a positive potential with respect to the collector. Then,

$$eV = \frac{1}{2}mv_{\max}^2$$

$$\therefore eV = h\nu - \phi$$

or

$$V = \frac{h}{e}v - \frac{\phi}{e} \quad \dots(1)$$

$\phi$  is constant for a given metal;  $h$  and  $e$  are also constants.

Hence, Eq. (1) represents a straight line.  $V$  is measured for different values of  $v$ . A graph is then plotted between the stopping potential ( $V$ ) taken along the Y-axis and the frequency of light ( $v$ ) taken along the X-axis. The graph is a straight line (Fig. 6.9). The slope of the straight line

$$\tan \theta = \frac{h}{e}$$

$\therefore$

$$h = e \tan \theta \quad \dots(2)$$

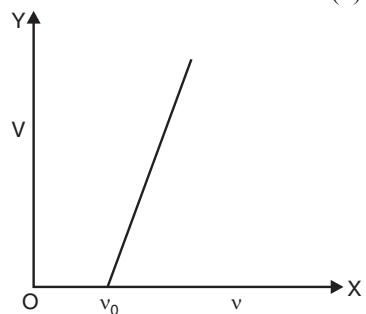


Fig. 6.9

Hence the value of  $h$  (Planck's constant) can be calculated. The intercept on the X-axis gives the threshold frequency  $v_0$  for the given emitter. From this, photoelectric work function  $= \phi = h v_0$  is calculated.

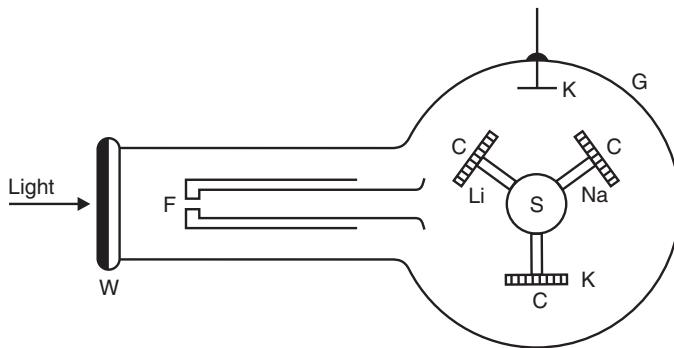


Fig. 6.10

**Experiment.** Millikan's apparatus is shown in Fig. 6.10. Alkali metals are employed as emitters, since they readily exhibit photoelectric emission even with visible light. Cylindrical blocks ( $C$ ) of sodium, potassium or lithium are mounted on a spindle  $S$  at the centre of the glass flask  $G$ . The flask is evacuated to a very high vacuum to free the metals from all absorbed gases and to prevent their oxidation. The spindle can be rotated from outside by an electromagnet. As each metal block passes by the adjustable sharp edge  $K$ , a thin layer of it is removed, thus exposing a fresh surface of the metal to the irradiating light entering the flask through a quartz window  $W$ . Monochromatic light provided by a spectroscope is used to illuminate the fresh metal surfaces. The photoelectrons are collected by a Faraday cylinder  $F$ . The Faraday cylinder is made of copper oxide which is not photosensitive. The photocurrent is measured by an electrometer connected to the Faraday cylinder.

*Robert Andrews Millikan (1868–1953)*

The stopping potential of the liberated photoelectrons is measured by raising the emitter surface to a positive potential, just sufficient to prevent any of the electrons from reaching the collector ( $F$ ). The stopping potential is the positive potential applied to the emitter, which corresponds to zero current in the electrometer. The stopping potential ( $V$ ) is determined for different wavelengths of the incident light. The value of  $V$  should be corrected for any contact potential between the metal ( $C$ ) and Faraday cylinder ( $F$ ). On plotting  $V$  against  $v$ , we get a straight line. Measuring the slope of the



straight line, the value of  $h/e$  is obtained. Then substituting the known value of  $e$ ,  $h$  is calculated. The value of  $h$  calculated in this way agrees fairly well with the value obtained by other methods. Thus the Einstein's equation can be verified experimentally.

## 6.6 PHOTOELECTRIC CELLS

*Photoelectric cell is an arrangement to convert light energy into electrical energy.* There are three types of photocells, photoemissive, photovoltaic and photoconductive.

### (i) Photo-emissive Cell.

**Construction.** This consists of a glass or quartz bulb ( $B$ ) according as it is to be used with visible or ultraviolet light (Fig. 6.11).  $C$  is the silver cathode in the form of a semi-cylindrical plate. The anode ( $A$ ) is a rod mounted vertically at the centre of the bulb and parallel to its axis. A positive potential of about 100 volts is applied to the anode, the negative being connected to the cathode through a galvanometer  $G$ .

**Working.** When light falls on the cathode  $C$ , electrons are ejected from the cathode. A small current flows through the cell and can be measured by the galvanometer.

- The photoemissive cell is used for reproduction of sound from photo-films.

### (ii) Photo-voltaic Cell.

**Construction.** It consists of a layer of semiconductor material spread over a metallic surface by heat treatment. The metal plate is made of copper. The semiconductor is cuprous oxide ( $Cu_2O$ ).

On the other side of the semiconductor, there is a very thin layer of a translucent deposit which allows the semiconductor to be illuminated by radiations (Fig. 6.12).

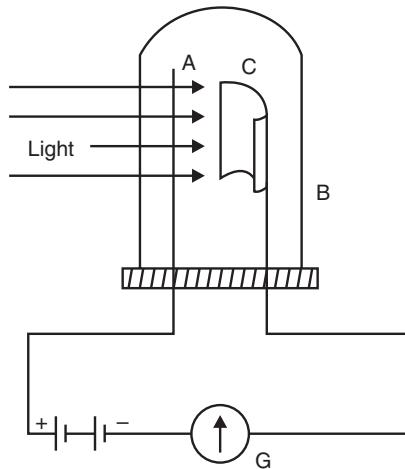


Fig. 6.11

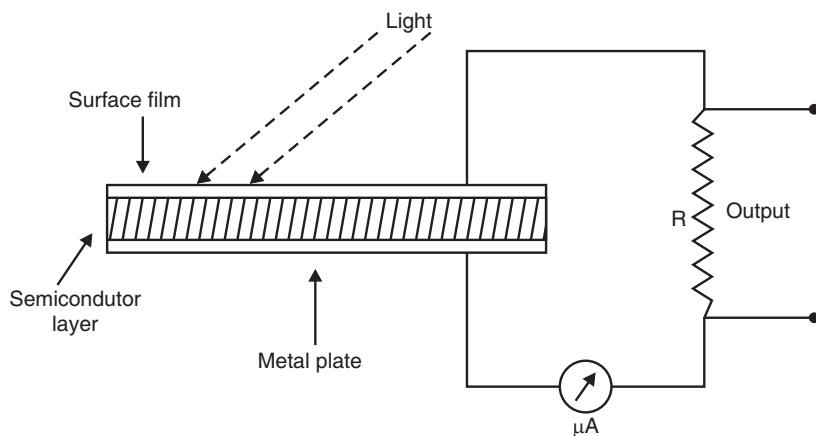


Fig. 6.12

**Working.** Light falling on the surface film (of gold or silver) penetrates into it and ejects photoelectrons from the semiconductor layer. These electrons travel in a direction opposite to the direction of the incident light. The conventional current, therefore, flows in the direction of the incident light. For small values of the resistance of the galvanometer, the current is directly

proportional to the intensity of light. No external battery is required to operate a photovoltaic cell as the cell itself generates an e.m.f.

### (iii) Photoconductive Cell.

**Principle:** The photoconductive cell is based on the principle that the electrical resistance of a semiconductor material decreases with the increase of intensity of radiation incident upon it and conductivity is increased.

The photoconductive materials used are cadmium sulfide ( $\text{CdS}$ ) and cadmium selenide ( $\text{CdSe}$ ).

**Construction:** Fig. 6.13 shows the construction and graphical symbol of a photoconductive cell. It is a two-terminal semiconductor device. The terminal resistance varies (linearly) with the intensity of the incident light. A thin layer of the photoconductive material connected between terminals is simply exposed to the incident light energy.

**Working:** As the illumination on the device increases in intensity, the energy state of a larger number of electrons in the structure will also increase because of the increased availability of the photon packages of energy. The result is an increasing number of relatively "free" electrons in the structure and a decrease in the terminal resistance.

Fig. 6.14 shows the *terminal characteristics* of a photoconductive cell.

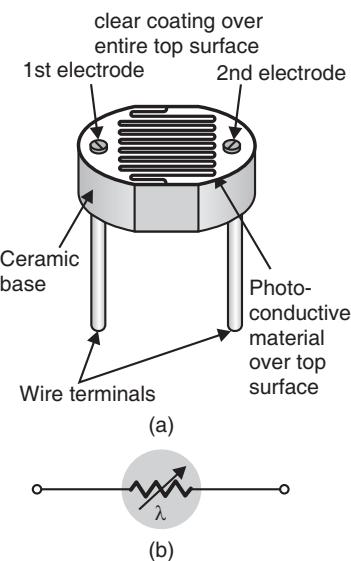


Fig. 6.13. Photoconductive cell:  
(a) construction; (b) symbol.

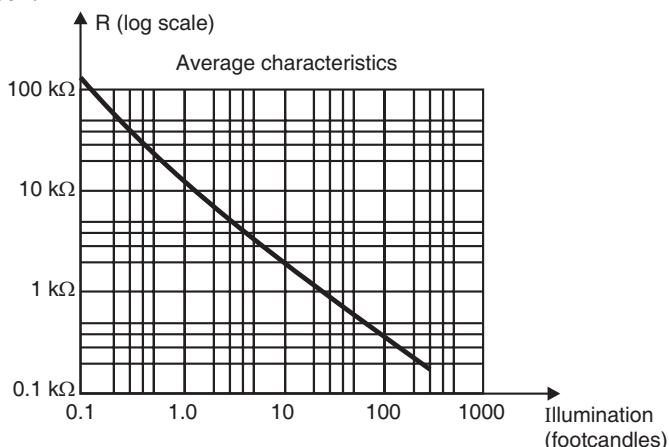


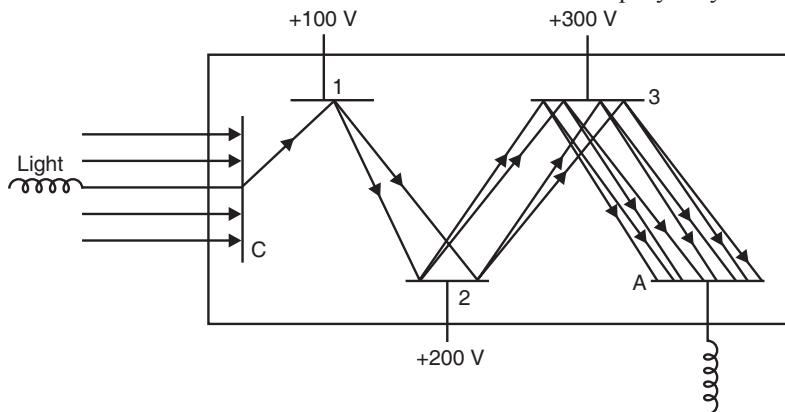
Fig. 6.14

Note the linearity (when plotted using a log-log scale) of the resulting curve and the large change in resistance ( $100 \text{ k}\Omega \rightarrow 100 \Omega$ ) for the indicated change in illumination.

**Applications of Photoelectric cells.** (i) *Exposure meters in photography.* An exposure meter is a device to calculate the correct time of exposure. The photoelectric cell in the instrument produces a current proportional to the light falling on it. The current operates a galvanometer, the scale of which is calibrated to read the time of exposure.

(ii) **Photo-multiplier.** It is based on the principle of *secondary emission*. When light strikes the surface of photosensitive metal plate  $C$ , it causes the ejection of photoelectrons from it (Fig. 6.15). These electrons are then attracted to a metal surface called a dynode, by setting a P.D. between the cathode  $C$  and the dynode 1. High energy electrons striking a metal surface can cause

the ejection of one or more secondary electrons from the surface. Suppose that a photoelectron striking dynode 1 produces  $x$  electrons by secondary emission. These electrons are then directed towards dynode 2 by making its potential higher than that of dynode 1. Suppose  $x$  electrons are again ejected by secondary emission for each incident electron. Then, for each electron emitted by the photosensitive plate, there are now  $x^2$  electrons and so on. If there are several dynodes, each at a potential higher than the preceding one, an avalanche of electrons reaches the collector plate  $A$ . A strong current then flows in the outer circuit. This device is used to amplify very weak light signals.



**Fig. 6.15**

(iii) Photoelectric cells are used to compare the illuminating powers of two light sources. They are also used in the measurement of the intensity of illumination of a light source.

(iv) **Sound reproduction in films.** The film is provided with a sound track at one edge. Light passing through the sound track of the film falls on a photocell. Current is produced, which fluctuates correspondingly with the intensity of sound recorded in the sound track. The current impulses are converted to sound by speakers.

(v) **Automatic operation of street lights.** A photoelectric cell, located in a street light circuit, switches off the street light when sunlight is incident on the cell. When sunlight fades and it becomes dark, the photoelectric cell switches on the street lights.

## 6.7 PLANCK'S QUANTUM THEORY

**The distribution of energy in the spectrum of a black body.** If the radiation emitted by a black body at a fixed temperature is analysed by means of a suitable spectroscopic arrangement, it is found to spread up into a continuous spectrum. The total energy is not distributed uniformly over the entire range of the spectrum.

**Experimental arrangement.** The distribution of energy in various parts of the spectrum was experimentally studied by Lummer and Pringsheim. The radiation from the black body was rendered into a parallel beam by the concave mirror [Fig. 6.16]. It is then allowed to fall on a prism of fluorspar to resolve it into a spectrum. The spectrum is brought to focus by another concave mirror on to a linear bolometer. The bolometer is connected to a galvanometer. The deflections in the galvanometer corresponding to different  $\lambda$  are noted by rotating the prism table. Then curves are plotted for  $E_\lambda$  versus  $\lambda$ . The experiment is done with the black body at different temperatures. The curves obtained are shown in Fig. 6.17.

**Results.** (i) At any given temperature,  $E_\lambda$  first increases with  $\lambda$ , reaching a maximum value corresponding to a particular wavelength  $\lambda_m$  and then decreases for longer wavelengths.

- (ii) The value of  $E_\lambda$  for any  $\lambda$  increases as temperature increases.
- (iii) The wavelength corresponding to the maximum energy shifts to shorter wavelength side as the temperature increases. This confirms Wien's displacement law  $\lambda_m T = \text{constant}$ .

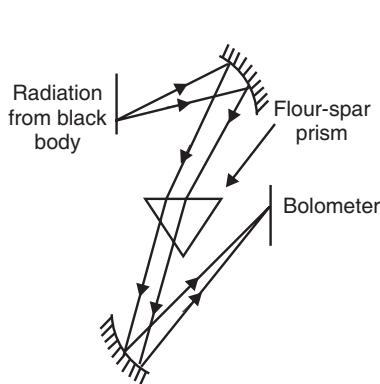


Fig. 6.16

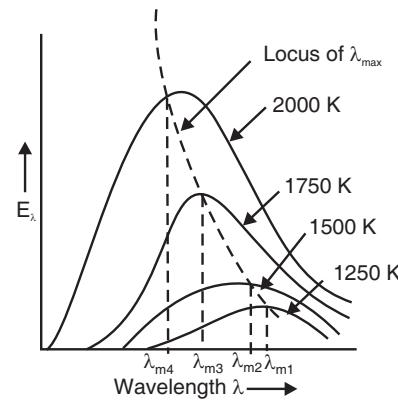


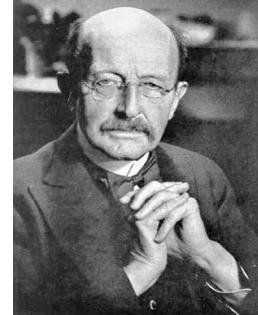
Fig. 6.17

(iv) Total energy emitted per unit area of the source per second at a given temperature is  $\int_0^\infty E_\lambda d\lambda$ . It will be represented by the total area between the curve for that temperature and the  $\lambda$ -axis. This area is found to be proportional to the fourth power of the absolute temperature. This verifies Stefan's law.

**Wien's Displacement Law.** *The wavelength of the most strongly emitted radiation in the continuous spectrum from a full radiator is inversely proportional to the absolute temperature of that body, i.e.,  $\lambda_m T = b$ .*

Here,  $b$  is Wien's constant  $2.898 \times 10^{-3}$  mK.

**Planck's hypothesis.** According to the classical theory of radiation, energy changes of radiators take place continuously. The classical theory failed to explain the experimentally observed distribution of energy in the spectrum of a black body. Planck succeeded in deriving a formula which agrees extremely well with experimental results. He discarded both the idea of radiation being a continuous stream as well as the law of equipartition of energy. He suggested the *quantum theory of radiation*. His assumptions are:



(1) A black-body radiation chamber is filled up not only with radiation, but also with *simple harmonic oscillators* or resonators of molecular dimensions. They can vibrate with all possible frequencies. **Max Planck (1858–1947)**

(2) The oscillators or resonators cannot radiate or absorb energy continuously. But an oscillator of frequency  $v$  can only radiate or absorb energy in units or quanta of magnitude  $hv$ .  $h$  is a universal constant called Planck's constant. The emission of radiation corresponds to a decrease and absorption to an increase in the energy and amplitude of an oscillator.

**Derivation of Planck's law of radiation.** Let  $N$  be the total number of Planck's resonators and  $E$  their total energy. Then, average energy per oscillator  $= \bar{\epsilon} = E / N$ .

$$N = N_0 + N_0 e^{-\epsilon/kT} + N_0 e^{-2\epsilon/kT} + \dots + N_0 e^{-r\epsilon/kT} + \dots$$

Here

$N_0$  = number of resonators having 0 energy.

$N_0 e^{-\epsilon/kT}$  = number of resonators having energy  $\epsilon$ ,

$N_0 e^{-2\epsilon/kT}$  = number of resonators having energy  $2\epsilon$ ,

$N_0 e^{-r\epsilon/kT}$  = number of resonators having energy  $r\epsilon$  and so on.

Putting

$$\begin{aligned} \varepsilon/kT &= x, \\ N &= N_0 + N_0 e^{-x} + N_0 e^{-2x} + \dots \dots \dots N_0 e^{-rx} \\ \therefore N &= \frac{N_0}{1-e^{-x}} \end{aligned} \quad \dots(1)$$

The total energy of Planck's resonators is

$$\begin{aligned} E &= 0 \times N_0 + \varepsilon \times N_0 e^{-x} + 2\varepsilon \times N_0 e^{-2x} + \dots + r\varepsilon \times N_0 e^{-rx} + \dots \\ Ee^{-x} &= \varepsilon N_0 e^{-2x} + 2\varepsilon N_0 e^{-3x} + \dots + r\varepsilon N_0 e^{-(r+1)x} \\ \text{Subtracting, } E(1-e^{-x}) &= \varepsilon N_0 e^{-x} + \varepsilon N_0 e^{-2x} + \varepsilon N_0 e^{-3x} + \dots \\ &= \frac{\varepsilon N_0 e^{-x}}{1-e^{-x}} \\ \therefore E &= \frac{\varepsilon N_0 e^{-x}}{(1-e^{-x})^2} \end{aligned} \quad \dots(2)$$

$$\left. \begin{array}{l} \text{Average energy} \\ \text{of a resonator} \end{array} \right\} = \bar{\varepsilon} = \frac{E}{N} = \frac{\varepsilon e^{-x}}{1-e^{-x}} = \frac{\varepsilon}{e^x - 1}$$

According to Planck's hypothesis,  $\varepsilon = h\nu$ . Further  $\nu = c/\lambda$ .

Hence,

$$\begin{aligned} \varepsilon &= \frac{hc}{\lambda} \quad \text{and} \quad x = \frac{\varepsilon}{kT} = \frac{hc}{\lambda kT} \\ \therefore \bar{\varepsilon} &= \frac{hc/\lambda}{(e^{hc/\lambda kT} - 1)} \end{aligned} \quad \dots(3)$$

Number of oscillators per unit volume in the wavelength range  $\lambda$  and  $\lambda + d\lambda = 8\pi \lambda^{-4} d\lambda$ .

... (4)

Hence, energy density of radiation between wavelengths  $\lambda$  and  $\lambda + d\lambda$  = (average energy of a Planck's oscillator)  $\times$  (number of oscillators per unit volume).

$$\begin{aligned} E_\lambda d\lambda &= \frac{hc/\lambda}{(e^{hc/\lambda kT} - 1)} \times 8\pi \lambda^{-4} d\lambda \\ E_\lambda d\lambda &= \frac{8\pi hc \lambda^{-5}}{(e^{hc/\lambda kT} - 1)} d\lambda \end{aligned} \quad \dots(5)$$

$$\text{or } E_\nu dv = \frac{8\pi h\nu^3}{c^3 (e^{h\nu/kT} - 1)} dv \quad \dots(6)$$

Here  $E_\nu dv$  is the energy density belonging to the range  $dv$ .  
Eq. (5) represents Planck's radiation law in terms of wavelength.  
Planck's formula fits the experimental curve very closely [Fig. 6.18].

- Planck's formula reduces to Wien's formula for small wavelengths. When  $\lambda$  is small,  $e^{hc/\lambda kT}$  is large when compared to 1. Hence Eq. (5) reduces to

$$E_\lambda d\lambda = 8\pi h c \lambda^{-5} e^{-hc/\lambda kT} d\lambda \quad \dots(7)$$

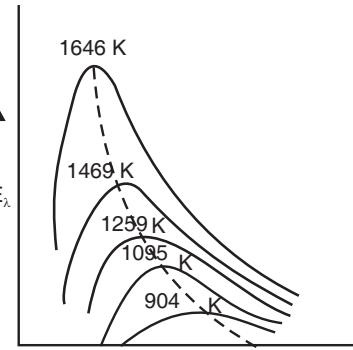


Fig. 6.18

This is *Wien's law*.

- Planck's formula reduces to Rayleigh Jean's formula for longer wavelengths.

When  $\lambda$  is large,  $e^{hc/\lambda kT} \approx 1 + (hc/\lambda kT)$ .

Hence Planck's law reduces to

$$E_\lambda d\lambda = 8\pi \frac{hc\lambda^{-5}}{(hc/\lambda kT)} d\lambda = 8\pi kT\lambda^{-4} d\lambda \quad \dots(8)$$

This is *Rayleigh-Jeans formula*.

### EXERCISE

1. The velocity of photoelectron depends upon the ..... of the incident photon only.  
(B.U., April 2012) [Ans. Frequency]
2. ..... is known as the Einstein's photoelectric equation.  
(B.U., April 2012) [Ans.  $h\nu = \phi + 1/2 mv^2$ ]
3. At threshold frequency, the kinetic energy of emitted photoelectron is .....  
(B.U., April 2014) [Ans. zero]
4. The photoelectric work function  $\phi = \dots$   
(B.U., April 2013) [Ans.  $\phi = h\nu_0$ ]
5. Millikan's experiment is based on .....  
(a) developing potential (b) high potential (c) stopping potential (d) negative potential.  
(B.U., April 2013) [Ans. (c)]
6. In photoconductive cell, as the intensity of radiation increases, ..... of semiconductor material decreases.  
(a) elasticity (b) hardness (c) resistance (d) none of these  
(B.U., April 2014) [Ans. (c)]
7. Explain the Richardson and Compton experiment to study the photoelectric phenomena.  
(B.U., April 2012)
8. State and explain laws of photoelectric emission.  
(B.U., April 2013)
9. Derive Einstein's photoelectric equation.  
(B.U., April 2011)
10. Calculate the work function of sodium, in electron-volts, given that the threshold wavelength is 6800 Å, and  $h = 6.625 \times 10^{-34}$  Js.  
(B.U., April 2012)  

[Sol.  $\phi = h\nu_0 = hc/\lambda_0$   
 Here,  $h = 6.625 \times 10^{-34}$  Js;  $c = 3 \times 10^8$  ms<sup>-1</sup> and  $\lambda_0 = 6800 \times 10^{-10}$  m.  
 $\therefore \phi = \frac{(6.625 \times 10^{-34})(3 \times 10^8)}{(6800 \times 10^{-10})} \text{ J} = \frac{(6.625 \times 10^{-34})(3 \times 10^8)}{(6800 \times 10^{-10})(1.6 \times 10^{-19})} \text{ eV}$   
 $= 1.827 \text{ eV.}]$
11. The photoelectric threshold for a metal is 3000 Å. Find the kinetic energy of an electron ejected from it by radiation of wavelength 1200 Å.  

[Sol. K.E. of the electron =  $\frac{1}{2}mv^2 = h(v - v_0)$   
 $= \frac{hc(\lambda_0 - \lambda)}{\lambda \lambda_0} = \frac{(6.62 \times 10^{-34})(3 \times 10^8)(1800 \times 10^{-10})}{(3000 \times 10^{-10})(1200 \times 10^{-10})} = 9.93 \times 10^{-19} \text{ J} = 6.2 \text{ eV}]$
12. Discuss the relation between photoelectric current and retarding potential in photoelectric effect.  
(B.U., April 2011)
13. With necessary theory, discuss the Millikan's experimental verification of Einstein's photoelectric equation.  
(B.U., April 2012)
14. Explain the action of a photoelectric cell.  
(B.U., April 2013)
15. Describe construction and working of photovoltaic cell.  
(B.U., April 2013)
16. What is photoconductive cell? Explain its working and important applications.  
(B.U., April 2011)
17. Discuss the distribution of energy in the spectrum of a black body and its results.  
(B.U., April 2012)
18. Derive an expression for Planck's law of radiation.  
(B.U., April 2013)

## PART - III

# QUANTUM MECHANICS

- 7. Wave Properties of Particles
- 8. Schrödinger Equation and its Applications
- 9. Quantum Theory of the Hydrogen Atom
- 10. Operator Formalism of Quantum Mechanics
- 11. Identical Particles and Spin
- 12. Scattering Theory
- 13. Perturbation Theory

## WAVE PROPERTIES OF PARTICLES

**After reading this chapter, you should be able to**

- ◆ Identify and understand the kinds of experimental results which are incompatible with classical physics and which required the development of a quantum theory of matter and light
- ◆ Outline the de-Broglie's theory of matter waves and discuss its experimental verification
- ◆ Distinguish between phase velocity and group velocity in wave motion
- ◆ Utilize both versions of Heisenberg's uncertainty principle in calculations
- ◆ Estimate the ground state energy and the radius of the hydrogen atom using uncertainty principle
- ◆ Explain why wave nature of matter is not observed in our daily experience and discuss the compatibility of the quantum picture with our experience in the macroscopic domain.

## 7.1 INADEQUACY OF CLASSICAL MECHANICS

Classical mechanics failed to explain the following phenomena:

- (i) It does not hold in the region of atomic dimensions, i.e., it can not explain the non-relativistic motion of atoms, electrons, protons etc.
- It could not explain the **stability of atoms**.
- It could not explain the origin of discrete spectra of atoms since, according to *classical mechanics, the energy changes are always continuous*.
- (ii) It could not explain observed spectrum of black body radiation.
- (iii) Classical mechanics could not explain observed phenomena like photoelectric effect, Compton effect etc.
- (iv) It could not explain the observed variation of specific heat capacity of solids.
- The inadequacy of classical mechanics led to the development of *Quantum Mechanics*.

### (i) The Hydrogen Atom and the Bohr Model

As the first example of the failure of classical physics to account for observed phenomena, we consider the case of the hydrogen atom. Rutherford model failed to explain two main observational features of the hydrogen atom:

(a) its stability and (b) the spectrum of its radiation. Let us consider these one at a time.

(a) An electron in a curved orbit is accelerated and hence must radiate. As it radiates its energy away, the radius of its orbit must decrease until eventually it collapses into the nucleus. Thus, the atom cannot be stable. But most of the atoms are stable.

(b) The second discrepancy involves the observed radiation spectrum. The frequency of the radiated energy should be the same as the orbiting frequency. As the electron orbit collapses, its orbiting frequency increases continuously. We might thus expect the spectrum of radiation emitted by excited hydrogen atoms to be continuous. In contrast, the experimentally observed spectrum consists of families of discrete lines.

Bohr provided an explanation for both the spectral discreteness and the observed stability. He proposed that in solving for the orbital motion of the electron in its hydrogenic orbit one should impose an added condition:

*The angular momentum of the electron must be equal to some integer multiple of  $\hbar$ .*

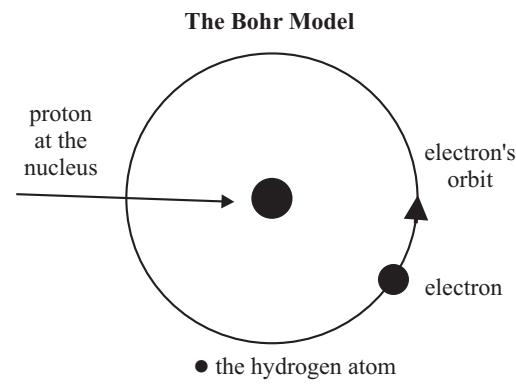
$$l = n\hbar \quad (n = 1, 2, 3\dots)$$

The quantisation implies that the laws of classical mechanics and of classical electro-magnetism are inapplicable at the atomic level.

### (ii) Black Body Radiation

The observed variation of the spectral intensity  $I(v)$  (power per unit area per unit frequency) of blackbody radiation as a function of frequency  $v$  is shown in Fig. 7.1. From the curves we note that

- (i) The intensity reaches a maximum at some frequency  $v_m$ .
- (ii) The frequency  $v_m$ , as well as the height of the peak, increase with temperature.



**Bohr's model of the hydrogen atom.**

The application of statistical thermodynamics and the ordinary laws of mechanics and electromagnetic theory led to the Rayleigh-Jeans formula

$$I(v) = \frac{8\pi v^2 kT}{c^3}.$$

This law, except for very low frequencies, is in total disagreement with experimental results. The law predicts an infinite amount of radiated intensity. Actually, the total radiated intensity is finite.

Max Planck resolved this controversy by postulating that the exchange of energy between atoms and radiation involves *discrete* amounts of energy. At a given frequency  $v$ , the smallest amount of energy that can be exchanged is equal to

$$E = hv.$$

*Only multiples of  $hv$  are involved in the interaction.*

Applying this postulate to the problem of black body radiation theory, Planck obtained

$$I(v) = \frac{8\pi v^2}{c^3} \frac{hv}{e^{hv/kT} - 1}.$$

Planck's formula agrees with experimental curves. *The notion that field energy is quantized rather than a continuous quantity was a new and profound addition to physics.*

### (iii) The Photoelectric Effect

A direct confirmation of the energy quantization of electromagnetic fields was provided by the phenomenon of photoelectric emission. When light of frequency  $v$  is incident on a metal, electrons are emitted from the metallic surface only when  $hv > \phi$ . When  $hv > \phi$ , the electrons are emitted with a kinetic energy  $T$ , where

$$hv = \phi + T.$$

The energy of the photo-electron is independent of the intensity of light. The intensity of light determines merely the number (per second) of the emitted electrons. The explanation was provided by Einstein and invoked the electromagnetic field particles, photons, each carrying an energy  $hv$ . Electrons in the material are held back from the vacuum by an energy barrier of height  $\phi$ . The impinging photon can transmit its energy  $hv$  to an electron near the surface. If  $hv < \phi$ , this energy is insufficient to surmount the barrier and no electrons are ejected. If  $hv > \phi$ , the excess  $T = hv - \phi$  is the kinetic energy of the electron leaving the surface.

### (iv) The Compton Effect

The scattering of monochromatic X-rays from targets composed of light elements was studied by Compton. He found that the scattered radiation consists of two lines, one of the same wavelength as the incident radiation and one of slightly longer wavelength. The change in wavelength depended on the scattering angle. Compton was able to explain the measured angular dependence of wavelength shift perfectly when he assumed that individual photons of energy  $E = hv$  and momentum  $p = \frac{hv}{c}$

collided with individual electrons in such a way that momentum and energy were conserved. Compton's formula

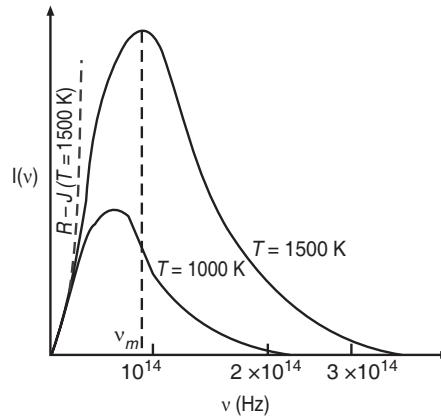


Fig. 7.1

$$\lambda' - \lambda = \Delta\lambda = \frac{h}{m_0 c} (1 - \cos \theta)$$

gives the wavelength of the scattered photon in terms of incident wavelength and the angle at which the scattered photon is detected. This equation was verified by Compton experimentally.

The quantity  $\frac{h}{m_0 c}$  in Compton's formula has dimensions of length and is called the Compton wavelength of the electron. Its magnitude is  $2.43 \times 10^{-12}$  m. This constant, which is characteristic of the X-ray scattering with wavelength shift, could not be reproduced by applying Maxwell's theory of electromagnetism to the scattering process.

#### (v) The Heat Capacity of Solids

Another experimental observation that could not be explained by ordinary physics was that of the heat capacity of solids. Classical equipartition of energy led to the prediction that the specific heat per kilo mole of solid elements should be  $3R$ . Dulong and Petit verified this result for many elements at high temperatures. However, the prediction failed completely in the low-temperature domain.

This disagreement between theory and experiment was resolved by Einstein and Debye who applied the Planck quantization condition

$$E = h\nu$$

to the mechanical oscillations of the lattice. The most profound effect of applying Planck's postulate to an oscillator is that its average thermal equilibrium excitation energy is no longer equal to  $kT$  (the classical value) but to

$$E = \frac{h\nu}{e^{h\nu/kT} - 1} \quad \dots(1)$$

Note that  $E \rightarrow kT$  for  $kT \gg h\nu$ . The application of Eq. (1) to the heat capacity problem resulted in excellent agreement with experiment.

## 7.2 MATTER WAVES

According to de Broglie a *moving particle, whatever its nature, has wave properties associated with it*. He proposed that the wavelength  $\lambda$  associated with any moving particle of momentum  $p$  (mass  $m$  and velocity  $v$ ) is given by

$$\lambda = \frac{h}{p} = \frac{h}{mv}, \quad \dots(1)$$

Here,  $h$  is Planck's constant. Such waves associated with the matter particles are called *matter waves* or *de Broglie waves*.

Bohr's Theory of the hydrogen atom led de Broglie to the conception of matter waves. According to Bohr's theory, the stable states of electrons in the atom are governed by "integer rules". The only phenomena involving integers in physics are those of interference and modes of vibration of stretched strings, both of which imply wave motion. Hence de Broglie thought that the electrons may also be characterised by a periodicity. So he proposed that *matter, like radiation, has dual nature*. Eq. (1) was verified by experiments involving the diffraction of electrons by crystals.



**Louis de Broglie**  
(1892-1987)

### The de Broglie Wavelength

A photon of light of frequency  $\nu$  has the momentum

$$p = h\nu/c.$$

But  $\nu = c/\lambda$ . Therefore, the momentum of the photon can be expressed in terms of wavelength  $\lambda$  as

$$p = h/\lambda.$$

The wavelength of a photon is, therefore, specified by its momentum according to the relation

$$\lambda = h/p. \quad \dots(1)$$

de Broglie suggested that Eq. (1) is a completely general one that applies to material particles as well as to photons. The momentum of a particle of mass  $m$  and velocity  $v$  is  $p = mv$ , and its de Broglie wavelength is accordingly

$$\lambda = h/mv.$$

#### 7.2.1. Expression for de Broglie Wavelength

De Broglie derived an expression for the wavelength of matter waves on the analogy of radiation.

Based on Planck's theory of radiation, the energy of a photon (quantum) is

$$E = h\nu = \frac{hc}{\lambda} \quad \dots(1)$$

Here,  $c$  is the vacuum velocity of light and  $\lambda$  its wavelength.

According to Einstein mass-energy relation,

$$E = mc^2 \quad \dots(2)$$

From Eqs. (1) and (2),

$$\begin{aligned} mc^2 &= \frac{hc}{\lambda} \\ \text{or} \quad \lambda &= \frac{h}{mc} = \frac{h}{p} = \frac{h}{\text{momentum}} \end{aligned} \quad \dots(3)$$

Here,  $mc = p$  is the momentum associated with photon.

De Broglie suggested that this equation for wavelength is a perfectly general one, applying to material particles as well as to photons. In the case of particles,

$$\text{momentum} = mv.$$

Hence the de Broglie wavelength of a particle is

$$\lambda = \frac{h}{mv} = \frac{h}{p} \quad \dots(4)$$

#### 7.2.2. Other Expressions for de-Broglie Wavelength

(i) **For a Free particle:** If  $E_k$  is the kinetic energy of the material particle, then

$$p = \sqrt{(2mE_k)}$$

Therefore, de Broglie wavelength of particle of K.E. =  $E_k$  is given by

$$\lambda = \frac{h}{\sqrt{2mE_k}} \quad \dots(1)$$

(ii) **For a charged particle accelerated through a potential difference  $V$ :** If a charged particle carrying charge  $q$  is accelerated through a potential difference  $V$  volts, then kinetic energy  $E_k = qV$ .

*∴ The de Broglie wavelength for charged particle of charge  $q$  and accelerated through a potential difference of  $V$  volts is*

$$\lambda = \frac{h}{\sqrt{2mqV}} \quad \dots(2)$$

(iii) **For thermal neutrons:** Kinetic energy of thermal neutrons is  $E_k = kT$ .

Here,

$k$  = Boltzmann constant,

$T$  = temperature at which neutrons are enclosed in the chamber.

Therefore, the de-Broglie wavelength of a material particle at temperature  $T$  is

$$\lambda = \frac{h}{\sqrt{(2mE_k)}} = \frac{h}{\sqrt{2mkT}} \quad \dots(3)$$

**EXAMPLE 1.** Find the de Broglie wavelength associated with

- (i) A 46 gm golf ball with velocity 36 m/s.
- (ii) an electron with a velocity  $10^7$  m/s.

Which of these two show wave character and why?

(Garhwal 1994)

**SOL.** (i) Since  $v \ll c$ , we can take  $m = m_0 \rightarrow$  the rest mass. Hence

$$\begin{aligned} \lambda &= \frac{h}{mv} = \frac{6.63 \times 10^{-34} \text{ J.s}}{(0.046 \text{ kg})(36 \text{ m/s})} \\ &= 4.0 \times 10^{-34} \text{ m.} \end{aligned}$$

Thus the wavelength associated with golf ball is much smaller as compared to its dimensions. Hence **no wave aspects** can be expected in its behaviour.

(ii) Again  $v \ll c$ , so  $m = m_0 = 9.1 \times 10^{-31} \text{ kg.}$

$$\therefore \lambda = \frac{h}{mv} = \frac{6.63 \times 10^{-34}}{(9.1 \times 10^{-31}) \times 10^7} = 7.3 \times 10^{-11} \text{ m}$$

This wavelength is comparable with the atomic dimensions. Hence a **moving electron exhibits a wave character**.

**EXAMPLE 2.** Show that the de Broglie wavelength associated with an electron of energy  $V$  electron-volts is approximately  $(1.227/\sqrt{V}) \text{ nm}$ .

**SOL.** The de Broglie wavelength  $\lambda$  associated with an electron of mass  $m$  and energy  $E$  is given by

$$\lambda = \frac{h}{\sqrt{(2mE)}}$$

Here, kinetic energy  $E_k = V \text{ eV} = 1.6 \times 10^{-19} \text{ VJ}$

$$\begin{aligned} \lambda &= \frac{6.62 \times 10^{-34}}{\sqrt{(2 \times 9.1 \times 10^{-31} \times 1.6 \times 10^{-19} V)}} \\ &= \frac{1.227 \times 10^{-9} \text{ m}}{\sqrt{V}} = \frac{1.227}{\sqrt{V}} \text{ nm} \end{aligned}$$

**EXAMPLE 3.** Find the kinetic energy of a proton whose de Broglie wavelength is 1 fm.

**SOL.**  $pc = (mv)c = hc/\lambda$

$$= \frac{(4.136 \times 10^{-15} \text{ eV.s}) (3 \times 10^8 \text{ ms}^{-1})}{1 \times 10^{-15} \text{ m}} = 1.241 \text{ GeV.}$$

Rest energy of proton =  $E_0 = 0.938 \text{ GeV}$ .

$pc > E_0$ . Hence a relativistic calculation is needed.

The total energy of the proton is

$$E = \sqrt{E_0^2 + p^2 c^2} = \sqrt{(0.938 \text{ GeV})^2 + (1.241 \text{ GeV})^2} = 1.556 \text{ GeV.}$$

The kinetic energy of the proton is

$$KE = E - E_0 = (1.556 - 0.938) \text{ GeV} = \mathbf{0.618 \text{ GeV}}$$

**EXAMPLE 4.** Show that the de Broglie wavelength for a material particle of rest mass  $m_0$  and charge  $q$ , accelerated from rest through a potential difference of  $V$  volts relativistically is given by

$$\lambda = \frac{h}{\sqrt{2m_0 qV \left(1 + \frac{qV}{2m_0 c^2}\right)}}$$

**SOL.** We use the relativistic formula to find momentum.

$$\text{Kinetic energy } E_k = qV.$$

$$E^2 = p^2 c^2 + m_0^2 c^4, E = E_k + m_0 c^2 = qV + m_0 c^2$$

So,

$$p^2 c^2 = E^2 - m_0^2 c^4 = (qV + m_0 c^2)^2 - m_0^2 c^4 = q^2 V^2 + 2m_0 c^2 qV$$

$$p^2 = 2m_0 qV \left(1 + \frac{qV}{2m_0 c^2}\right) \text{ or } p = \sqrt{2m_0 qV \left(1 + \frac{qV}{2m_0 c^2}\right)}$$

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2m_0 qV \left(1 + \frac{qV}{2m_0 c^2}\right)}}$$

**Special Case.** If the charged particle is an electron, then  $q = e = 1.6 \times 10^{-19} \text{ C}$ ,  $m_0 = 9.1 \times 10^{-31} \text{ kg}$ .

$$\begin{aligned} \lambda &= \frac{6.62 \times 10^{-34}}{\sqrt{(2 \times 9.1 \times 10^{-31} \times 1.6 \times 10^{-19} V) \left(1 + \frac{1.6 \times 10^{-19} V}{2 \times 9.1 \times 10^{-31} \times (3 \times 10^8)^2}\right)}} \\ &= \frac{1.227}{\sqrt{V}} \times \frac{1}{\sqrt{1 + 9.768 \times 10^{-7} V}} \text{ nm} \end{aligned}$$

### 7.2.3. Phase velocity (or Wave Velocity) of de Broglie Waves

The phase velocity  $v_p$  of a monochromatic wave is the velocity with which a definite phase of the wave, such as its crest or trough, is propagated in a medium.

The equation of motion of a plane wave of frequency  $\nu$  and wavelength  $\lambda$  moving in the  $x$ -direction is

$$y = A \cos(\omega t - kx).$$

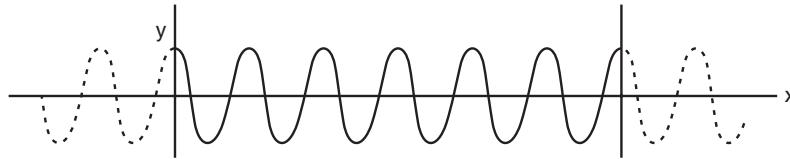


Fig. 7.2

Here,  $A$  is the amplitude of the wave (Fig. 7.2).

$\omega (= 2\pi\nu)$  is the angular frequency and  $k (= 2\pi/\lambda)$  is the propagation constant.

This wave moves with a phase velocity

$$v_p = v\lambda = \omega/k$$

- A particle of mass  $m$  moving with velocity  $v$  has a wave associated with it whose wavelength is given by

$$\lambda = \frac{h}{mv}.$$

Let  $E$  be the total energy of the particle. Let  $\nu$  be the frequency of the associated wave. We equate the quantum expression  $E = hv$  with the relativistic formula for total energy  $E = mc^2$ . So we get

$$hv = mc^2 \text{ or } \nu = mc^2/h.$$

Let  $v_p$  be the de Broglie wave velocity. Then,

$$v_p = v\lambda = \left(\frac{mc^2}{h}\right) \left(\frac{h}{mv}\right) = \frac{c^2}{v}.$$

But the particle velocity  $v$  is always less than  $c$  (the velocity of light). Therefore, the de Broglie wave velocity  $v_p$  must be greater than  $c$ .

#### 7.2.4. Group Velocity

It is possible to form a concentrated *wave packet* by taking a sum of a large number of plane waves with slightly different wavelengths and frequencies. When such a group travels in a dispersive medium, the phase velocities of its different components are different. The observed velocity is, however, the velocity with which the maximum amplitude of the group advances in the medium. This is called the “group velocity”. Thus, *the group velocity is the velocity with which the energy in the group is transmitted*. The individual waves travel “inside” the group with their phase velocities.

- *The velocity with which the centre of a wave group i.e., maximum amplitude, moves is called the group velocity of the wave group.*

#### 7.2.5. Expression for Group Velocity

Consider two waves that have the same amplitude  $A$  but differ by an amount  $\Delta\omega$  in angular frequency and an amount  $\Delta k$  in wave number. They can be represented by the equations

$$y_1 = A \cos(\omega t - kx)$$

$$y_2 = A \cos[(\omega + \Delta\omega)t - (k + \Delta k)x]$$

The superposition of the two waves will yield a single *wave packet* or *wave group*.

Let us find the velocity  $v_g$  with which the wave group travels.

The resultant displacement  $y$  at any time  $t$  and any position  $x$  is the sum of  $y_1$  and  $y_2$ .

$$\begin{aligned} y &= y_1 + y_2 \\ &= A \cos(\omega t - kx) + A \cos[(\omega + \Delta\omega)t - (k + \Delta k)x] \end{aligned}$$

$$= 2A \cos \frac{1}{2} [2\omega + \Delta\omega]t - (2k + \Delta k)x \cos \frac{1}{2} (\Delta\omega t - \Delta kx)$$

$\Delta\omega$  and  $\Delta k$  are small compared with  $\omega$  and  $k$  respectively. Therefore,

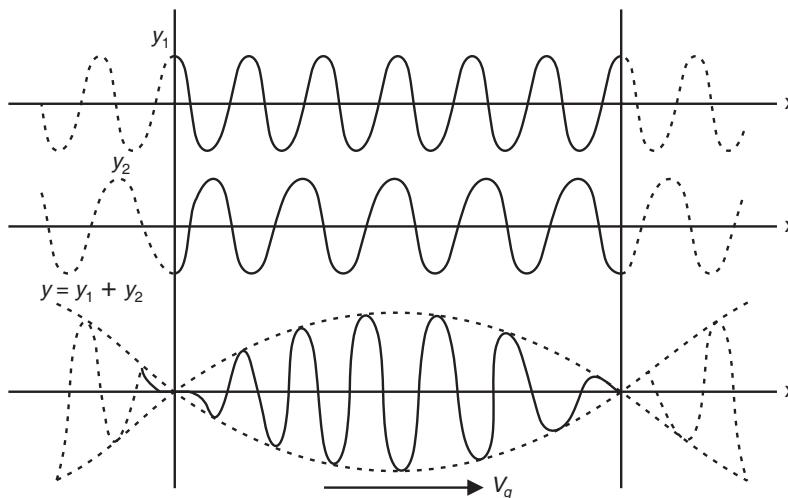


Fig. 7.3

$$2\omega + \Delta\omega \approx 2\omega$$

$$2k + \Delta k \approx 2k.$$

$$\therefore y = 2A \cos \left( \frac{\Delta\omega}{2} t - \frac{\Delta k}{2} x \right) \cos (\omega t - kx) \quad \dots(1)$$

This is the analytical expression for resultant wave (wave packet) due to superposition of the two waves. The second cosine function is the original wave. The coefficient of this cosine can be considered to be an amplitude that varies with  $x$  and  $t$ . This variation of amplitude is called the *modulation* of the wave.

Hence Eq. (1) represents a wave of angular frequency  $\omega$  and wave number  $k$  that has superimposed upon it a modulation of angular frequency  $\frac{1}{2} \Delta\omega$  (Fig. 7.3).

The velocity  $v_g$  of the wave groups is

$$v_g = \frac{\Delta\omega}{\Delta k} \quad \dots(2)$$

When  $\omega$  and  $k$  have continuous spreads, the group velocity is given by

$$v_g = \frac{d\omega}{dk} \quad \dots(3)$$

This is the expression for the group velocity.

#### Group Velocity of de Broglie Waves

A particle moving with a velocity  $v$  is supposed to consist of a group of waves, according to de Broglie hypothesis.

The group velocity is given by,

$$v_g = \frac{d\omega}{dk}$$

The angular frequency and wave number of the de Broglie waves associated with a particle of rest mass  $m_0$  moving with the velocity  $v$  are given by

$$\omega = 2\pi v = \frac{2\pi mc^2}{h} = \frac{2\pi m_0 c^2}{h\sqrt{1-v^2/c^2}} \quad \dots(1)$$

and

$$k = \frac{2\pi}{\lambda} = \frac{2\pi mv}{h} = \frac{2\pi m_0 v}{h\sqrt{1-v^2/c^2}} \quad \dots(2)$$

By differentiation, we obtain

$$\begin{aligned}\frac{d\omega}{dv} &= \frac{2\pi m_0 v}{h(1-v^2/c^2)^{3/2}} \\ \frac{dk}{dv} &= \frac{2\pi m_0}{h(1-v^2/c^2)^{3/2}}.\end{aligned}$$

The group velocity  $v_g$  of the de Broglie waves associated with particle is

$$v_g = \frac{d\omega}{dk} = \frac{d\omega/dv}{dk/dv} = v$$

Hence the de Broglie wave group associated with a moving particle travels with the same velocity as the particle.

#### Relation between Group Velocity ( $v_g$ ) and Wave Velocity or Phase Velocity ( $v_p$ )

We have the relations,

$$\text{Wave velocity, } v_p = \frac{\omega}{k} \quad \dots(1)$$

$$\text{Group velocity, } v_g = \frac{d\omega}{dk} \quad \dots(2)$$

The wave number is given by

$$\begin{aligned}k &= \frac{2\pi}{\lambda} \\ \therefore \frac{dk}{d\lambda} &= -\frac{2\pi}{\lambda^2} \quad \dots(3)\end{aligned}$$

$$\text{Also } \omega = 2\pi v = 2\pi \frac{v_p}{\lambda}$$

$$\therefore \frac{d\omega}{d\lambda} = 2\pi \left[ -\frac{v_p}{\lambda^2} + \frac{1}{\lambda} \frac{dv_p}{d\lambda} \right]$$

$$\text{or } \frac{d\omega}{d\lambda} = -\frac{2\pi}{\lambda^2} \left[ v_p - \lambda \frac{dv_p}{d\lambda} \right] \quad \dots(4)$$

Dividing Eq. (4) by Eq. (3), we get

$$\frac{d\omega}{d\lambda} \cdot \frac{d\lambda}{dk} = \frac{-\frac{2\pi}{\lambda^2} \left[ v_p - \lambda \frac{dv_p}{d\lambda} \right]}{-\frac{2\pi}{\lambda^2}}$$

or  $\frac{d\omega}{dk} = v_p - \lambda \frac{dv_p}{d\lambda}$   
 $\therefore v_g = v_p - \lambda \frac{dv_p}{d\lambda}$  ... (5)

Eq. (5) gives the relationship between group velocity ( $v_g$ ) and phase velocity or wave velocity ( $v_p$ ).

From this equation the following two cases arise.

(i) *For dispersive medium*

$v_p = f(\lambda)$ . Usually  $dv_p/d\lambda$  is positive (normal dispersion).

$$\therefore v_g < v_p.$$

This is the case with de Broglie waves.

(ii) *For non-dispersive medium*

$$v_p \neq f(\lambda). \frac{dv_p}{d\lambda} = 0. \therefore v_g = v_p$$

This result is true for electromagnetic waves in vacuum.

- If the phase velocity is the same for all wavelengths, as is true for light waves in empty space, the group and phase velocities are the same.

### 7.3 EXPERIMENTAL STUDY OF MATTER WAVES

#### Davisson and Germer's Experiment

**Experimental arrangement.** Fig. 7.4 shows the experimental arrangement. Electrons are produced by heating a filament (F) by a low tension battery (L.T.). The electrons are then accelerated through a potential difference  $V$  in the electron gun. This electron beam falls on a large single crystal of nickel. The electrons are scattered in all directions by the atoms in the crystal. The detector can be moved to any angle  $\phi$  relative to the incident beam. The energy of the electrons in the primary beam, the angle at which they reach the target, and the position of the detector could all be varied.

**Experimental procedure.** The accelerating potential  $V$  is given a low value. The beam of electrons falls normally on the surface of the crystal. The detector is moved to various positions and the intensity of the diffracted beam at each position is noted. The intensity of diffracted beam is plotted against the angle between the incident beam and the beam entering the detector. The method of plotting is such that the intensity at any angle is proportional to the distance of the curve at that angle from the point of scattering. The observations are repeated for different accelerating voltages and a number of curves are drawn. Fig. 7.5 shows typical polar graphs of electron intensity.

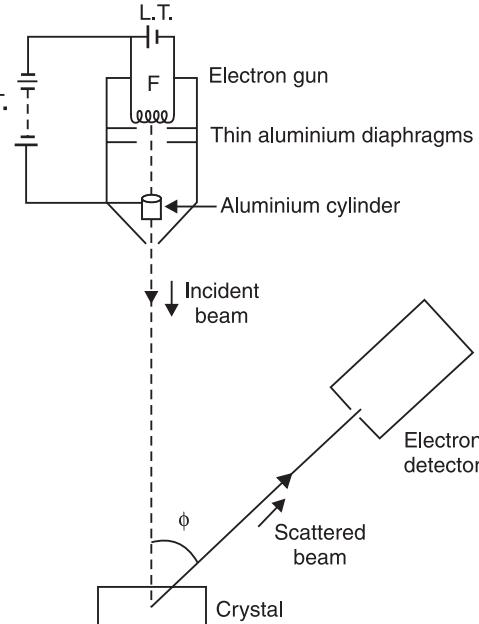


Fig. 7.4

- (i) The graph remains fairly smooth, till the accelerating voltage becomes  $44V$  when a '**bump**' appears on the curve.
- (ii) As the accelerating voltage is increased, the length of the '**bump**' increases.
- (iii) The '**bump**' becomes most prominent in the curve for  $54$  volt electrons at  $\phi = 50^\circ$ .
- (iv) As the accelerating voltage is further increased, the bump decreases in length and finally disappears at  $68V$ .

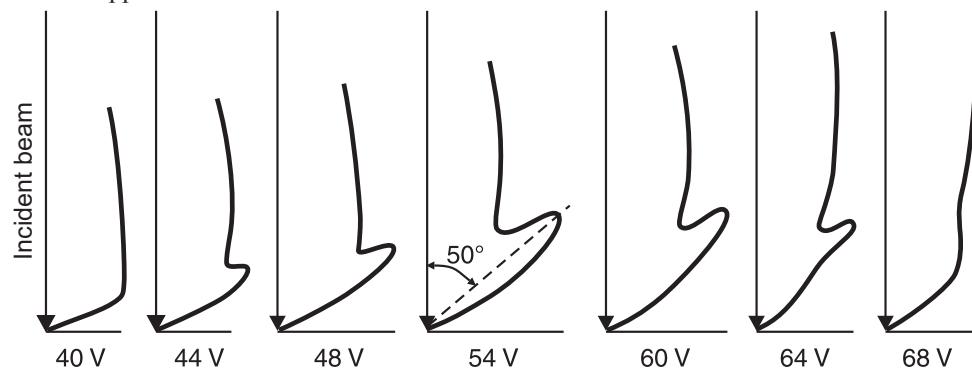


Fig. 7.5

### Calculation of Observed Wavelength

For nickel, the spacing of the atomic planes which can be measured by x-ray diffraction, is  $d = 0.091 \text{ nm}$ .

There is an intense reflection of the beam at the angle  $\phi = 50^\circ$ .

The angle of incidence relative to the family of Bragg planes shown in Fig. 7.6 is  $\theta = \frac{180 - 50}{2} = 65^\circ$ .

The Bragg equation for maxima in the diffraction pattern is

$$n \lambda = 2d \sin \theta \quad \dots(i)$$

For  $n = 1$ , the de Broglie wavelength  $\lambda$  of the diffracted electrons is

$$\lambda = 2d \sin \theta = (2)(0.091 \text{ nm}) (\sin 65^\circ) = 0.165 \text{ nm}$$

### Calculation of the expected wavelength of the electrons using de Broglie's formula.

The electron kinetic energy KE is  $54 \text{ eV}$ . The electron wavelength is therefore

$$\lambda = \frac{h}{mv} = \frac{h}{\sqrt{2mKE}} = \frac{6.63 \times 10^{-34} \text{ J.s}}{\sqrt{(2)(9.1 \times 10^{-31} \text{ kg})(54 \text{ eV})(1.6 \times 10^{-19} \text{ J/eV})}} = 0.166 \text{ nm}$$

The observed wavelength agrees well with the expected wavelength. The Davisson-Germer experiment thus directly verifies de Broglie's hypothesis of the wave nature of moving bodies.

### G.P. Thomson's Experiment

**Experimental arrangement.** The experimental arrangement is shown in Fig. 7.7. A beam of cathode rays is produced in a discharge tube  $AC$  by means of an induction coil. The electrons

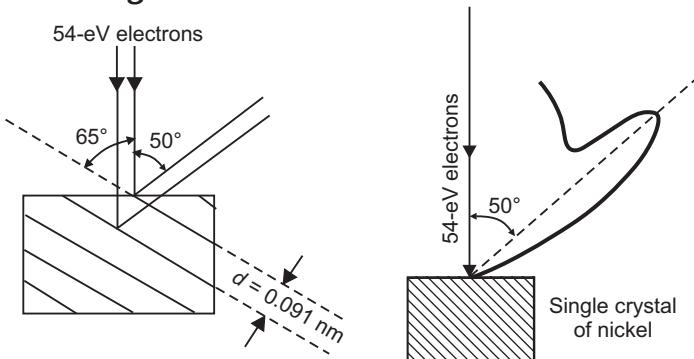


Fig. 7.6

passing through a fine hole  $A$ , are incident on a thin gold foil  $F$ . The thickness of the foil is about  $10^{-8}$  m. The emergent beam of electrons is received on a photographic plate  $P$ . Visual examination of the pattern is made possible by the fluorescent screen  $S$ . A very high vacuum is maintained in the camera part  $FP$  of the apparatus while air is allowed to leak into the discharge tube section through a needle valve. Thus a beam of required voltage can be continuously produced, the discharge tube being sufficiently soft.

**Experimental procedure.** A beam of electrons of known velocity is made to fall on the photographic plate, after traversing the thin gold foil. When the plate is developed a symmetrical pattern consisting of concentric rings about a central spot, is obtained [Fig. 7.8]. This pattern is similar to the pattern produced by X-rays in the powdered crystal method. When the cathode rays in the discharge tube are deflected by a magnetic field, the entire pattern on the screen  $S$  is found to shift correspondingly. Thus the pattern is confirmed as due to diffracted electrons and not due to secondary X-rays, generated by the electrons going through the foil. Further, on removing the film  $F$  the pattern disappears, showing that the presence of the film is essential. If the electrons behaved as corpuscles, the electrons passing through the foil should have been scattered through a wide angle. Clearly, this experiment demonstrates that the electron beam behaves as waves, since diffraction patterns can be produced only by waves.

**Verification of the de Broglie equation.** G.P. Thomson employed very high voltages of the order of 50000 volts to accelerate the electrons. With very high speed electrons, relativistic correction for the mass of the electron has to be applied. It can be shown that

$$\lambda = \frac{12.27}{\sqrt{V}} \frac{1}{\sqrt{1 + 9.768 \times 10^{-7} V}} \text{ Å.} \quad \dots(1)$$

The correction factor is very small, except for very large values of  $V$ .

**To calculate  $\lambda$  from the radii of the rings.** In the polycrystalline film, there will be some crystals set at the correct angle to give a Bragg reflection. If there are enough crystals distributed at random, the result of such reflections will be a series of rings, arising from the intersection of the cones of diffraction with the photographic plate. Let  $AB$  be the incident beam passing through the film at  $B$ .  $BP$  is the direction of the beam which has suffered a Bragg reflection in some crystal in the film at  $B$ . This reflected beam falls at the point  $P$  on the photographic plate at a distance  $R$  from the central point  $O$  (Fig. 7.9). Let the distance  $BO$ , from the film to the plate, be  $L$ .  $\angle PBO = 2\theta$  where  $\theta$  is the glancing angle given by the Bragg relation,

$$n\lambda = 2d \sin \theta.$$

$R/L = \tan 2\theta = 2\theta$ , since  $\theta$  is small.

$$\therefore \theta = R/2L$$

$$\text{But } 2d \sin \theta = 2d\theta = n\lambda \text{ or } \theta = n\lambda /2d$$

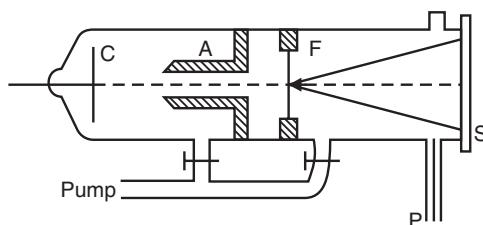


Fig. 7.7

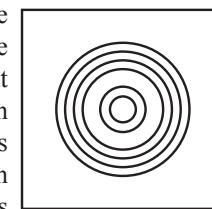


Fig. 7.8

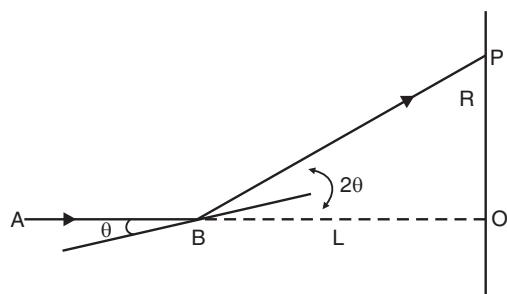


Fig. 7.9

$$\therefore \frac{n\lambda}{2d} = \frac{R}{2L}$$

or  $\lambda = \frac{Rd}{nL}$  ...(2)

From this the wavelength is calculated. This agrees with the calculated value given by Eqn. (1). This provides ultimate confirmation for the wave nature of the electron.

### Two Slit Interference Pattern with Electrons

**Principle.** The interference of waves associated with a beam of energetic electrons is demonstrated by a *double slit experiment*. The interference pattern so obtained resembles the pattern in case of visible light and proves the association of wave-packets with electrons.

**Experimental Arrangement.** Fig. 7.10 shows the experimental arrangement.

The electron gun (*G*) supplies a mono-energetic beam of electrons. The electrons emerging from the gun are allowed to pass through two slits  $S_1$  and  $S_2$ . The interference pattern is observed on a photographic film *P*.

The entire apparatus is enclosed in a high vacuum chamber. So electrons emerging from the gun travel to the screen without collisions in the path.

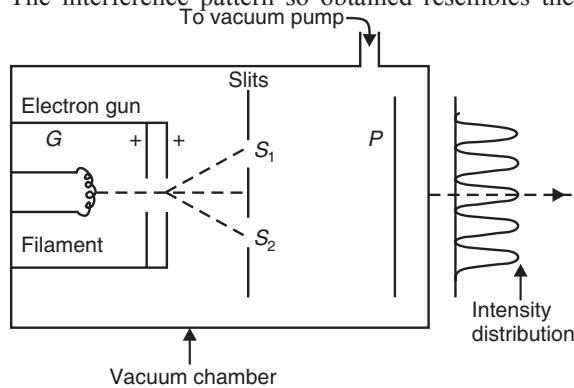


Fig. 7.10

**Result.** When photographic film is observed with an electron microscope, it is observed that interference fringes are obtained on the film. The distribution of intensity on the film is shown on the right. The formation of interference fringes clearly indicates the wave nature of electrons.

Knowing the separation between  $S_1$ ,  $S_2$  and the distance between the plane of  $S_1$ ,  $S_2$  and the photographic plate *P*, the value of wavelength associated with electrons is calculated. This value comes out to be same as given by de-Broglie relation.

#### 7.3.1. Standing de Broglie Waves of Electron in Bohr Orbits

The most recent development in the theory of atomic structure shows that the Bohr picture of the atoms with sharply defined electron orbits is not correct. According to the Wave mechanics developed by Heisenberg, Schrodinger and Dirac, the electrons move around the nucleus as *wave packets*. The *wave packets* are formed in a somewhat similar manner as standing waves are formed in sound.

This atom model has electrons in various orbits. The electrons behave as matter waves of wavelength

$$\lambda = \frac{h}{mv}.$$

The electron exists as a standing wave in each orbit. The energy level and ‘orbits’ of Bohr model are retained. Even the quantum conditions of Bohr follow from wave mechanics.

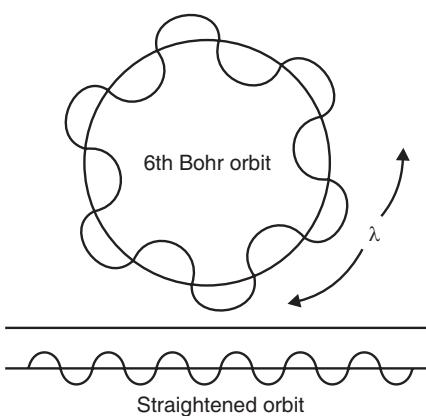


Fig. 7.11

For example, consider the electron in the hydrogen atom as a *standing wave* extending in a circle round the nucleus. In order that this wave may just occupy the circumference of a circle, the circle must contain an integral number of wavelengths (Fig. 7.11). *i.e.*,

$$2\pi r = n\lambda.$$

Here,  $r$  is the radius of the circle and  $n$ , an integer.

But

$$\lambda = \frac{h}{mv}.$$

$\therefore$

$$2\pi r = n \frac{h}{mv} \quad \text{or} \quad mv r = n \frac{h}{2\pi}.$$

$\therefore$

$$mv r = n\hbar.$$

But  $mv r$  is the angular momentum of the electron, regarded as a particle.

Hence the angular momentum is equal to an integral multiple of  $\hbar$  (Bohr's postulate).

#### 7.4 ELECTRON MICROSCOPE

**Principle.** A particle of mass  $m$  moving with a large velocity  $v$  behaves like a wave of wavelength  $\lambda = h/mv$ . This wave nature of the particle is used in the construction of an electron microscope. An electron microscope is, in principle, similar to that of an optical microscope. In optical microscope, using visible light, focusing is done by suitable lenses. But in an electron microscope, a beam of electrons is used in place of ordinary light and focusing is done by electric and magnetic fields. The wave theory of matter shows, that electrons accelerated to a potential of 60000 volt have a wavelength of about  $5 \times 10^{-12}$  m. This wavelength is  $10^5$  times shorter than the wavelength for visible light. As the resolving power of a microscope is proportional to the wavelength of light used, the resolving power of an electron microscope is much higher than that of an optical microscope.

**Electrostatic lens.** An electron beam passing through an electric field undergoes divergence or convergence, depending upon the potential gradients at different points in the field. The arrangement shown in Fig. 7.12 works like a lens for a beam of electrons. Metallic cylinder  $A$  is connected to the negative terminal of a source of high potential difference and  $B$  to its positive terminal. The electric lines of force between them are almost semicircular in shape. A beam of electrons moving coaxially to  $A$ , gets converged, when it passes through the electrostatic field between  $A$  and  $B$ . This system of cylinders is called an electrostatic lens and its focal length depends on the potential difference between the cylinders.

The focal length of the electrostatic lens can be changed by adjusting the P.D. of the cylinders. But the focal length of a glass lens is fixed, depending upon its radius of curvature and its material.

**Magnetic lens.** In a magnetic field, an electron beam undergoes divergence or convergence depending upon the potential gradient of different points in the field. An arrangement of magnetic coils used for converging or diverging an electron beam is called a *magnetic lens*. The focusing action of a magnetic field on electrons is identical with that of an optical lens on a beam of light and hence the arrangement is called *magnetic lens*.

Fig. 7.13 shows the electrons being emitted by a source  $O$ , then passing across a magnetic field and getting converged to a point  $I$ . The magnetic field is generally

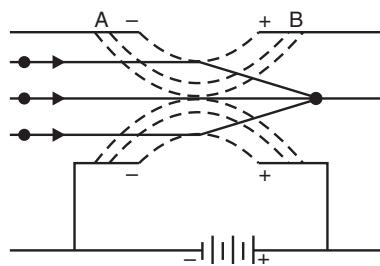


Fig. 7.12

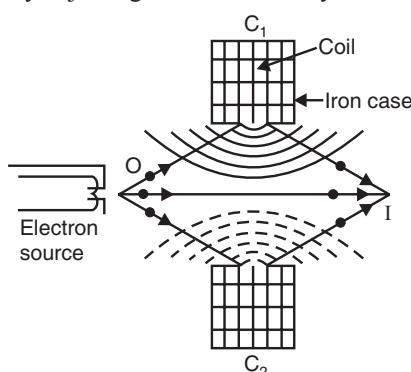


Fig. 7.13

provided by coils carrying current. Two coils  $C_1$  and  $C_2$  are enclosed in two iron cases which have one hole each. When the holes face each other, the magnetic field in the space between the two produces convergence. The focal length of such a magnetic lens depends upon the quality of the coil and the current it carries. Adjusting the currents in the coils, the intensity of the magnetic field can be adjusted and thus the focal length of the system can be adjusted. In electron microscope, generally, magnetic lenses are used.

**Construction.** The essential parts of an electron microscope are shown in Fig. 7.14. An optical microscope is shown alongside for comparison.

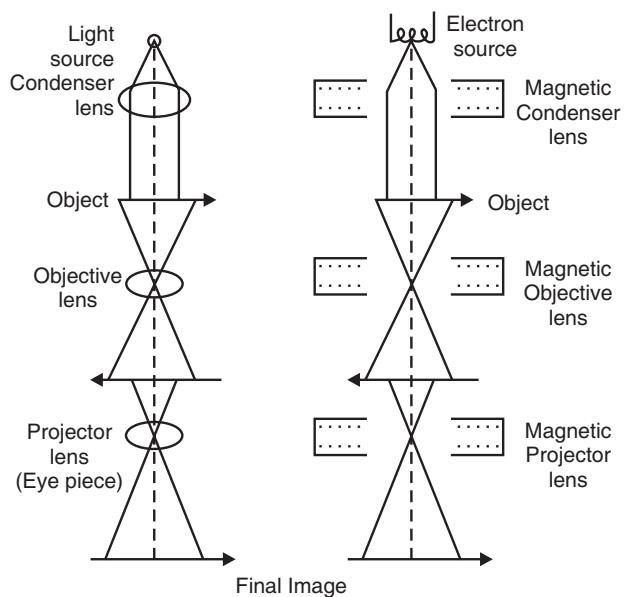


Fig. 7.14

**Working.** Electrons liberated from the source are accelerated under potentials of about 60000 V. The accelerated electron beam is then focused by a magnetic condensing lens on the object. The electrons which are scattered by the object pass on, go through the magnetic objective lens which forms the first magnified image. The projection lens forms a real magnified image on a fluorescent screen.

- Uses.**
- (1) It is used in industry to study the structure of textile fibres, surface of metals, composition of paper and paints etc.
  - (2) In medicine and biology it is used to study virus, the disease causing agent beyond the reach of ordinary microscope. The *bacteria* are shown in greater detail in an electron microscope. A knowledge of the structure of viruses enables scientist to devise methods of their destruction and thereby control diseases.
  - (3) In Physics, it has been used in the investigation of atomic structure and structure of crystals.
  - (4) In Chemistry, small particles forming colloids have become open to study and analysis.

## 7.5 HEISENBERG'S UNCERTAINTY PRINCIPLE

**Statement.** *It is impossible to determine precisely and simultaneously the values of both the members of a pair of physical variables which describe the motion of an atomic system.*

Such pairs of variables are called *canonically conjugate* variables.

**Example 1.** According to this principle, the position and momentum of a particle (say electron) cannot be determined simultaneously to any desired degree of accuracy.

Taking  $\Delta x$  as the error in determining its position and  $\Delta p$  the error in determining its momentum at the same instant, these quantities are related as follows:

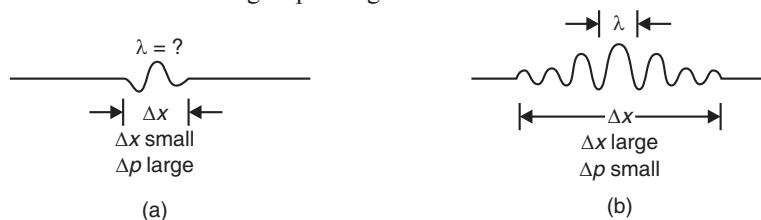
$$\Delta x \Delta p \geq \hbar/2.$$

The product of the two errors is approximately of the order of Planck's constant  $\hbar$ .

- If  $\Delta x$  is small,  $\Delta p$  will be large and vice versa.

If one quantity is measured accurately, the other quantity becomes less accurate.

**Example 2.** Consider the wave group of Fig. 7.15.



Werner Heisenberg  
(1901 – 1976)

Fig. 7.15

(a) *A narrow de Broglie wave group.* The narrower its wave group, the more precisely a particle's position can be specified (Fig. 7.15a).

But, the wavelength of the waves in a narrow packet is not well defined because there are not enough waves to measure  $\lambda$  accurately. Since  $\lambda = \frac{h}{mv}$ , the particle's momentum  $mv$  is not a precise quantity. If we make a series of momentum measurements, we will find a broad range of values.

(b) *A wide wave group.* Now the wavelength  $\lambda$  can be precisely determined (Fig. 7.15b). The momentum  $mv$  is therefore a precise quantity. But the position of the particle is not well defined.

- It is impossible to know both the exact position and exact momentum of an object at the same time.

### Illustration (i) : Determination of position with $\gamma$ -ray microscope.

Suppose we try to measure the position and linear momentum of an electron using an imaginary microscope with a very high resolving power (Fig. 7.16).

#### Uncertainty in position measurement

The electron can be observed if atleast one photon is scattered by it into the microscope lens.

The resolving power of the microscope is given by the relation

$$\Delta x = \frac{\lambda}{2 \sin \theta} \quad \dots(1)$$

Here,  $\Delta x$  is the distance between two points which can be just resolved by microscope. This is the range in which the electron would be visible when disturbed by the photon. Hence

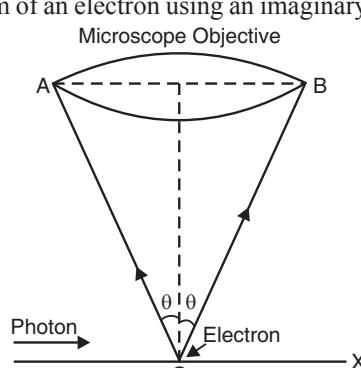


Fig. 7.16

$\Delta x$  is the uncertainty involved in the position measurement of the electron.

#### Uncertainty in momentum measurement

The incoming photon will interact with the electron through the Compton effect. To be able to see this electron, the scattered photon should enter the microscope within the angle  $2\theta$ . The momentum imparted by the photon to the electron during the impact is of the order of  $h/\lambda$ .

The component of this momentum along  $OA = -\frac{h}{\lambda} \sin \theta$

The component of this momentum along  $OB = \frac{h}{\lambda} \sin \theta$ .

The uncertainty in the momentum measurement in the  $x$ -direction is

$$\Delta p_x = \frac{h}{\lambda} \sin \theta - \left( -\frac{h}{\lambda} \sin \theta \right) = \frac{2h}{\lambda} \sin \theta. \quad \dots(2)$$

$$\therefore \Delta x \times \Delta p_x = \frac{\lambda}{2 \sin \theta} \times \frac{2h}{\lambda} \sin \theta = h.$$

A more sophisticated approach will show that  $\Delta x \Delta p_x \geq \hbar/2$ .

#### Illustration (ii) : Diffraction of a beam of electrons by a slit.

A beam of electrons is transmitted through a slit and received on a photographic plate  $P$  kept at some distance from the slit (Fig. 7.17).

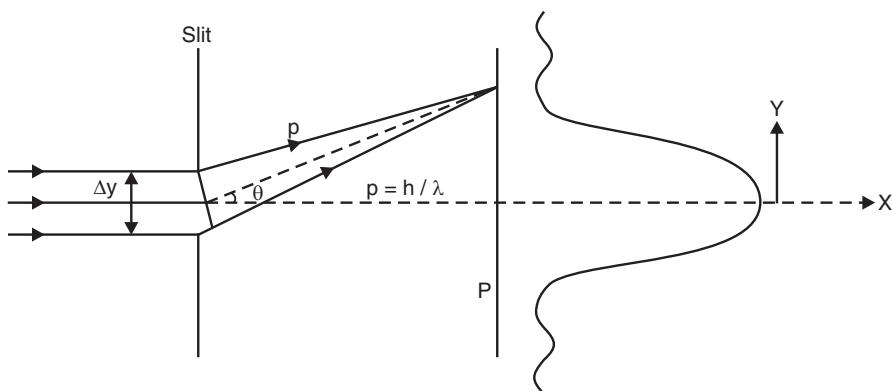


Fig. 7.17

#### ● Uncertainty in position measurement

We can only say that the electron must have passed through the slit and cannot specify its exact location in the slit as the electron crosses it. Hence the position of any electron recorded on the plate is uncertain by an amount equal to the width of the slit ( $\Delta y$ ).

$\lambda$  is the wavelength of the electrons.  $\theta$  is the angle of deviation corresponding to first minimum.

From the theory of diffraction in optics,

$$\Delta y = \frac{\lambda}{\sin \theta}. \quad \dots(1)$$

This is the uncertainty in determining the position of electron along  $y$ -axis.

#### ● Uncertainty in momentum measurement

Initially the electrons are moving along  $X$ -axis. So they have no component of momentum along  $y$ -axis. As the electrons are deviated at the slit from their initial path to form the pattern, they acquire an additional component of momentum along  $y$ -axis.

$p$  is the momentum of the electron on emerging from the slit.

The component of momentum of electron along  $y$ -axis is  $p \sin \theta$ .

The electron may be anywhere within the pattern from angle  $-\theta$  to  $+\theta$ . So, the  $y$ -component of momentum of the electron may be anywhere between  $-p \sin \theta$  and  $+p \sin \theta$ .

Therefore, the uncertainty in the  $y$ -component of momentum of the electron

$$\Delta p_y = 2p \sin \theta = \frac{2h}{\lambda} \sin \theta \quad \left( \because \lambda = \frac{h}{p} \right) \quad \dots(2)$$

$$\therefore \Delta y \Delta p_y = \frac{\lambda}{\sin \theta} \times \frac{2h}{\lambda} \sin \theta = 2h.$$

i.e.,  $\Delta y \Delta p_y \geq \frac{\hbar}{2}$ , which is Heisenberg's uncertainty principle.

### 7.5.1. Energy-Time Uncertainty Relation

Consider a free particle of mass  $m$  moving with velocity  $v$ .

Its kinetic energy is

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

Here,  $p = mv$  = momentum of particle.

Uncertainty in energy is given by

$$\Delta E = \frac{2p \Delta p}{2m} = \frac{p \Delta p}{m} = \frac{mv \Delta p}{m} = v \Delta p.$$

But

$$v = \frac{\Delta x}{\Delta t}$$

$\therefore$

$$\Delta E = \frac{\Delta x}{\Delta t} \Delta p$$

$$\Delta E \Delta t = \Delta x \Delta p$$

Uncertainty relation in terms of position and momentum is

$$\Delta x \Delta p \geq \hbar/2$$

$\therefore$

$$\Delta E \Delta t \geq \frac{\hbar}{2}.$$

### 7.5.2. Applications of Uncertainty Principle

- Non-existence of Electron in the Nucleus

**EXAMPLE 1.** Prove the nonexistence of electron in the nucleus on the basis of uncertainty principle.

**SOL.** Nuclear diameter is of the order of  $10^{-15}$  m. For an electron to exist inside a nucleus, the uncertainty in its position must be at least of this order. That is,  $\Delta x \approx 10^{-15}$  m.

The uncertainty in the electron momentum is

$$\Delta p \approx \frac{\hbar}{2(\Delta x)} \approx \frac{1.055 \times 10^{-34}}{2 \times 10^{-15}} = 0.527 \times 10^{-19} \text{ kg ms}^{-1}$$

Energy of the electron,  $E = (c^2 p^2 + m_0^2 c^4)^{1/2} \approx cp$

$$E = (3 \times 10^8 \text{ ms}^{-1}) \times (0.527 \times 10^{-19} \text{ kg ms}^{-1}) = 1.581 \times 10^{-11} \text{ J}$$

$$E = \frac{1.581 \times 10^{-11}}{1.6 \times 10^{-19}} \text{ eV} = 98.8 \text{ MeV}$$

For an electron to be a constituent of the nucleus, it should have an energy of the order of 98.8 MeV. However, the energy of an electron emitted in a  $\beta$ -decay experiment is of the order of 3 MeV. Hence, we conclude that electrons do not reside in the nucleus.

- **Zero-Point Energy of a Harmonic Oscillator**

**EXAMPLE 2.** Using the uncertainty principle, estimate the ground state energy of the harmonic oscillator.

**SOL.** Assume that the particle position is uncertain by  $\Delta x$ . Then, from the uncertainty principle, the uncertainty in particle momentum is  $\Delta p = \hbar / (2\Delta x)$ .

Total energy of the linear harmonic oscillator is

$$E = \frac{1}{2}k(\Delta x)^2 + \frac{(\Delta p)^2}{2m} = \frac{1}{2}k(\Delta x)^2 + \frac{\hbar^2}{8m(\Delta x)^2} \quad \dots(1)$$

The value of  $E$  will be minimum when  $\frac{dE}{d(\Delta x)} = 0$

$$k\Delta x - \frac{\hbar^2}{4m(\Delta x)^3} = 0 \text{ or } \Delta x = \left( \frac{\hbar^2}{4mk} \right)^{1/4}.$$

Substituting the value of  $\Delta x$  in Eq. (1), we get

$$E_{\min} = \frac{1}{2}k\left(\frac{\hbar^2}{4mk}\right)^{1/2} + \frac{\hbar^2}{8m}\left(\frac{4mk}{\hbar^2}\right)^{1/2} = \frac{\hbar}{2}\left(\frac{k}{m}\right)^{1/2} = \frac{\hbar\omega}{2}.$$

- **The Ground State Energy and the Radius of the Hydrogen Atom**

**EXAMPLE 3.** Consider an electron of momentum  $p$  in the Coulomb field of a proton. The total energy is

$$E = \frac{p^2}{2m} - \frac{e^2}{(4\pi\epsilon_0)r},$$

where  $r$  is the distance of the electron from the proton. Assuming that the uncertainty  $\Delta r$  of the radial coordinate is  $\Delta r \approx r$  and that  $\Delta p \approx p$ , use Heisenberg's uncertainty principle  $\Delta r \Delta p = \hbar$  to obtain an estimate of the size and the energy of the hydrogen atom in the ground state.

**SOL.** From the uncertainty principle,  $\Delta p \geq \frac{\hbar}{\Delta r} = \frac{\hbar}{r}$ .

The momentum  $p$  cannot be less than the uncertainty  $\Delta p$ . Hence, the minimum possible momentum is

$$p = \frac{\hbar}{r} \quad \dots(1)$$

The total energy of the electron is given by

$$E = \frac{p^2}{2m} - \frac{e^2}{4\pi\epsilon_0 r} \quad \dots(2)$$

$$\therefore E = \frac{\hbar^2}{2mr^2} - \frac{e^2}{4\pi\epsilon_0 r} \quad \dots(3)$$

The system will be in the state of lowest energy at the value of  $r$  given by

$$\frac{dE}{dr} = 0$$

$$\text{or } -\frac{\hbar^2}{mr^3} + \frac{e^2}{4\pi\epsilon_0 r^2} = 0$$

$$\text{or } r = \frac{(4\pi\epsilon_0)\hbar^2}{me^2}$$

This is same as the expression obtained for the first Bohr radius  $a_0$ . Its value is

$$r = a_0 = 0.53 \text{ \AA} \quad \dots(4)$$

Substituting in Eq. (3), the ground state energy of the hydrogen atom is

$$E_0 = \frac{\hbar^2}{2m} \left( \frac{me^2}{4\pi\epsilon_0\hbar^2} \right)^2 - \frac{e^2}{4\pi\epsilon_0} \left( \frac{me^2}{4\pi\epsilon_0\hbar^2} \right) = -\frac{me^4}{32\pi^2\epsilon_0^2\hbar^2}$$

$$\therefore E_0 = -13.6 \text{ eV} \quad \dots(5)$$

#### ● Evidence for finite width of the spectral lines

**EXAMPLE 4.** The lifetime of an excited state of an atom is about  $10^{-8}$  sec. Calculate the minimum uncertainty in the determination of the energy of the excited state.

**SOL.** We have,  $\Delta E \Delta t \geq \hbar/2$ .

$$\therefore \Delta E \geq \frac{\hbar}{2\Delta t} = \frac{1.054 \times 10^{-34}}{2(10^{-8})}$$

$$\therefore \Delta E \geq 0.527 \times 10^{-26} \text{ J} = 3.29 \times 10^{-8} \text{ eV.}$$

This is known as the *energy width* of an excited state.

The width of the spectral line when the atom de-excites to the ground state is

$$\begin{aligned} \Delta v &= \frac{\Delta E}{h} \\ &= \frac{0.527 \times 10^{-26}}{6.63 \times 10^{-34}} \\ \Delta v &= 8 \times 10^6 \text{ Hz} \end{aligned}$$

This is the limit to the accuracy with which the frequency of the radiation emitted by an atom can be determined.

#### 7.5.3. Mathematical Proof of Uncertainty Principle for One Dimensional Wave-packet

We shall derive the position-momentum uncertainty relation by using the theory of Fourier analysis. A moving particle corresponds to a single wave group. An isolated wave group is the result of superposing an infinite number of waves with different angular frequencies  $\omega$ , continuous range of wave numbers  $k$  and amplitudes (Fig. 7.18). The composition produces oscillations confined to a *single* region of space and thus provides an idealized picture of a localized matter wave.

At a certain time  $t$ , the wave group  $\psi(x)$  can be represented by the *Fourier integral*

$$\psi(x) = \int_0^\infty g(k) \cos kx dk$$

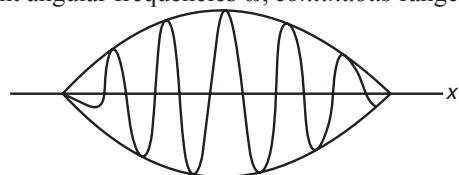


Fig. 7.18

Here the amplitude function  $g(k)$  describes how the amplitudes of the waves that contribute to  $\psi(x)$  vary with wave number  $k$ .  $\psi(x)$  and  $g(k)$  are just *Fourier transforms* of each other. Fig. 7.19

shows Gaussian distributions for the amplitude function  $g(k)$  and the wave packet  $\psi(x)$ . The relationship between the distance  $\Delta x$  and the wave number spread  $\Delta k$  depends upon the shape of the wave group and upon how  $\Delta x$  and  $\Delta k$  are defined. The widths  $\Delta x$  and  $\Delta k$  obey a reciprocal relation in which the product  $\Delta x \Delta k$  is equal to a pure number. The minimum value of the product  $\Delta x \Delta k$  occurs when the envelope of the group has the bell shape of a Gaussian function (Fig. 7.19). Thus, the Gaussian wave packets happen to be minimum uncertainty wave packets. If  $\Delta x$  and  $\Delta k$  are taken as the standard deviations of the respective functions  $\psi(x)$  and  $g(k)$ , then this minimum value is  $1/2$ . Wave groups in general do not have Gaussian forms. So we can write

$$\Delta x \Delta k \geq \frac{1}{2} \quad \dots(1)$$

Let  $\lambda$  be the de Broglie wavelength of the particle. We see from

$$k = \frac{2\pi}{\lambda} = \frac{2\pi p}{h}$$

that the momentum of the particle is determined by the wave number  $k$ .

$$\therefore p = \frac{\hbar k}{2\pi}.$$

$$\Delta p = \frac{\hbar \Delta k}{2\pi} = \hbar \Delta k.$$

Hence an uncertainty  $\Delta k$  in the wave number of the de Broglie waves associated with the particle results in an uncertainty  $\Delta p$  in the particle's momentum.

From Eq. (1),  $\Delta x \Delta k \geq 1/2$  or  $\Delta k \geq 1/2 \Delta x$

$$\therefore \Delta x \Delta p \geq \hbar/2 \quad \dots(2)$$

This is the *Heisenberg uncertainty relation for position and momentum*.

Eq. (2) states that the product of the uncertainty  $\Delta x$  in the position of an object at some instant and the uncertainty  $\Delta p$  in its momentum component in the  $x$  direction at the same instant is equal to or greater than  $\hbar/2$ .

The three-dimensional form of the Heisenberg uncertainty relations for position and momentum is now

$$\Delta x \Delta p_x \geq \frac{\hbar}{2}, \Delta y \Delta p_y \geq \frac{\hbar}{2}, \Delta z \Delta p_z \geq \frac{\hbar}{2} \quad \dots(3)$$

#### Time-Energy uncertainty relation

The theory of Fourier analysis may also be invoked to obtain a *time-energy uncertainty relation*. Indeed, according to Fourier analysis, a wave packet of duration  $\Delta t$  must be composed of plane-wave components whose angular frequencies extend over a range  $\Delta\omega$  such that  $\Delta t \Delta\omega \geq 1/2$ . Since  $E = \hbar\omega$ , we, therefore, have

$$\Delta t \Delta E \geq \hbar/2 \quad \dots(4)$$

This is the *Heisenberg uncertainty relation for time and energy*. It connects the uncertainty  $\Delta E$  in the determination of the energy of a system with the time interval  $\Delta t$  available for this energy determination. Thus, if a system does not stay longer than a time  $\Delta t$  in a given state of motion, its energy in that state will be uncertain by an amount  $\Delta E \geq \hbar/2 \Delta t$ .

#### 7.5.4. Elementary proof of Uncertainty Relation between Displacement and Momentum

Consider a particle in motion along the  $x$ -axis.

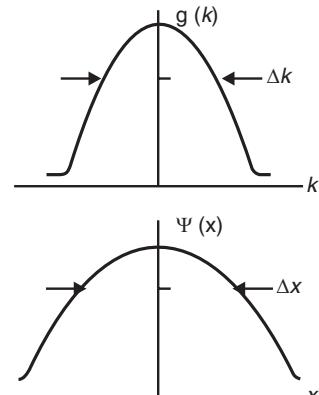


Fig. 7.19

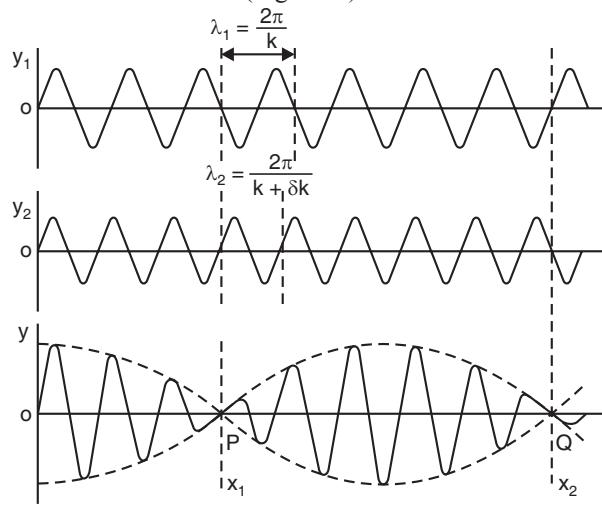
- The de Broglie relation between the wavelength  $\lambda$  of the associated wave and the momentum  $p_x$  of the particle is

$$p_x = \frac{h}{\lambda} = \hbar k \quad \dots(1)$$

Here,  $k = 2\pi/\lambda$  is called the *propagation constant*.

- The particle in motion is represented by a wave-packet.

We consider the wave-packet as the superposition of two simple harmonic plane waves of propagation constant  $k$  and  $k + \delta k$  (Fig. 7.20).



**Fig. 7.20**

The length of the wave-packet is

$$\Delta x = \frac{2\pi}{\Delta k} \quad \dots(2)$$

- The particle must be somewhere within the wave-packet.

Uncertainty in the position of the particle =  $\Delta x$ .

Uncertainty in the propagation constant =  $\Delta k$ .

From Eq. (1), the uncertainty  $\Delta p_x$  in the momentum is given by

$$\Delta p_x = \hbar \Delta k \quad \dots(3)$$

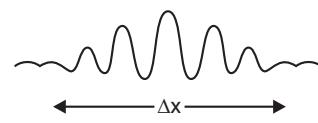
Multiplying Eq. (2) by Eq. (3), we get the product of the uncertainties.

$$\Delta x \cdot \Delta p_x = \left( \frac{2\pi}{\Delta k} \right) (\hbar \Delta k) = 2\pi\hbar = h.$$

If wave-packets have shapes different from that shown in Fig. 7.20, then the sign of equality is replaced by the sign  $\gtrsim$ .

$$\therefore \Delta x \cdot \Delta p_x \gtrsim h \quad \dots(4)$$

- A **single wave-packet** is formed by superposition of an **infinite number of plane waves** of different wave numbers  $k$  (Fig. 7.21). This wave has virtually no amplitude outside a rather narrow region of space  $\Delta x$ .



**Fig. 7.21**

For a **single wave-packet**  $\Delta x$  and  $\Delta k$  are related by:

$$\Delta x \gtrsim \frac{1}{\Delta k} \quad \dots(5)$$

Multiplying Eq. (3) by Eq. (5), we get

$$\Delta x \cdot \Delta p_x \gtrsim \hbar \quad \dots(6)$$

### 7.5.5. Elementary Proof of the Uncertainty Relation between Energy and Time

A particle in motion is represented by a wave-packet (Fig. 7.22).

$\Delta x$  is the width of the wave-packet moving along the  $x$ -axis.

$v_g$  = group velocity of the wave-packet, and

$v_x$  = particle velocity along  $x$ -axis.

The group velocity of the wave-packet is equal to the particle velocity.

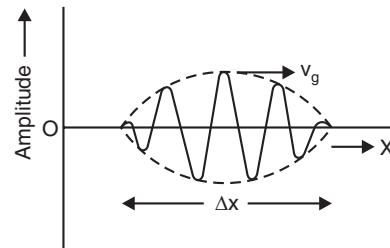


Fig. 7.22

Suppose the wave-packet moves through  $\Delta x$  in time  $\Delta t$ .

$$\begin{aligned} \Delta t &= \frac{\Delta x}{v_g} = \frac{\Delta x}{v_x} \\ \therefore \Delta t &= \frac{\Delta x}{v_x} \end{aligned} \quad \dots(1)$$

Since  $\Delta x$  is the uncertainty in the  $x$ -coordinate of the particle,  $\Delta t$  is the uncertainty in time  $t$  at which the particle passes through a point with velocity  $v_g$ .

$$\text{The kinetic energy of the particle, } E = \frac{p_x^2}{2m} \quad \dots(2)$$

Taking differential of Eq. (2), the uncertainty  $\Delta E$  in the kinetic energy is given by

$$\begin{aligned} \Delta E &= \frac{2p_x \Delta p_x}{2m} = \frac{p_x}{m} \Delta p_x = \frac{mv_x}{m} \Delta p_x \\ \Delta E &= v_x \Delta p_x \end{aligned} \quad \dots(3)$$

Multiplying Eq. (1) by Eq. (3), we get

$$\Delta E \cdot \Delta t = \Delta x \cdot \Delta p_x \quad \dots(4)$$

According to the uncertainty relation between position and momentum, we have

$$\Delta x \cdot \Delta p_x \gtrsim \hbar.$$

$\therefore$  from Eq. (4), we have

$$\Delta E \cdot \Delta t \gtrsim \hbar \quad \dots(5)$$

This is the energy-time uncertainty relation.

### 7.5.6. Bohr's Complementarity Principle

**Statement.** In any experimental situation in which a physical entity (matter or radiation) exhibits its wave properties, it is impossible to attribute corpuscular characteristics to it.

- The particle and wave aspects of a physical entity are complementary and cannot be exhibited at the same time.

**Example.** In a two slit interference experiment, if we try to define electron trajectory either by closing one slit or by placing some detecting apparatus immediately behind one of the slits, the interference pattern disappears. It follows that any experiment which can be devised either displays wave like or particle like characteristics of the system. The wave and particle pictures give complementary descriptions of the same system.

**Illustration.** Consider an experimental arrangement in which light is diffracted by a double slit and is detected on a screen (Fig. 7.23). Suppose the light beam is very weak and there is only one photon at a time in the apparatus. We can regard each photon as having a wave associated with it. The two slits transmit the photon in the manner of a wave passing through *both* slits at once. When it strikes the screen, light is behaving as a particle does. Thus the wave and particle pictures give complementary descriptions of the same system. Bohr introduced the **complementarity principle**.

*It states that the wave and particle aspects of matter are complementary rather than being contradictory, both equally essential for a full description of the phenomenon.*

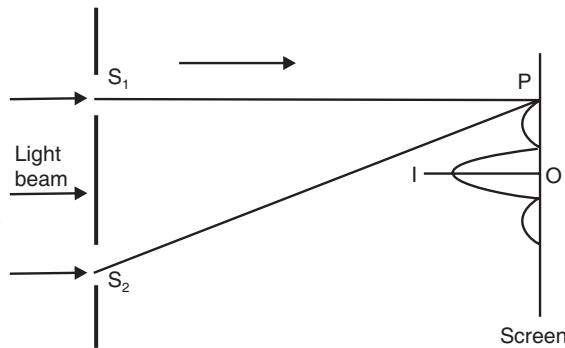


Fig. 7.23

### EXERCISE

1. The de Broglie wavelength of a particle of mass  $m$  and momentum  $p$  is given by  
 (a)  $\lambda = \frac{h}{p}$       (b)  $\lambda = \frac{p}{h}$       (c)  $\lambda = \frac{h}{2p}$       (d)  $\lambda = \frac{2p}{h}$ . (B.U. 2014)
2. If the momentum of a particle is doubled, then its de Broglie wavelength  
 (a) zero      (b) halves      (c) double      (d) unchanged.
3. De-Broglie wavelength for thermal neutrons  
 (a)  $\lambda = \frac{h}{3mkT}$       (b)  $\lambda = \frac{h}{\sqrt{2mkT}}$       (c)  $\lambda = \frac{h}{mkT}$       (d)  $\lambda = \frac{h}{\sqrt{mkT}}$ . (B.U. 2011)
4. The group velocity is given by  
 (a)  $v_g - v_p$       (b)  $v_g = \lambda v_p$       (c)  $v_g = v_p - \lambda \frac{dv_p}{d\lambda}$       (d)  $v_g = v_p - \frac{dv_p}{d\lambda}$ . (B.U. 2015)
5. The phase velocity of de Broglie wave  
 (a)  $v_p = \frac{v}{c}$       (b)  $v_p = \frac{c}{v}$       (c)  $v_p = \frac{c^2}{v}$       (d)  $v_p = \frac{v}{c^2}$ . (B.U. 2015)
6. Which one of the following is true?  
 (a)  $\Delta J \Delta \theta \geq h$       (b)  $\Delta J \Delta x \geq h$       (c)  $\Delta J \Delta p \geq h$       (d)  $\Delta J \Delta k \geq h$   
**Ans. 1. (a), 2. (b), 3. (b), 4. (c), 5. (c), 6. (a)**
7. Explain the concept of matter waves. (B.U. 2015)
8. Derive de Broglie relation for matter waves. (B.U. 2012)
9. Derive an expression for phase velocity. (B.U. 2014)
10. Derive an expression for group velocity. (B.U. 2013)
11. Show that the group velocity of the de-Broglie wave associated with a particle is the same as the particle velocity. (B.U. 2013)
12. Describe Davisson and Germer experiment for study of electron diffraction. (B.U. 2013)

- 
- 13. Describe GP. Thomson experiment on electron diffraction. (B.U. 2015)
  - 14. With principle explain the working of an electron microscope. (Madras 2006)
  - 15. State Heisenberg's uncertainty principle and give two illustrations. (B.U. 2012)
  - 16. Derive Heisenberg uncertainty relation. (B.U. 2014)
  - 17. Explain the concept of energy and time uncertainty. (B.U. 2015)
  - 18. Explain non-existence of free electrons in the nucleus.
  - 19. The lifetime of an excited state of an atom is about  $10^{-8}$  sec. Calculate the minimum uncertainty in the determination of the energy of the excited state. (B.U. 2013)

**Solution.** We have,  $\Delta E \Delta t \geq \hbar$

$$\therefore \Delta E \geq \frac{\hbar}{\Delta t} = \frac{1.055 \times 10^{-34}}{(10^{-8})}$$
$$\therefore \Delta E \geq 1.0 \times 10^{-26} \text{ J} = 6.5 \times 10^{-8} \text{ eV.}$$

This is known as the *energy width* of an excited state.

- 20. A microscope using photons is employed to locate an electron in an atom to within a distance of  $0.2 \text{ \AA}$ . What is the uncertainty in the momentum of the electron located in this way? (Madras, Nov. 2006)

[Hint:  $\Delta x \Delta p \sim \hbar$ . Given:  $\Delta x = 0.2 \times 10^{-10} \text{ m}$   
 $\Delta p \sim 5.27 \times 10^{-24} \text{ kg m/s}$ ]

# SCHRÖDINGER EQUATION AND ITS APPLICATIONS

**After reading this chapter, you should be able to**

- ◆ Explain why Schrödinger's wave equation is not valid for relativistic particles
- ◆ Outline the probability interpretation of the wave function and apply operators to it to obtain information about a particle's physical properties such as position, momentum and energy
- ◆ Utilize Schrödinger equation to find the allowed energy levels of quantum mechanical systems
- ◆ Solve the Schrödinger equation to obtain wave functions for some basic, physically important types of potential (potential wells, barriers and harmonic oscillator), and estimate the shape of the wavefunction based on the shape of the potential
- ◆ Explain the alpha decay of a nucleus on the basis of the tunnel effect.

## 8.1 DERIVATION OF TIME-DEPENDENT FORM OF SCHRÖDINGER EQUATION

The quantity that characterises the de Broglie waves is called the *wave function*. It is denoted by  $\psi$ . It may be a complex function. Let us assume that  $\psi$  is specified in the  $x$  direction by

$$\psi = Ae^{-i\omega(t-x/v)} \quad \dots(1)$$

If  $v$  is the frequency, then  $\omega = 2\pi v$  and  $v = \lambda/\nu$ .

$$\therefore \psi = Ae^{-2\pi i(vt-x/\lambda)} \quad \dots(2)$$

Let  $E$  be the total energy and  $p$  the momentum of the particle.

$$\text{Then, } v = \frac{E}{h} \text{ and } \lambda = \frac{h}{p}.$$

Making these substitutions in Eq. (2),

$$\begin{aligned} \psi &= Ae^{-(2\pi i/h)(Et-px)} \\ \psi &= Ae^{-(i/h)(Et-px)} \end{aligned} \quad \dots(3)$$

Eq. (3) is a mathematical description of the wave equivalent of an unrestricted particle of total energy  $E$  and momentum  $p$  moving in the  $+x$  direction.



*Erwin Schrödinger  
(1887 – 1961)*

Differentiating Eq. (3) twice with respect to  $x$ , we get

$$\begin{aligned} \frac{\partial^2 \psi}{\partial x^2} &= -\frac{p^2}{\hbar^2} \psi \\ p^2 \psi &= -\hbar^2 \frac{\partial^2 \psi}{\partial x^2} \end{aligned} \quad \dots(4)$$

Differentiating Eq. (3) once with respect to  $t$ , we get

$$\begin{aligned} \frac{\partial \psi}{\partial t} &= -\frac{iE}{\hbar} \psi \\ E\psi &= -\frac{\hbar}{i} \frac{\partial \psi}{\partial t} \end{aligned} \quad \dots(5)$$

At speeds small compared with that of light, the total energy  $E$  of a particle is the sum of its kinetic energy  $p^2/2m$  and its potential energy  $V$ .  $V$  is in general a function of position  $x$  and time  $t$ .

$$\therefore E = \frac{p^2}{2m} + V \quad \dots(6)$$

Multiplying both sides of Eq. (6) by  $\psi$ , we get

$$E\psi = \frac{p^2\psi}{2m} + V\psi \quad \dots(7)$$

Substituting the expressions for  $E\psi$  and  $p^2\psi$  from Eqs. (5) and (4) into Eq. (7), we obtain

$$\begin{aligned} -\frac{\hbar}{i} \frac{\partial \psi}{\partial t} &= -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} + V\psi \\ \text{or} \quad i\hbar \frac{\partial \psi}{\partial t} &= -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} + V\psi \end{aligned} \quad \dots(8)$$

Eq. (8) is the *time-dependent form of Schrödinger's equation in one dimension*.

In three dimensions, the time-dependent form of Schrödinger's equation is

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \left( \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) + V\psi.$$

Here, the particle's potential energy  $V$  is some function of  $x, y, z$ , and  $t$ .

### Schrödinger's equation : Steady-state form — Time independent Schrödinger equation

In a great many situations, the potential energy of a particle does not depend upon time explicitly. The forces that act upon it, and hence  $V$ , vary with the position of the particle only. When this is true, Schrödinger's equation may be simplified by removing all reference to  $t$ .

The one-dimensional wave function  $\psi$  of an unrestricted particle may be written in the form

$$\begin{aligned}\psi &= Ae^{-(i/\hbar)(Et - px)} = Ae^{-(iE/\hbar)t} \cdot e^{+(ip/\hbar)x} \\ \therefore \psi &= \psi_0 e^{-(iE/\hbar)t}\end{aligned}\dots(1)$$

Here,  $\psi_0 = Ae^{+(ip/\hbar)x}$ . That is,  $\psi$  is the product of a position dependent function  $\psi_0$  and a time-dependent function  $e^{-(iE/\hbar)t}$ .

Differentiating Eq. (1) with respect to  $t$ , we get

$$\frac{\partial\psi}{\partial t} = -\frac{iE}{\hbar}\psi_0 e^{-(iE/\hbar)t}\dots(2)$$

Differentiating Eq. (1) twice with respect to  $x$ , we get

$$\frac{\partial^2\psi}{\partial x^2} = \frac{\partial^2\psi_0}{\partial x^2} e^{-(iE/\hbar)t}\dots(3)$$

We can substitute these values in the time-dependent form of Schrödinger's equation

$$\begin{aligned}i\hbar \frac{\partial\psi}{\partial t} &= -\frac{\hbar^2}{2m} \frac{\partial^2\psi}{\partial x^2} + V\psi \\ \therefore E\psi_0 e^{-(iE/\hbar)t} &= -\frac{\hbar^2}{2m} \frac{\partial^2\psi_0}{\partial x^2} e^{-(iE/\hbar)t} + V\psi_0 e^{-(iE/\hbar)t}\end{aligned}$$

Dividing through by the common exponential factor, we get

$$\frac{\partial^2\psi_0}{\partial x^2} + \frac{2m}{\hbar^2} (E - V) \psi_0 = 0\dots(4)$$

Eq. (4) is the *steady-state form of Schrödinger's equation*.

In three dimensions, it is

$$\nabla^2 \psi_0 + \frac{2m}{\hbar^2} (E - V) \psi_0 = 0\dots(5)$$

Usually it is written in the form

$$\nabla^2 \psi + \frac{2m}{\hbar^2} (E - V) \psi = 0$$

Steady-state form of Schrödinger's equation in three dimensions is

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} + \frac{2m}{\hbar^2} (E - V) \psi = 0$$

### Eigenvalues and Eigenfunctions

The values of energy  $E_n$  for which Schrödinger's steady-state equation can be solved are called **eigenvalues**. The corresponding wave functions  $\psi_n$  are called **eigenfunctions**.

**Degeneracy.** If there is more than one linearly independent wave function belonging to the same energy eigenvalue  $E$ , the energy level is said to be degenerate. If there are  $g$  linearly independent wave functions ( $\psi_1, \psi_2, \dots, \psi_g$ ) belonging to the same energy state, then the energy level is said to be  $g$ -fold degenerate.

The degree of degeneracy is defined as the number of linearly independent eigen functions corresponding to the same eigen value.

### Operator for Momentum

The wave function for a free particle moving in the positive  $x$ -direction is

$$\psi(x, t) = Ae^{i/\hbar(p_x x - Et)} \quad \dots(1)$$

Differentiating Eq. (1) with respect to  $x$ , we get

$$\begin{aligned} \frac{\partial \psi}{\partial x} &= A \left( \frac{i}{\hbar} \right) p_x e^{i/\hbar(p_x x - Et)} = \frac{i}{\hbar} p_x \psi \\ \text{or} \quad \frac{\hbar}{i} \frac{\partial \psi}{\partial x} &= p_x \psi \\ \therefore -i\hbar \frac{\partial}{\partial x} \psi &= p_x \psi \end{aligned} \quad \dots(2)$$

Evidently the dynamical quantity  $p_x$  in some sense corresponds to the differential operator  $-i\hbar \frac{\partial}{\partial x}$ .

We denote operators by using a caret.

$\hat{p}_x$  is the operator that corresponds to  $x$ -component of the momentum  $p_x$ .

From Eq. (2), for the  $x$ -component of the momentum, the operator is

$$\hat{p}_x = -i\hbar \frac{\partial}{\partial x} \quad \dots(3)$$

The **operator representation** of  $p$  is

$$\hat{p} = -i\hbar \nabla \quad \dots(4)$$

### To obtain the expectation value of the momentum $p_x$

The expectation value of  $p_x$  is found by using the corresponding differential operator.

$$\langle p_x \rangle = \int_{-\infty}^{+\infty} \psi^*(x, t) \left( -i\hbar \frac{\partial}{\partial x} \right) \psi(x, t) dx$$

#### 8.1.1. Properties of the Wave Function

##### Physical significance of $\psi$

- By applying Schrödinger's equation to the motion of a particle, we get the wave function  $\psi$ . The future course of the particle's motion—like its initial state—is a matter of probabilities instead of certainties.
- The quantity whose variations make up matter waves is called the **wave function**,  $\Psi$ . This function can be either real or complex.

The wave function  $\Psi$  itself, however, has no direct physical significance.

- The only quantity having a physical meaning is the square of its magnitude  $P = |\psi|^2 = \psi\psi^*$ .

Here,  $\psi^*$  is the complex conjugate of  $\psi$ .

The quantity  $P$  is the probability density.

- $|\Psi|^2$ , the square of the absolute value of the wave function, is called **probability density**.

The probability of experimentally finding the body described by the wave function  $\Psi$  at the point  $x, y, z$  at the time  $t$  is proportional to the value of  $|\Psi|^2$  there at  $t$ .

A large value of  $|\Psi|^2$  means the strong possibility of the body's presence.

A small value of  $|\Psi|^2$  means the slight possibility of its presence.

This interpretation was made by Max Born.

- The probability of finding a particle in a volume  $dx dy dz$  is  $|\psi|^2 dx dy dz$ .
- The particle is certainly to be found somewhere in space.

$$\therefore \iiint |\psi|^2 dx dy dz = 1.$$

The triple integral extends over all possible values of  $x, y, z$ .

A wave function ( $\psi$ ) satisfying this relation is called a *normalised wave function*.

#### Orthogonal wave function

If the product of a function  $\psi_1(x)$  and the complex conjugate  $\psi_2^*(x)$  of a function  $\psi_2(x)$  vanishes when integrated with respect to  $x$  over the interval  $a \leq x \leq b$ , that is, if

$$\int_a^b \psi_2^*(x) \psi_1(x) dx = 0,$$

then  $\psi_1(x)$  and  $\psi_2(x)$  are said to be orthogonal in the interval  $(a, b)$ .

#### Normalised wave function

The probability of finding a particle in the volume element  $dV$  is given by  $\psi\psi^* dV$ .

The total probability of finding the particle in the entire space is unity, i.e.,

$$\int \psi\psi^* dV = \int |\psi|^2 dV = 1.$$

Here, the integration extends over all space.

Any wave function satisfying the above equation is said to be normalised.

Very often  $\psi$  is not a normalized wave function. We know that it is possible to multiply  $\psi$  by a constant  $A$ , to give a new wave function,  $A\psi$ , which is also a solution of the wave equation. Now the problem is to choose the proper value of  $A$  such that the new wave function is a normalized function. In order that it is a normalized function, it must meet the requirement

$$\int (A\psi)^* A\psi dx dy dz = 1$$

$$\text{or, } |A|^2 \int \psi\psi^* dx dy dz = 1$$

$$\text{or } |A|^2 = \frac{1}{\int \psi\psi^* dx dy dz}$$

$|A|$  is called normalizing constant.

**Orthonormal functions.** The functions which are orthogonal and also normalised are called orthonormal functions.

#### Requirements of wave function

To arrive at results consistent with physical observations, several additional requirements are imposed on the wave function  $\psi(x)$ :

1. It must be well behaved, that is, single-valued and continuous everywhere.
2. If  $\psi_1(x), \dots, \psi_n(x)$  are solutions of Schrödinger equation, then the linear combination  $\psi(x) = a_1 \psi_1(x) + a_2 \psi_2(x) + \dots + a_n \psi_n(x)$  must be a solution.
3. The wave function  $\psi(x)$  must approach zero as  $x \rightarrow \pm\infty$

## APPLICATIONS OF SCHRÖDINGER'S EQUATION

### 8.2 THE FREE PARTICLE

A free particle is defined as one which is subject to no forces of any kind, and so, moves in a region of constant potential. The particle motion is confined to the  $x$ -axis only. We shall consider the potential to be zero,  $V=0$ .

The time-independent Schrödinger equation becomes

$$\frac{d^2 \psi}{dx^2} + \frac{2mE}{\hbar^2} \psi = 0 \quad \dots(1)$$

Define a quantity  $k$ , called the wave number, by

$$k = \sqrt{\frac{2mE}{\hbar^2}} \quad \dots(2)$$

$$\therefore \frac{d^2 \psi}{dx^2} + k^2 \psi = 0 \quad \dots(3)$$

The solution of Eq. (3) is

$$\psi = A e^{ikx} + B e^{-ikx} \quad \dots(4)$$

Let us set  $B = 0$ . Then

$$\psi = A e^{ikx} \quad \dots(5)$$

Eq. (5) describes a particle moving in the positive  $x$ -direction.

A problem appears in normalising the wave function. We ought to have

$$\int_{-\infty}^{\infty} \psi^* \psi dx = A^2 \int_{-\infty}^{\infty} dx = 1 \quad \dots(6)$$

$\int_{-\infty}^{\infty} dx$  is infinite. So  $A$  must be zero. This difficulty arises since we are considering an ideal

case of infinite length. In practice, the particle is confined to a finite length so that normalisation is possible.

The probability density for the particle is  $\psi^* \psi = A^2$ , a constant independent of  $x$ . Consequently, the particle is equally likely to be found anywhere. So we have an infinite amount of uncertainty in its position, *i.e.*,  $\Delta x = \infty$ . According to the expression  $\Delta p \Delta x \geq \hbar/2$ ,  $\Delta p$  must be zero.

Let us prove this.

$$\langle p \rangle = \int_{-\infty}^{\infty} \psi^* \left( -i\hbar \frac{\partial}{\partial x} \right) \psi dx = \hbar k \int_{-\infty}^{\infty} \psi^* \psi dx = \hbar k$$

Notice that  $\langle p \rangle$  is positive, meaning that our particle has a momentum of exactly  $\hbar k$  and is moving toward the right.

$$\begin{aligned} \langle p^2 \rangle &= \int_{-\infty}^{\infty} \psi^* \left( -\hbar^2 \frac{\partial^2}{\partial x^2} \right) \psi dx = -\hbar^2 (-k^2) \int_{-\infty}^{\infty} \psi^* \psi dx \\ &= \hbar^2 k^2 \\ \therefore (\Delta p)^2 &= \langle p^2 \rangle - \langle p \rangle^2 = 0. \end{aligned}$$

Thus the momentum of the particle is precisely defined.

Note that for the particle in a square well, the energy is quantised. But for the free particle, there is no such restriction on the energy. Generally, bound systems (such as electrons trapped in an atom)

will give rise to discrete energy levels. Unbound systems (such as two atoms that collide) will give rise to an energy continuum.

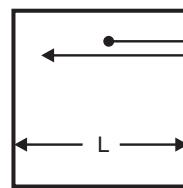
### 8.3 THE PARTICLE IN A BOX: INFINITE SQUARE WELL POTENTIAL

Consider a particle moving inside a box along the  $x$ -direction. The particle is bouncing back and forth between the walls of the box. The box has insurmountable potential barriers at  $x = 0$  and  $x = L$ . i.e., the box is supposed to have walls of infinite height at  $x = 0$  and  $x = L$  (Fig. 8.1). The particle has a mass  $m$  and its position  $x$  at any instant is given by  $0 < x < L$ .

The potential energy  $V$  of the particle is infinite on both sides of the box. The potential energy  $V$  of the particle can be assumed to be zero between  $x = 0$  and  $x = L$ .

In terms of the boundary conditions imposed by the problem, the potential function is

$$\begin{aligned} V &= 0 \text{ for } 0 < x < L \\ V &= \infty \text{ for } x \leq 0 \\ V &= \infty \text{ for } x \geq L \end{aligned}$$



The particle cannot exist outside the box and so its wave function  $\psi$  is 0 for  $x \leq 0$  and  $x \geq L$ . Our task is to find what  $\psi$  is within the box, viz., between  $x = 0$  and  $x = L$ .

Within the box, the Schrödinger's equation becomes

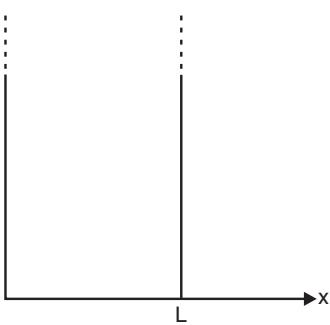
$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2} E\psi = 0 \quad \dots(1)$$

The general solution of this equation is

$$\psi = A \sin \frac{\sqrt{2mE}}{\hbar} x + B \cos \frac{\sqrt{2mE}}{\hbar} x \quad \dots(2)$$

$A$  and  $B$  are constants to be evaluated.

Fig. 8.1



The boundary conditions are used to evaluate the constants  $A$  and  $B$  in Eq. (2).

(i)  $\psi = 0$  at  $x = 0$ . Hence  $B = 0$ .

(ii)  $\psi = 0$  at  $x = L$ .

$\psi$  will be 0 at  $x = L$  only when

$$\frac{\sqrt{2mE}}{\hbar} L = n\pi \quad n = 1, 2, 3, \dots \quad \dots(3)$$

From Eq. (3) it is clear that the energy of the particle can have only certain values, which are the eigen values.

These eigenvalues are found by solving Eq. (3) for  $E_n$ .

$$\therefore E_n = \frac{n^2\pi^2\hbar^2}{2mL^2} \quad n = 1, 2, 3, \dots \quad \dots(4)$$

For each value of  $n$ , there is an energy level. Each value of  $E_n$  is called an *eigenvalue*. Thus inside the box, the particle can only have the *discrete* energy values specified by Eq. (2). Note also that the particle cannot have zero energy.

#### Wave Functions

The wave functions of a particle in a box whose energies are  $E_n$  are, from Eq. (2) with  $B = 0$ .

$$\psi_n = A \sin \frac{\sqrt{2mE_n}}{\hbar} x \quad \dots(5)$$

Substituting Eq. (4) for  $E_n$  gives

$$\psi_n = A \sin \frac{n\pi x}{L} \quad \dots(6)$$

$\psi_n$  are the eigenfunctions corresponding to the energy eigenvalues  $E_n$ .

#### Evaluation of $A$ and Normalization of the Wave Function

It is certain that the particle is somewhere inside the box. Hence for a normalised wave function

$$\int_0^L \psi * \psi dx = 1 \text{ i.e., } A^2 \int_0^L \sin^2 \left( \frac{n\pi x}{L} \right) dx = 1$$

$$\text{i.e., } A^2 \int_0^L \left( \frac{1 - \cos 2n\pi x/L}{2} \right) dx = 1 \text{ or } A^2 \frac{L}{2} = 1$$

$$\text{or } A = \sqrt{\frac{2}{L}}$$

The normalized wave functions of the particle are

$$\psi_n = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L} \quad n = 1, 2, 3, \dots \quad \dots(7)$$

The normalized wave functions  $\psi_1$ ,  $\psi_2$ , and  $\psi_3$  together with the probability densities  $|\psi_1|^2$ ,  $|\psi_2|^2$ , and  $|\psi_3|^2$  are plotted in Fig. 8.2.

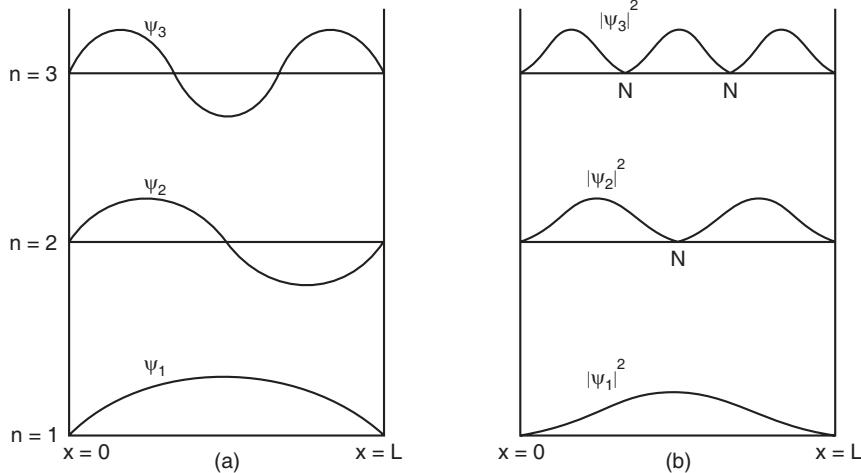


Fig. 8.2

**EXAMPLE 1.** Calculate the permitted energy levels of an electron, in a box 1 Å wide.

**SOL.** Here,  $m$  = mass of the electron =  $9.1 \times 10^{-31}$  kg;  $L = 1\text{Å} = 10^{-10}$  m,  $E_n = ?$

$$\therefore \text{The permitted electron energies} = E_n = \frac{n^2 \pi^2 \hbar^2}{2mL^2}$$

$$E_n = \frac{n^2 \pi^2 (1.054 \times 10^{-34})^2}{2(9.1 \times 10^{-31})(10^{-10})^2} = 6 \times 10^{-18} n^2 \text{ J} = 38n^2 \text{ eV.}$$

The minimum energy, the electron can have, is  $E_1 = 38 \text{ eV}$ , corresponding to  $n = 1$ .

The other values of energy are  $E_2 = 4E_1 = 152 \text{ eV}$ ,  $E_3 = 9E_1 = 342 \text{ eV}$  and so on.

**EXAMPLE 2.** Calculate the expectation value  $\langle p_x \rangle$  of the momentum of a particle trapped in a one-dimensional box.

**SOL.** The normalized wave functions of the particle are

$$\begin{aligned}\psi_n^* &= \psi_n = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L} \\ \frac{d\psi}{dx} &= \sqrt{\frac{2}{L}} \left( \frac{n\pi}{L} \right) \cos \frac{n\pi x}{L} \\ \text{Now, } \langle p_x \rangle &= \int_{-\infty}^{\infty} \psi^* \left( -i\hbar \frac{d}{dx} \right) \psi dx \\ &= -i\hbar \frac{2}{L} \frac{n\pi}{L} \int_0^L \sin \frac{n\pi x}{L} \cos \frac{n\pi x}{L} dx \\ &= 0\end{aligned}$$

The expectation value  $\langle p_x \rangle$  of the particle's momentum is 0.

**EXAMPLE 3.** Find the expectation value  $\langle x \rangle$  of the position of a particle trapped in a box  $L$  wide.

$$\begin{aligned}\text{SOL. } \langle x \rangle &= \int_{-\infty}^{\infty} x |\psi|^2 dx = \frac{2}{L} \int_0^L x \sin^2 \frac{n\pi x}{L} dx \\ &= \frac{2}{L} \left[ \frac{x^2}{4} - \frac{x \sin(2n\pi x/L)}{4n\pi/L} - \frac{\cos(2n\pi x/L)}{8(n\pi/L)^2} \right]_0^L \\ \therefore \langle x \rangle &= \frac{2}{L} \left( \frac{L^2}{4} \right) = \frac{L}{2}.\end{aligned}$$

This result means that the average position of the particle is the middle of the box in all quantum states.

**EXAMPLE 4.** Verify the orthogonal property of wave functions of a particle in a one dimensional box.

**SOL.** The wave functions of a particle in one dimensional box are

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L}$$

We want to verify

$$\int_0^L \psi_n^*(x) \psi_m(x) dx = 0 \text{ for } n \neq m.$$

$$\begin{aligned}\text{L.H.S.} &= \int_0^L \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L} \sqrt{\frac{2}{L}} \sin \frac{m\pi x}{L} dx = \frac{1}{L} \int_0^L \left( 2 \sin \frac{n\pi x}{L} \sin \frac{m\pi x}{L} \right) dx \\ &= \frac{1}{L} \int_0^L \left[ \cos(n-m) \frac{\pi x}{L} - \cos(n+m) \frac{\pi x}{L} \right] dx\end{aligned}$$

$$\therefore \text{L.H.S.} = 0. \text{ Hence verified.}$$

### 8.4 FINITE SQUARE POTENTIAL WELL

Fig. 8.3 shows a square potential well of width  $L$ . Consider a particle with energy  $E$  moving along the  $x$ -axis. The energy  $E$  of the trapped particle is less than the height  $U$  of the barriers. Schrodinger's steady-state equation for regions I and III is

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E-U)\psi = 0$$

Put  $\frac{\sqrt{2m(U-E)}}{\hbar} = a$  ... (1)

$$\frac{d^2\psi}{dx^2} - a^2\psi = 0 \quad x < 0, x > L, \quad \dots(2)$$

The solutions to Eq. (2) are :

$$\psi_I = A e^{ax} + B e^{-ax} \quad \dots(3)$$

$$\psi_{III} = C e^{ax} + D e^{-ax} \quad \dots(4)$$

For  $\psi$  to remain finite as  $x \rightarrow \pm\infty$ , we must have  $B = 0$  and  $C = 0$ .

$$\therefore \psi_I = A e^{ax} \quad \dots(5)$$

$$\psi_{III} = D e^{-ax} \quad \dots(6)$$

These wave functions decrease exponentially inside the barriers at the sides of the well.

Schrödinger's equation for region II is

$$\frac{d^2\psi}{dx^2} + \frac{2mE}{\hbar^2}\psi = 0 \quad \dots(7)$$

The solution of Eq. (7) is

$$\psi_{II} = F \sin \frac{\sqrt{2mE}}{\hbar} x + G \cos \frac{\sqrt{2mE}}{\hbar} x \quad \dots(8)$$

At the boundaries  $x = 0$  and  $x = L$ , the wave functions inside and outside must have the same value and the same slope.

$$\psi_{II} = \psi_I \text{ and } \frac{d\psi_{II}}{dx} = \frac{d\psi_I}{dx} \text{ at } x = 0.$$

$$\psi_{II} = \psi_{III} \text{ and } \frac{d\psi_{II}}{dx} = \frac{d\psi_{III}}{dx} \text{ at } x = L.$$

These conditions when applied to Eqs. (5), (6) and (8) yield the following relations :

$$G = A \quad \dots(9)$$

$$F \frac{\sqrt{2mE}}{\hbar} = Aa \quad \dots(10)$$

$$F \sin \frac{\sqrt{2mE}}{\hbar} L + G \cos \frac{\sqrt{2mE}}{\hbar} L = D e^{-aL} \quad \dots(11)$$

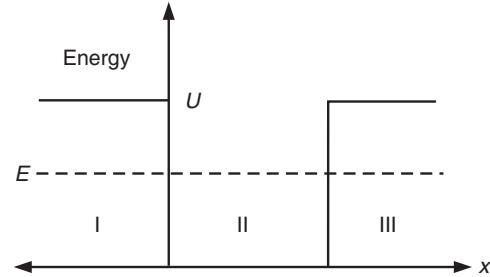


Fig. 8.3

$$F \frac{\sqrt{2mE}}{\hbar} \cos\left(\frac{\sqrt{2mE}}{\hbar} L\right) - G \frac{\sqrt{2mE}}{\hbar} \sin\left(\frac{\sqrt{2mE}}{\hbar} L\right) = -Da e^{-aL} \quad \dots(12)$$

Eqs. (9) and (10) give

$$F = Ga \frac{\hbar}{\sqrt{2mE}} \quad \dots(13)$$

Eq. (13) together with Eqs. (11) and (12) yields

$$\begin{aligned} \tan\left(\frac{\sqrt{2mE}}{\hbar} L\right) &= \frac{2a(\sqrt{2mE}/\hbar)}{\left(\frac{2mE}{\hbar^2} - a^2\right)} \\ \therefore \quad \tan\left(\frac{\sqrt{2mE}}{\hbar} L\right) &= \frac{2\sqrt{E(U-E)}}{2E-U} \end{aligned} \quad \dots(14)$$

Only those values of  $E$  which satisfy this relation are the allowed energy states. Thus the energy is quantised. The allowed energy values are found by numerical or graphical methods.

The wave functions for the first three allowed energy values and the corresponding probability densities for finding the particle at different locations are shown in Figure 8.4 (a) and (b) respectively. The particle has a certain probability of being found outside the wall.

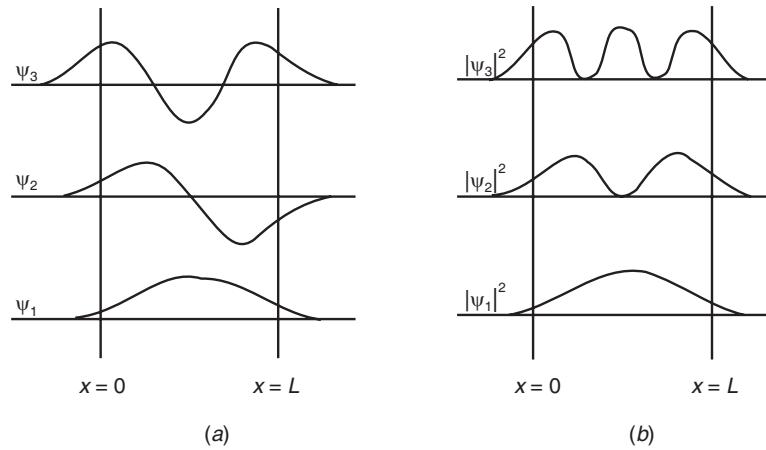


Fig. 8.4

## 8.5 RECTANGULAR POTENTIAL WELL

A one dimensional rectangular potential well is defined by

$$V(x) = \begin{cases} 0 & (x < 0), \\ -V_0 & (0 < x < a), \\ 0 & (x > a), \end{cases} \quad \dots(1)$$

$V_0$ , the *depth of the well*, is a positive number (Fig. 8.5). When total energy of the particle is negative ( $-V_0 < E < 0$ ), the particle is confined to the well and forms bound states.  $m$  is the mass of the particle.

The Schrödinger equation for the three regions are

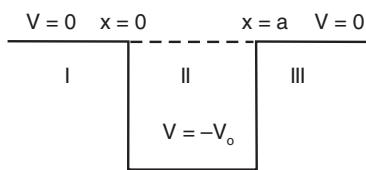


Fig. 8.5

$$\frac{d^2 \psi_1}{dx^2} + \frac{2mE}{\hbar^2} \psi_1 = 0 \text{ or } \frac{d^2 \psi_1}{dx^2} - \lambda^2 \psi_1 = 0 \quad (x < 0)$$

$$\frac{d^2 \psi_2}{dx^2} + \frac{2m(E+V_0)}{\hbar^2} \psi_2 = 0 \text{ or } \frac{d^2 \psi_2}{dx^2} + k^2 \psi_2 = 0 \quad (0 < x < a)$$

$$\frac{d^2 \psi_3}{dx^2} + \frac{2mE}{\hbar^2} \psi_3 = 0 \text{ or } \frac{d^2 \psi_3}{dx^2} - \lambda^2 \psi_3 = 0 \quad (x > a)$$

Here,  $\lambda = \sqrt{\frac{-2mE}{\hbar^2}}$  and  $k = \sqrt{\frac{2m(E+V_0)}{\hbar^2}}$

The solutions of the Schrödinger equation outside the potential well are now  $e^{\lambda x}$  and  $e^{-\lambda x}$ . Here  $\lambda$  is real. When  $x \rightarrow \infty$ ,  $e^{\lambda x}$  is not bounded. Hence it must not appear in  $\psi$  for  $x > a$ . Similarly,  $\psi$  must not contain  $e^{-\lambda x}$  for  $x < 0$ .

The (unnormalised) wave functions are

$$\begin{aligned} \psi_1 &= e^{\lambda x} & (x < 0), \\ \psi_2 &= A e^{ikx} + B e^{-ikx} & (0 < x < a) \\ \psi_3 &= C e^{-\lambda x} & (x > a) \end{aligned} \quad \dots(2)$$

The continuity of  $\psi$  and  $d\psi/dx$  at  $x = 0$  and  $x = a$  require that

$$\begin{aligned} \psi_1 &= \psi_2 \text{ at } x = 0 \\ \frac{d\psi_1}{dx} &= \frac{d\psi_2}{dx} \text{ at } x = 0 \\ \psi_2 &= \psi_3 \text{ at } x = a \\ \frac{d\psi_2}{dx} &= \frac{d\psi_3}{dx} \text{ at } x = a. \end{aligned}$$

These conditions give four relations between the three constants  $A, B, C$ .

$$\begin{aligned} 1 &= A + B \\ \lambda &= ik(A - B) \\ Ce^{-\lambda a} &= Ae^{ik a} + Be^{-ik a} \\ -\lambda Ce^{-\lambda a} &= ik(Ae^{ik a} - Be^{-ik a}) \end{aligned} \quad \dots(3)$$

These equations will give a unique solution, only if

$$2 \cot ka = \frac{k}{\lambda} - \frac{\lambda}{k} \quad \dots(4)$$

This is a transcendental equation in  $k$ .

If this condition is satisfied, the solution is

$$A = B^* = \frac{1}{2} \left( 1 - i \frac{\lambda}{k} \right), C = \frac{1}{2} \left( \frac{k}{\lambda} + \frac{\lambda}{k} \right) e^{\lambda a} \sin ka \quad \dots(5)$$

To solve Eq. (4) graphically, introduce the quantities

$$\gamma = \sqrt{\frac{2mV_0 a^2}{\hbar^2}}, \alpha = \frac{ka}{\gamma} = \sqrt{1 + \frac{E}{V_0}} \quad \dots(6)$$

Then Eq. (4) can be expressed as

$$\gamma \alpha = (n-1)\pi + 2\cos^{-1}\alpha, (n=1, 2, \dots) \quad \dots(7)$$

Fig. 8.6 illustrates this relation graphically for the two cases  $\gamma = 1, 4$ . The abscissae of the points of intersection (denoted by dots) of the two graphs give the energies of the corresponding states according to the formula

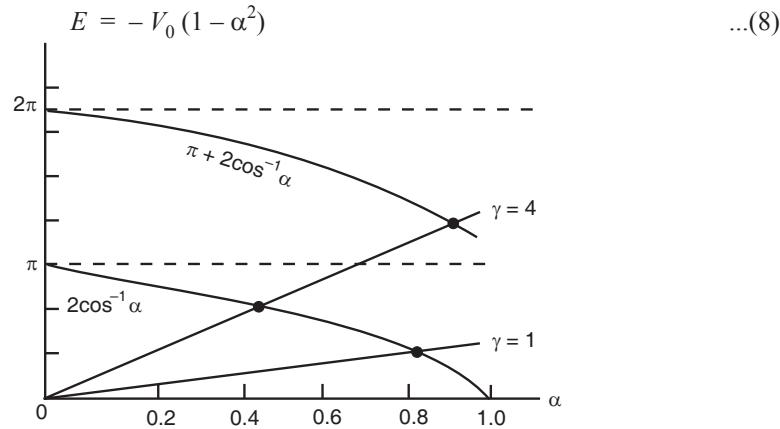


Fig. 8.6

The number of states for a given well depth is clearly the greatest integer contained in the quantity  $(\gamma/\pi + 1)$ . This quantity increases as  $V_0$  is made larger. The values of  $E$  corresponding to the stationary states decreases as the well depth increases. A new level appears at zero energy each time  $\gamma$  assumes the value  $n\pi$ . Energy-level diagrams for the two cases of Fig. 8.6 are drawn in Fig. 8.7.  $n$  is the *quantum number* ranking the energy levels in increasing order.

Substituting the values of  $A$ ,  $B$  and  $C$  from Eq. (5) in Eq. (2),

$$\begin{aligned} \psi_1 &= e^{\lambda x} && (x < 0), \\ \psi_2 &= \frac{1}{\alpha} \sin \left[ k \left( x - \frac{a}{2} \right) + \frac{n\pi}{2} \right] && (0 < x < a), \\ \psi_3 &= (-1)^{n+1} e^{-\lambda(x-a)} && (x > a) \end{aligned} \quad \dots(9)$$

$\psi$  is zero at  $x = \pm\infty$  and at  $(n-1)$  points within the region  $0 < x < a$ . So the total number of zeros of  $\psi$  is  $n+1$ . The wave function for  $n=5$  is shown in Fig. 8.8.

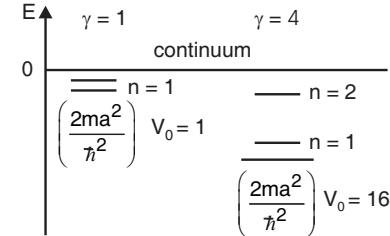


Fig. 8.7

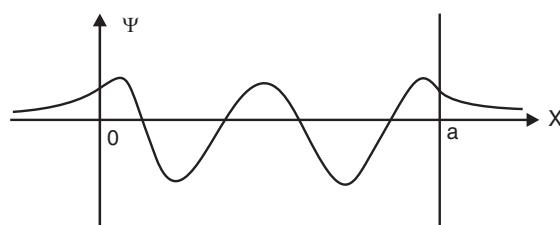


Fig. 8.8

For an *infinitely deep potential well*, the straight line of Fig. 8.6 approaches the vertical axis. The points of intersection approach the values

$$k a = n \pi \quad (n = 1, 2, \dots) \quad \dots(10)$$

Hence the energy levels are

$$\frac{\hbar^2 k^2}{2m} = \frac{n^2 \pi^2 \hbar^2}{2ma^2} \quad \dots(11)$$

For  $E > 0$ , well behaved solutions can be obtained for all values of  $E$ . So the energy spectrum is a continuum in this case.

## 8.6 THE SQUARE WELL IN THREE DIMENSIONS

Consider a particle of mass  $m$  which is restricted to move in a box with sides  $a$ ,  $b$  and  $c$  (Fig. 8.9). The potential function  $V(x, y, z)$  is having a constant value of zero in the regions given as follows :

$$\begin{aligned} V(x, y, z) &= 0, & 0 < x < a, \\ V(x, y, z) &= 0, & 0 < y < b, \\ V(x, y, z) &= 0, & 0 < z < c. \end{aligned}$$

The potential outside the box is infinite.

The Schrödinger time independent wave equation for the particle inside the box is given by

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + \frac{2mE}{\hbar^2} \psi = 0 \quad \dots(1)$$

Let us solve this equation by the method of separation of variables. Assume that the function  $\psi$  can be written as a product of three functions,  $X$ ,  $Y$ , and  $Z$ . Each function depends on only one of the coordinates.

$$\psi(x, y, z) = X(x) Y(y) Z(z) = XYZ \quad \dots(2)$$

$$\frac{\partial^2 \psi}{\partial x^2} = X'' YZ, \frac{\partial^2 \psi}{\partial y^2} = XY'' Z, \frac{\partial^2 \psi}{\partial z^2} = XYZ''. \quad \dots(2)$$

Substituting this in Eq. (2) and dividing by  $XYZ$ , we get

$$\frac{X''}{X} + \frac{Y''}{Y} + \frac{Z''}{Z} + \frac{2mE}{\hbar^2} = 0 \quad \dots(3)$$

We notice that each term of this equation depends on a different variable and that the three variables are independent. The last term is constant. The only way for the equation to remain valid for all values of  $x$ ,  $y$ , and  $z$  in our interval is for *each* term of Eq. (3) to be constant. Therefore

$$\frac{X''}{X} = -\alpha^2, \frac{Y''}{Y} = -\beta^2, \frac{Z''}{Z} = -\gamma^2$$

where  $\alpha$ ,  $\beta$  and  $\gamma$  are constants.  $\alpha^2 + \beta^2 + \gamma^2 = 2mE/\hbar^2$

These equations can be written as

$$\frac{d^2 X}{dx^2} + \alpha^2 X = 0 \quad \dots(4)$$

$$\frac{d^2 Y}{dy^2} + \beta^2 Y = 0 \quad \dots(5)$$

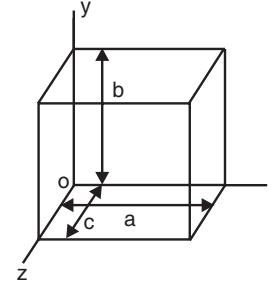


Fig. 8.9

$$\frac{d^2 Z}{dz^2} + \gamma^2 Z = 0 \quad \dots(6)$$

The general solutions of (4), (5) and (6) are given by

$$X = A_1 \sin \alpha x + B_1 \cos \alpha x$$

$$Y = A_2 \sin \beta y + B_2 \cos \beta y$$

$$Z = A_3 \sin \gamma z + B_3 \cos \gamma z$$

Values of the constants  $A_1, A_2, A_3, B_1, B_2, B_3$  can be found by applying boundary conditions. The boundary conditions require that the wave function vanishes at the box walls where the potential is infinite, i.e.,

$$\psi(0, y, z) = \psi(x, 0, z) = \psi(x, y, 0) = 0 \quad \dots(7)$$

$$\psi(a, y, z) = \psi(x, b, z) = \psi(x, y, c) = 0 \quad \dots(8)$$

Applying boundary conditions (7) to the above equations,

$$B_1 = B_2 = B_3 = 0$$

Applying boundary conditions (8) to the above equations

$$\sin \alpha a = 0, \text{ i.e., } \alpha a = n_x \pi \text{ or } \alpha = n_x \pi/a$$

$$\sin \beta b = 0, \text{ i.e., } \beta b = n_y \pi \text{ or } \beta = n_y \pi/b$$

$$\sin \gamma c = 0, \text{ i.e., } \gamma c = n_z \pi \text{ or } \gamma = n_z \pi/c$$

Here  $n_x, n_y, n_z$  are integers of which none is equal to zero.

$$\therefore X = A_1 \sin \frac{n_x \pi x}{a} \quad \dots(9)$$

$$Y = A_2 \sin \frac{n_y \pi y}{b} \quad \dots(10)$$

$$Z = A_3 \sin \frac{n_z \pi z}{c} \quad \dots(11)$$

Substituting these values in Eq. (2), we get

$$\psi(x, y, z) = A_1 A_2 A_3 \sin \frac{n_x \pi x}{a} \sin \frac{n_y \pi y}{b} \sin \frac{n_z \pi z}{c}$$

$$= A \sin \frac{n_x \pi x}{a} \sin \frac{n_y \pi y}{b} \sin \frac{n_z \pi z}{c}$$

In the above equation  $A$  is the normalisation constant.

$A$  can be found by using the normalisation condition

$$\int_0^a \int_0^b \int_0^c \psi \psi^* dx dy dz = 1 \quad \dots(12)$$

$$A^2 \int_0^a \int_0^b \int_0^c \sin^2 \frac{n_x \pi x}{a} \sin^2 \frac{n_y \pi y}{b} \sin^2 \frac{n_z \pi z}{c} dx dy dz = 1$$

$$\therefore A^2 \frac{a}{2} \cdot \frac{b}{2} \cdot \frac{c}{2} = 1$$

$$\text{or } A = \frac{2\sqrt{2}}{\sqrt{(abc)}} \quad \dots(13)$$

The normalised wave function is

$$\psi(x, y, z) = \frac{2\sqrt{2}}{\sqrt{(abc)}} \sin \frac{n_x \pi x}{a} \sin \frac{n_y \pi y}{b} \sin \frac{n_z \pi z}{c} \quad \dots(14)$$

We have  $\alpha^2 + \beta^2 + \gamma^2 = 2mE/\hbar^2$

$$\begin{aligned} \frac{n_x^2 \pi^2}{a^2} + \frac{n_y^2 \pi^2}{b^2} + \frac{n_z^2 \pi^2}{c^2} &= \frac{2mE}{\hbar^2} \\ \therefore E_{n_x n_y n_z} &= \frac{\hbar^2 \pi^2}{2m} \left[ \frac{n_x^2}{a^2} + \frac{n_y^2}{b^2} + \frac{n_z^2}{c^2} \right] \end{aligned} \quad \dots(15)$$

Let us examine the case in which the particle is confined in a cubical region. Then  $a = b = c$ . The energy levels of Eq. (15) become

$$E_{\text{cube}} = \frac{\hbar^2 \pi^2}{2ma^2} (n_x^2 + n_y^2 + n_z^2) \quad \dots(16)$$

The *ground state energy* value is obtained by putting

$$\begin{aligned} n_x &= n_y = n_z = 1 \\ \therefore E_{111} &= \frac{3\pi^2 \hbar^2}{2ma^2} \end{aligned} \quad \dots(17)$$

There is only one set of quantum numbers that gives this energy state, and this level is said to be *non-degenerate*.

There are three possibilities for the first excited state :

$$\begin{array}{ll} n_x = 2 & n_y = n_z = 1 \\ n_y = 2 & n_x = n_z = 1 \\ n_z = 2 & n_x = n_y = 1 \end{array} \quad E_{211} = E_{121} = E_{112} = \frac{3\pi^2 \hbar^2}{ma^2} \quad \dots(18)$$

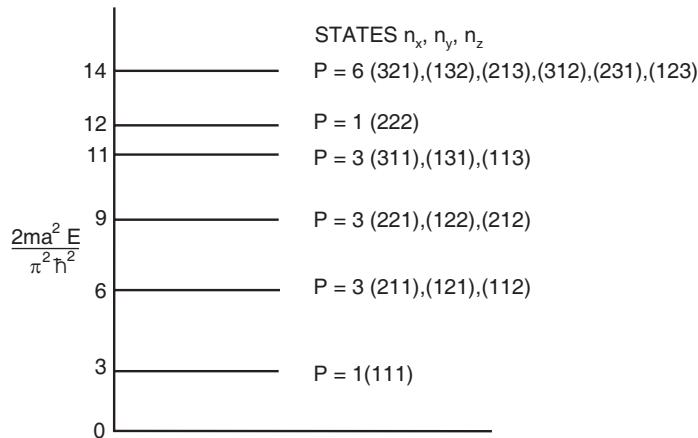


Fig. 8.10

Here there are three sets  $(211)$ ,  $(121)$  and  $(112)$  of the quantum numbers  $n_x$ ,  $n_y$ , and  $n_z$  that will give the same energy level. That is, several distinct quantum states possess the same energy. This property is called *degeneracy*. In this particular case, the level is *triply degenerate*. Fig. 8.10 shows energy levels, degree of degeneracy and quantum numbers of a particle in a cubical box.

Eq. (15) indicates that the energy values of a particle in an infinitely deep potential well are not continuous but *discrete*. The particle has different bound states inside the well.

## 8.7 POTENTIAL STEP

The potential function of a potential step is defined by

$$\begin{aligned} V(x) &= 0 & x < 0 \\ &= V_0 & x > 0 \end{aligned} \quad \dots(1)$$

Let electrons of energy  $E$  move from left to right, i.e., along the positive direction of  $x$ -axis (Fig. 8.11). It is desired to find the eigenfunction solutions of the time-independent Schrödinger equation

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E - V)\psi = 0 \quad \dots(2)$$

For I region  $V(x) = 0$ . Therefore, the Schrödinger equation takes the form

$$\frac{d^2\psi}{dx^2} + \frac{2mE}{\hbar^2}\psi = 0 \quad \dots(3)$$

The solution of Eq. (3) is

$$\psi_1 = Ae^{ip_1 x/\hbar} + Be^{-ip_1 x/\hbar} \quad \dots(4)$$

where  $A$  and  $B$  are constants.

$$p_1 = \sqrt{(2mE)}.$$

Some particles may be reflected by the potential barrier and some transmitted. The first and second terms respectively represent the *incident* and *reflected* particles.

The Schrödinger wave equation for II region is

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E - V_0)\psi = 0 \quad \dots(5)$$

The solution of Eq. (5) is

$$\psi_2 = Ce^{ip_2 x/\hbar} + De^{-ip_2 x/\hbar} \quad \dots(6)$$

where  $p_2 = \sqrt{[2m(E - V_0)]}$ ;  $C$  and  $D$  are constants.

In Eq. (6), the first term represents the *transmitted wave*. The second term represents a wave coming from  $+\infty$  in the negative direction. Clearly for  $x > 0$  no particles can flow to the left and  $D$  must be zero. Therefore, Eq. (6) becomes

$$\psi_2 = Ce^{ip_2 x/\hbar} \quad \dots(7)$$

The continuity of  $\psi$  implies that  $\psi_1 = \psi_2$  at  $x = 0$

$$\therefore A + B = C. \quad \dots(8)$$

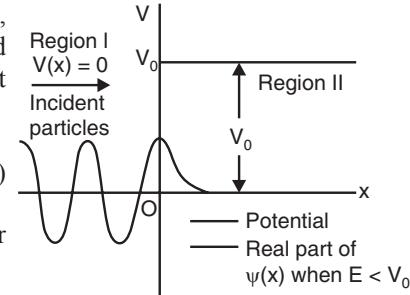
The continuity of  $\frac{d\psi}{dx}$  implies that  $\frac{d\psi_1}{dx} = \frac{d\psi_2}{dx}$  at  $x = 0$ .

$$\therefore p_1(A - B) = p_2C \quad \dots(9)$$

Solving (8) and (9) we get

$$B = \frac{p_1 - p_2}{p_1 + p_2}A \quad \dots(10)$$

$$\text{and } C = \frac{2p_1}{p_1 + p_2}A \quad \dots(11)$$



**Fig. 8.11**

$B$  and  $C$  represents the amplitudes of *reflected* and *transmitted* beams respectively in terms of the amplitude of the incident wave.

The *reflectance* and the *transmittance* at the potential discontinuity may be defined as follows:

$$\text{Reflectance } R = \frac{\text{magnitude of reflected current}}{\text{magnitude of incident current}}$$

$$\text{Transmittance } T = \frac{\text{magnitude of transmitted current}}{\text{magnitude of incident current}}$$

Two cases may arise : (i)  $E > V_0$  and (ii)  $E < V_0$

**Case (i):**  $E > V_0$ . When  $E > V_0$ ,  $p_2 = \sqrt{[2m(E - V_0)]}$  is real.

We will now derive expressions for the current density in the I and II regions.

The *probability current* is defined as

$$J = \frac{\hbar}{2im} [\psi^* \nabla \psi - \psi \nabla \psi^*] \quad \dots(12)$$

$$\begin{aligned} \therefore (J_x)_I &= \frac{\hbar}{2im} \left[ \psi_1^* \frac{d\psi_1}{dx} - \psi_1 \frac{d\psi_1^*}{dx} \right] \\ &= \frac{\hbar}{2im} \left[ \left\{ (A^* e^{-ip_1 x/\hbar} + B^* e^{ip_1 x/\hbar}) \times \left( \frac{ip_1}{\hbar} \right) (A e^{ip_1 x/\hbar} - B e^{-ip_1 x/\hbar}) \right\} \right. \\ &\quad \left. - \left\{ (A e^{ip_1 x/\hbar} + B e^{-ip_1 x/\hbar}) \times \left( -\frac{ip_1}{\hbar} \right) (A^* e^{-ip_1 x/\hbar} - B^* e^{ip_1 x/\hbar}) \right\} \right] \\ &= \frac{p_1 (AA^* - B^*B)}{m} = \frac{p_1}{m} [|A|^2 - |B|^2] \end{aligned} \quad \dots(13)$$

From the above relation it is evident that the current in the I region is equal to the difference between two terms. The first term which is proportional to  $p_1 |A|^2$  represents the *incident wave*. The second term which is proportional to  $p_1 |B|^2$  represents the *reflected wave*.

$$\left. \begin{array}{l} \text{The probability current} \\ \text{of the incident beam} \end{array} \right\} = |A|^2 \frac{p_1}{m} \quad \dots(14)$$

$$\left. \begin{array}{l} \text{The probability current} \\ \text{of the reflected beam} \end{array} \right\} = |B|^2 \frac{p_1}{m} \quad \dots(15)$$

The expression for the probability current in region II is

$$\begin{aligned} (J_x)_{II} &= \frac{\hbar}{2im} \left[ \psi_2^* \frac{d\psi_2}{dx} - \psi_2 \frac{d\psi_2^*}{dx} \right] \\ &= \frac{\hbar}{2im} \left[ \left\{ C^* e^{-ip_2 x/\hbar} \left( \frac{ip_2}{\hbar} \right) C e^{ip_2 x/\hbar} \right\} \right. \\ &\quad \left. - \left\{ C e^{ip_2 x/\hbar} \left( -\frac{ip_2}{\hbar} \right) C^* e^{-ip_2 x/\hbar} \right\} \right] \end{aligned}$$

$$\begin{aligned}
 &= \frac{p_2}{2m} [CC^* + CC^*] = \frac{p_2}{m} (CC^*) \\
 &= \frac{|C|^2 p_2}{m}
 \end{aligned} \quad \dots(16)$$

Eq. (16) represents the *transmitted current*.

$$\begin{aligned}
 R &= \frac{\text{magnitude of reflected current}}{\text{magnitude of incident current}} \\
 &= \frac{|B|^2 (p_1/m)}{|A|^2 (p_1/m)} \\
 R &= \frac{(p_1 - p_2)^2}{(p_1 + p_2)^2} \text{ from Eq. (10)} \quad \dots(17)
 \end{aligned}$$

$$\begin{aligned}
 T &= \frac{\text{magnitude of transmitted current}}{\text{magnitude of incident current}} \\
 &= \frac{|C|^2 (p_2/m)}{|A|^2 (p_1/m)} \\
 &= \left( \frac{2 p_1}{p_1 + p_2} \right)^2 \times \frac{p_2}{p_1} \text{ from Eq. (11)} \\
 \therefore T &= \frac{4 p_1 p_2}{(p_1 + p_2)^2} \quad \dots(18)
 \end{aligned}$$

**Case (ii).**  $E < V_0$ . When  $E < V_0$ ,  $p_2 = \sqrt{[2m(E - V_0)]}$  is imaginary.

Hence  $p_2 = i\sqrt{[2m(V_0 - E)]}$  and  $p_2^* = -i\sqrt{[2m(V_0 - E)]} = -p_2$ .

The probability current in this case is given by

$$\begin{aligned}
 J_x &= \frac{\hbar}{2im} \left[ \Psi_2^* \frac{d\Psi_2}{dx} - \Psi_2 \frac{d\Psi_2^*}{dx} \right] \\
 &= \frac{\hbar}{2im} \left[ C^* e^{-ip_2^* x/\hbar} \left( \frac{ip_2}{\hbar} \right) C e^{ip_2 x/\hbar} - C e^{ip_2 x/\hbar} \left( -\frac{ip_2^*}{\hbar} \right) C^* e^{-ip_2^* x/\hbar} \right]
 \end{aligned}$$

Substituting  $p_2^* = -p_2$  we get,

$$\begin{aligned}
 J_x &= \frac{\hbar}{2im} \left[ C^* e^{ip_2 x/\hbar} \left( \frac{ip_2}{\hbar} \right) C e^{ip_2 x/\hbar} - C C^* \left( \frac{ip_2}{\hbar} \right) e^{ip_2 x/\hbar} e^{ip_2 x/\hbar} \right] \\
 &= 0
 \end{aligned}$$

Thus the transmitted current is zero.

$$\begin{aligned}
 T &= \frac{\text{magnitude of transmitted current}}{\text{magnitude of incident current}} = 0 \\
 \therefore T &= 0 \quad \dots(19)
 \end{aligned}$$

By definition,  $R + T = 1$

$$\therefore R = 1 \quad \dots(20)$$

### 8.8 THE BARRIER PENETRATION PROBLEM

Consider a beam of particles of kinetic energy  $E$  incident from the left on a potential barrier of height  $V$  and width  $OA = L$  (Fig. 8.12).  $V > E$  and on both sides of the barrier,  $V = 0$ , which means that no forces act upon the particles there. This potential is described by

$$\begin{aligned} V &= 0 \text{ for } x < 0 && \text{(region I)} \\ V &= V \text{ for } 0 < x < L && \text{(region II)} \\ V &= 0 \text{ for } x > L && \text{(region III)} \end{aligned} \quad \dots(1)$$

Let  $\psi_1$ ,  $\psi_2$ , and  $\psi_3$  be the respective wave functions in regions I, II and III as indicated in the figure.

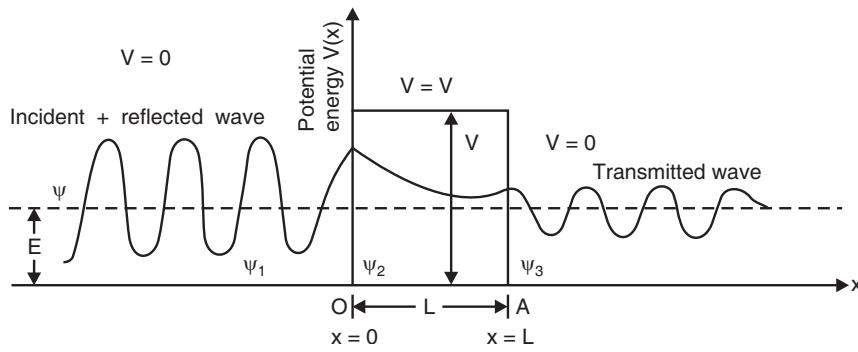


Fig. 8.12

The corresponding Schrödinger equations are

$$\begin{aligned} \text{region I} \quad & \frac{d^2\psi_1}{dx^2} + \frac{2mE}{\hbar^2}\psi_1 = 0 \quad \text{since } V = 0 \\ \text{region II} \quad & \frac{d^2\psi_2}{dx^2} + \frac{2m}{\hbar^2}(E-V)\psi_2 = 0 \quad (\because V = V) \\ \text{region III} \quad & \frac{d^2\psi_3}{dx^2} + \frac{2mE}{\hbar^2}\psi_3 = 0 \quad (\because V = 0) \end{aligned} \quad \dots(2)$$

$$\text{Put } \frac{2mE}{\hbar^2} = \alpha^2 \text{ and } \frac{2m}{\hbar^2}(V-E) = \beta^2.$$

Then the equations become

$$\begin{aligned} \text{region I} \quad & \frac{d^2\psi_1}{dx^2} + \alpha^2\psi_1 = 0 \\ \text{region II} \quad & \frac{d^2\psi_2}{dx^2} - \beta^2\psi_2 = 0 \\ \text{region III} \quad & \frac{d^2\psi_3}{dx^2} + \alpha^2\psi_3 = 0 \end{aligned} \quad \dots(3)$$

The solutions to these equations are

$$\begin{aligned} \text{region I} \quad & \psi_1 = Ae^{i\alpha x} + Be^{-i\alpha x} \\ \text{region II} \quad & \psi_2 = Fe^{-\beta x} + Ge^{\beta x} \end{aligned} \quad \dots(4)$$

$$\text{region III} \quad \psi_3 = Ce^{i\alpha x} + De^{-i\alpha x}$$

where the constants  $A$ ,  $B$  and so on are the amplitudes of the corresponding components of each wave. They may be recognized as follows:

$A$  is the amplitude of the wave, incident on the barrier from the left,

$B$  is the amplitude of the reflected wave in region I,

$F$  is the amplitude of the wave, penetrating the barrier in region II,

$G$  is the amplitude of the reflected wave (from the surface at  $A$ ) in region II,

$C$  is the amplitude of the transmitted wave, in region III, and

$D$  is the amplitude of a (nonexistent) reflected wave, in region III.

It should be noted that we have drawn the wave function through the three regions in Fig. 8.12 so that it is continuous and singly valued everywhere along the  $x$ -axis.

Since the probability density associated with a wave function is proportional to the square of the amplitude of that function, we can define the barrier transmission coefficient as

$$T = \frac{|C|^2}{|A|^2} \quad \dots(5)$$

and a reflection coefficient for the barrier surface at  $x = 0$  as

$$R = \frac{|B|^2}{|A|^2} \quad \dots(6)$$

If the barrier is high, compared to the total energy of the particle, or is thick compared to the wavelength of the wave function, then the transmission coefficient becomes

$$T \approx 16 \frac{E}{V} \left(1 - \frac{E}{V}\right) \exp\left[-\frac{2L}{\hbar} \sqrt{2m(V-E)}\right] \quad \dots(7)$$

where  $L$  is the physical thickness of the barrier. The ratio  $\frac{|C|^2}{|A|^2}$  is also called the ‘penetrability’ of

the barrier. It represents the probability that a particle incident on the barrier from one side will appear on the other side. Such a probability is zero classically, but a finite quantity in quantum mechanics. We thus conclude that if a particle with energy  $E$  is incident on a thin energy barrier of height greater than  $E$ , there is a finite probability of the particle penetrating the barrier. This phenomenon is called the *tunnel effect*. This effect was used by George Gamow in 1928 to explain the process of  $\alpha$ -decay exhibited by radioactive nuclei.

**EXAMPLE.** The potential barrier problem is a good approximation to the problem of an electron trapped inside but near the surface of a metal. Calculate the probability of transmission that a 1.0 eV electron will penetrate a potential barrier of 4.0 eV when the barrier width is 2.0 Å.

**SOL.** From equation (7), the transmission coefficient is

$$\begin{aligned} T &\approx 16 \left(\frac{1.0 \text{ eV}}{4.0 \text{ eV}}\right) \left(1 - \frac{1.0 \text{ eV}}{4.0 \text{ eV}}\right) \\ &\times \exp\left[-\frac{2 \times 2 \times 10^{-10} \text{ m}}{1.05 \times 10^{-34} \text{ Js}} \sqrt{2(9.1 \times 10^{-31} \text{ kg})(4-1)(1.6 \times 10^{-19} \text{ J})}\right] \\ &\approx 0.084 \end{aligned}$$

Thus, only about eight 1.0 eV electrons, out of every hundred, penetrate the barrier.

### Calculation of Transmission Coefficient T

The arbitrary constants  $A, B, F, G$  and  $C$  are found from the boundary conditions at  $x = 0$  and  $x = L$ .

The wave function and its derivative must be continuous at  $x = 0$  and  $x = L$ .

$$\psi_1 = A e^{i\alpha x} + B e^{-i\alpha x}$$

$$\psi_2 = F e^{-\beta x} + G e^{\beta x}$$

$$\psi_3 = C e^{i\alpha x}$$

At

$$x = 0, \quad \psi_1(0) = \psi_2(0). \quad \therefore \quad A + B = F + G \quad \dots(1)$$

At

$$x = 0, \quad \psi_1'(0) = \psi_2'(0). \quad \therefore \quad i\alpha(A - B) = -\beta(F - G)$$

or

$$A - B = \frac{i\beta}{\alpha} (F - G) \quad \dots(2)$$

At

$$x = L, \quad \psi_2(L) = \psi_3(L) \quad \therefore \quad F e^{-\beta L} + G e^{\beta L} = C e^{i\alpha L} \quad \dots(3)$$

At

$$x = L, \quad \psi_2'(L) = \psi_3'(L)$$

$$\therefore -\beta[F e^{-\beta L} - G e^{\beta L}] = i\alpha C e^{i\alpha L} \quad \dots(4)$$

Solving for  $F$  and  $G$  we get

$$F = \frac{C}{2} e^{i\alpha L} \left( 1 - \frac{i\alpha}{\beta} \right) e^{\beta L} = \frac{C}{2\beta} (\beta - i\alpha) e^{i\alpha L} e^{\beta L}$$

$$G = \frac{C}{2} e^{i\alpha L} \left( 1 + \frac{i\alpha}{\beta} \right) e^{-\beta L} = \frac{C}{2\beta} (\beta + i\alpha) e^{i\alpha L} e^{-\beta L}$$

Using Eqs. (1) and (2) we get

$$A = \frac{C}{4} e^{i\alpha L} \left[ -\frac{(\beta - i\alpha)^2}{i\alpha\beta} e^{\beta L} + \frac{(\beta + i\alpha)^2}{i\alpha\beta} e^{-\beta L} \right]$$

$$\frac{C}{A} = \frac{4i\alpha\beta e^{-i\alpha L}}{(\beta + i\alpha)^2 e^{-\beta L} - (\beta - i\alpha)^2 e^{\beta L}}$$

$$= \frac{4i\alpha\beta e^{-i\alpha L}}{2(\alpha^2 - \beta^2) \sinh \beta L + 4i\alpha\beta \cosh \beta L}$$

$$\therefore \left| \frac{C}{A} \right|^2 = \frac{4\alpha^2\beta^2}{(\alpha^2 - \beta^2)^2 \sinh^2 \beta L + 4\alpha^2\beta^2 \cosh^2 \beta L}$$

$$= \frac{4\alpha^2\beta^2}{4\alpha^2\beta^2 + (\alpha^2 + \beta^2)^2 \sinh^2 \beta L}$$

$$\therefore T = \left| \frac{C}{A} \right|^2 = \left[ 1 + \frac{(\alpha^2 + \beta^2)^2}{4\alpha^2\beta^2} \sinh^2 \beta L \right]^{-1}$$

$$\text{But } (\alpha^2 + \beta^2)^2 = \left( \frac{2m}{\hbar^2} \right)^2 V^2, \quad \alpha^2 \beta^2 = \left( \frac{2m}{\hbar^2} \right)^2 E(V - E)$$

$$\therefore T = \left[ 1 + \frac{V^2}{4E(V-E)} \sinh^2 \beta L \right]$$

If  $\beta L$  is large, then  $\sinh \beta L$  behaves as  $e^{\beta L}/2$ .

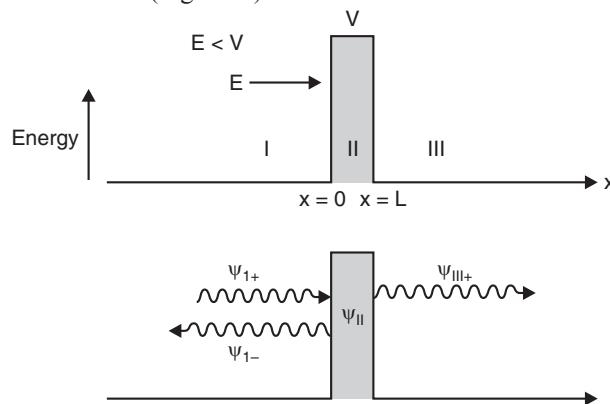
$$\sinh^2 \beta L \approx \frac{1}{4} e^{2\beta L} \gg 1; \quad \beta = \frac{\sqrt{2m(V-E)}}{\hbar}$$

$$\therefore T = 16 \frac{E}{V} \left( 1 - \frac{E}{V} \right) e^{-\frac{2L}{\hbar} \sqrt{2m(V-E)}}$$

### 8.8.1. Tunnel Effect

Consider a potential barrier of height  $V$ , with  $E < V$ .

The barrier has a finite width (Fig. 8.13).



**Fig. 8.13**

The de Broglie waves that correspond to the particle are partly reflected and partly transmitted.

This means that the particle has a finite chance of penetrating the barrier.

The particle has a certain probability of passing through the barrier and emerging on the other side. The particle lacks the energy to go over the top of the barrier. But the particle can tunnel through the walls. The higher the barrier and the wider it is, the less the chance that the particle can get through.

### Explanation of Alpha Decay

The decay of a nucleus by alpha particle emission is explained on the basis of quantum tunneling. An  $\alpha$ -particle whose kinetic energy is only a few MeV is able to escape from nucleus whose potential wall is 25 MeV high. The probability of escape is so small that the  $\alpha$ -particle might have to strike the wall  $10^{38}$  or more times before it emerges. But sooner or later it does get out.

### Approximate transmission probability

Consider a beam of identical particles all of which have the kinetic energy  $E$ . The beam is incident from region I on a potential barrier of height  $V$  and width  $L$  (Fig. 8.13). On both sides of the barrier  $V = 0$ .

This means that no forces act on particles in regions I and III.

- The wave function  $\psi_{I+}$  represents the incoming particles moving towards the barrier from region I.
- The wave function  $\psi_{I-}$  represents the reflected particles moving to the left away from the barrier.

- The wave function  $\psi_{III}$  represents the transmitted particles moving to the region III.

The transmission probability  $T$  for a particle to pass through the barrier is given by

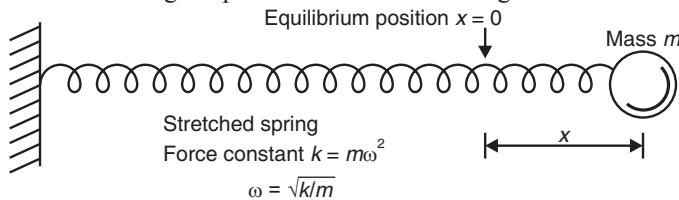
$$T = \frac{\text{Number of particles transmitted}}{\text{Number of particles incident}}.$$

This probability is approximately given by

$$T = \exp\left[-\frac{2L}{\hbar}\sqrt{2m(V-E)}\right]$$

## 8.9 LINEAR HARMONIC OSCILLATOR

Consider a particle executing simple harmonic motion along the  $x$  direction (Fig. 8.14).



**Fig. 8.14**

Let  $k$  be the restoring force per unit displacement.

$$\text{The P.E. of the particle} = \int_0^x kx \, dx = \frac{1}{2}kx^2.$$

The Schrödinger equation for the harmonic oscillator is

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2} \left(E - \frac{1}{2}kx^2\right) \psi = 0 \quad \dots(1)$$

It is convenient to simplify Eq. (1) by introducing the dimensionless quantities

$$y = \left(\frac{1}{\hbar} \sqrt{km}\right)^{1/2} x = \sqrt{\frac{2\pi m v}{\hbar}} x \quad \dots(2)$$

$$\text{and} \quad \alpha = \frac{2E}{\hbar} \sqrt{\frac{m}{k}} = \frac{2E}{\hbar v} \quad \dots(3)$$

where  $v$  is the classical frequency of the oscillation given by

$$v = \frac{1}{2\pi} \sqrt{\frac{k}{m}}.$$

In terms of  $y$  and  $\alpha$ , Schrödinger's equation becomes

$$\frac{d^2\psi}{dy^2} + (\alpha - y^2) \psi = 0 \quad \dots(4)$$

To solve this equation, a solution of the form below can be tried:

$$\psi = f(y) e^{-y^2/2} \quad \dots(5)$$

Here,  $f(y)$  is a function of  $y$  that remains to be found.

By inserting the  $\psi$  of Eq. (5) in Eq. (4) we obtain

$$\frac{d^2f}{dy^2} - 2y \frac{df}{dy} + (\alpha - 1)f = 0. \quad \dots(6)$$

which is the differential equation that  $f$  obeys.

Writing  $(\alpha - 1) = 2n$ , Eq. (6) becomes,

$$\frac{d^2 f}{dy^2} - 2y \frac{df}{dy} + 2n f = 0. \quad \dots(7)$$

This is a standard mathematical equation known as *Hermite's equation*.

The solutions of Eq. (7) are called *Hermite's Polynomials*, given by

$$H_n(y) = f(y) = (-1)^n \exp y^2 \frac{d^n}{dy^n} [\exp(-y^2)] \quad \dots(8)$$

The eigen functions of harmonic oscillator, therefore, are the following:

$$\psi_n(y) = NH_n(y) \exp(-y^2/2) \quad \dots(9)$$

Here,  $N$  is a normalisation constant.

**Energy Levels.** The eigen values (permitted values of the total energy) are given by,

$$E_n = \left(n + \frac{1}{2}\right) \hbar v, \quad n = 0, 1, 2, 3, \dots \quad \dots(10)$$

The energy of a harmonic oscillator is thus quantised in steps of  $\hbar v$ . The energy levels here are evenly spaced (Fig. 8.15).

**Zero-point energy.** When  $n = 0$ ,  $E_0 = \frac{1}{2} \hbar v$ .

This is the lowest value of energy the oscillator can have. This value is called the *zero point energy*.

**Wave Functions.** Each wave function  $\psi_n$  consists of a polynomial  $H_n(y)$  (called a *Hermite polynomial*), the exponential factor  $e^{-y^2/2}$  and a numerical coefficient which is needed for  $\psi_n$  to meet the normalisation condition

$$\int_{-\infty}^{\infty} |\psi_n|^2 dy = 1 \quad n = 0, 1, 2, \dots \quad \dots(11)$$

The general formula for the  $n$ th wave function is

$$\psi_n = \left(\frac{2mv}{\hbar/2\pi}\right)^{1/4} (2^n n!)^{-1/2} H_n(y) e^{-y^2/2} \quad \dots(12)$$

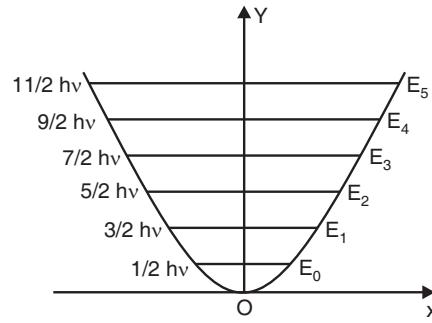


Fig. 8.15

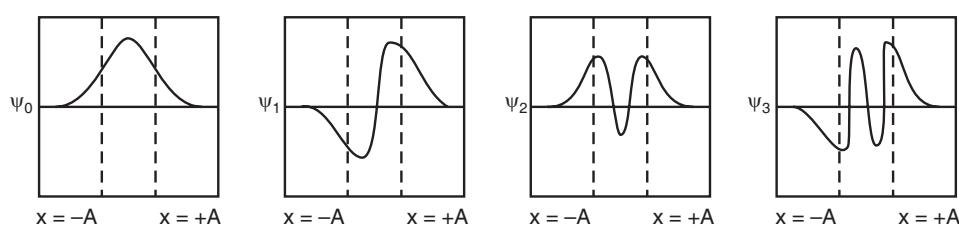


Fig. 8.16

The first four Hermite polynomials  $H_n(y)$  are listed in Table 8.1. The corresponding wave functions are plotted in Fig. 8.16. The vertical lines show the limits  $-A$  and  $+A$  between which a classical oscillator with the same energy would vibrate.

TABLE 8.1. Some Hermite polynomials

$n$	$H_n(y)$	$E_n$
0	1	$\frac{1}{2}hv$
1	$2y$	$\frac{3}{2}hv$
2	$4y^2 - 2$	$\frac{5}{2}hv$
3	$8y^3 - 12y$	$\frac{7}{2}hv$

## 8.10 THREE DIMENSIONAL HARMONIC OSCILLATOR—(SPHERICALLY SYMMETRIC CASE)

A three dimensional harmonic oscillator consists of a particle bound to origin by a force proportional to the displacement  $r$  from the fixed point, *i.e.*,

$$F = -kr.$$

Here,  $k$  is force constant.

We consider the special case in which frequency of oscillator along three axes  $X$ ,  $Y$  and  $Z$  are same, *i.e.*,

$$\begin{aligned} v_x &= v_y = v_z = v \text{ (say)} \\ \therefore k &= m\omega^2 = m(2\pi\nu)^2 = 4\pi^2 v^2 m. \end{aligned} \quad \dots(1)$$

Potential energy of oscillator assuming zero potential energy at  $r = 0$  is

$$V = - \int F dr = \int kr dr = \frac{1}{2}kr^2$$

But

$$r^2 = x^2 + y^2 + z^2$$

$\therefore$

$$V(r) = \frac{1}{2}kr^2 = \frac{1}{2}k(x^2 + y^2 + z^2).$$

The Schrödinger equation for this system is

$$\begin{aligned} \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + \frac{2m}{\hbar^2} \left[ E - \frac{1}{2}k(x^2 + y^2 + z^2) \right] \psi &= 0 \\ \text{or} \quad \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + 2\sqrt{\frac{m}{\hbar^2}} \sqrt{\frac{mk}{\hbar^2}} \\ &\times \left[ \frac{E}{\sqrt{k}} - \frac{1}{2}\sqrt{k}(x^2 + y^2 + z^2) \right] \psi = 0. \\ \text{or} \quad \frac{1}{\sqrt{\frac{mk}{\hbar^2}}} \left\{ \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right\} \\ &+ \left\{ 2\sqrt{\frac{m}{\hbar^2 k}} E - \sqrt{\frac{mk}{\hbar^2}} (x^2 + y^2 + z^2) \right\} \psi = 0 \end{aligned} \quad \dots(2)$$

Let  $\sqrt{\frac{mk}{\hbar^2}} = \alpha^2$  and  $2E \sqrt{\frac{m}{\hbar^2 k}} = \lambda$  ... (3)

$\therefore \frac{1}{\alpha^2} \left\{ \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right\} + [\lambda - \alpha^2 (x^2 + y^2 + z^2)] \psi = 0$  ... (4)

Let  $q_x = \alpha x, q_y = \alpha y$  and  $q_z = \alpha z$  ... (5)

$\therefore \frac{\partial^2 \psi}{\partial q_x^2} + \frac{\partial^2 \psi}{\partial q_y^2} + \frac{\partial^2 \psi}{\partial q_z^2} + [\lambda - (q_x^2 + q_y^2 + q_z^2)] \psi = 0$  ... (6)

This equation can be solved by the method of separation of variables.

Let  $\psi = Q_x(q_x) Q_y(q_y) Q_z(q_z)$ , ... (7)

Here,  $Q_x$  is function of  $q_x$ ,  $Q_y$  is function of  $q_y$  and  $Q_z$  is function of  $q_z$  only.

On substituting Eq. (7) in Eq. (6) and simplifying, we get

$$\frac{\partial^2 Q_x}{\partial q_x^2} + (\lambda_x - q_x^2) Q_x = 0 \quad \dots(8)$$

$$\frac{\partial^2 Q_y}{\partial q_y^2} + (\lambda_y - q_y^2) Q_y = 0 \quad \dots(9)$$

$$\frac{\partial^2 Q_z}{\partial q_z^2} + (\lambda_z - q_z^2) Q_z = 0 \quad \dots(10)$$

Here, the constants  $\lambda_x, \lambda_y$  and  $\lambda_z$  are related by

$$\lambda_x + \lambda_y + \lambda_z = \lambda \quad \dots(11)$$

Each of Eqs. (8), (9) and (10) is the differential equation for a one-dimensional harmonic oscillator (Section 8.9). By analogy, the normalised wave functions are given by

$$Q_x(q_x) = \frac{1}{[2^{n_x} n_x! \sqrt{\pi}]^{1/2}} e^{-q_x^2/2} H_{n_x}(q_x) \quad \dots(12)$$

$$Q_y(q_y) = \frac{1}{[2^{n_y} n_y! \sqrt{\pi}]^{1/2}} e^{-q_y^2/2} H_{n_y}(q_y) \quad \dots(13)$$

$$Q_z(q_z) = \frac{1}{[2^{n_z} n_z! \sqrt{\pi}]^{1/2}} e^{-q_z^2/2} H_{n_z}(q_z) \quad \dots(14)$$

$\lambda_x, \lambda_y$  and  $\lambda_z$  are restricted to the values

$$\lambda_x = 2n_x + 1, \lambda_y = 2n_y + 1 \text{ and } \lambda_z = 2n_z + 1 \quad \dots(15)$$

Here,  $n_x, n_y$  and  $n_z$  are non-negative integers.

The complete normalised wave function is given by

$$\begin{aligned} \psi = Q_x Q_y Q_z &= \frac{1}{[2^{(n_x+n_y+n_z)} n_x! n_y! n_z! \pi^{3/2}]^{1/2}} e^{-1/2(q_x^2+q_y^2+q_z^2)} \\ &\times H_{n_x}(q_x) H_{n_y}(q_y) H_{n_z}(q_z). \end{aligned} \quad \dots(16)$$

The eigen values are given by

$$E_n = \left( n_x + n_y + n_z + \frac{3}{2} \right) \hbar \omega = \left( n + \frac{3}{2} \right) \hbar \omega \quad \dots(17)$$

Here,  $n = n_x + n_y + n_z$  is called *total quantum number*.

Energy of this system depends only on the total quantum number,  $n$ .

The states of energy  $E_n$  are, therefore, *degenerate*. The degree of the degeneracy is  $\frac{1}{2}[(n+1)(n+2)]$ . Fig. 8.17 shows the energy levels and degree of degeneracy for three dimensional isotropic harmonic oscillator.

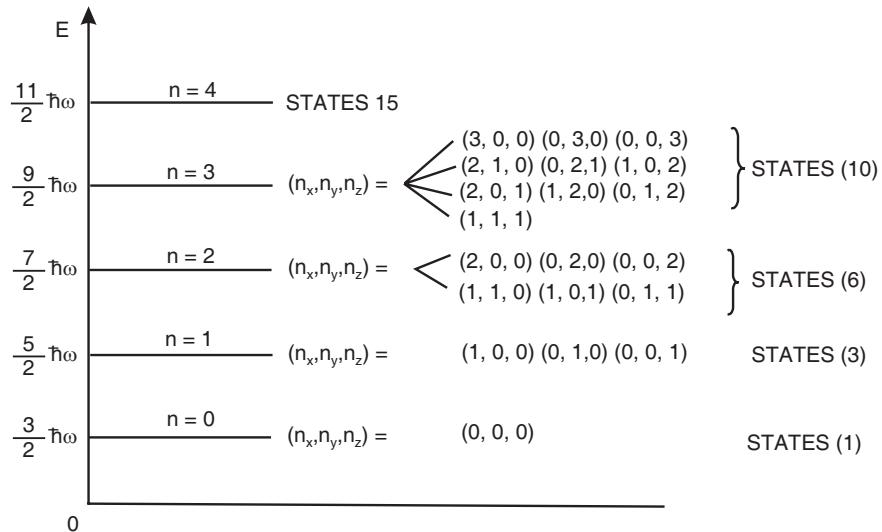


Fig. 8.17

### 8.11 THE RIGID ROTATOR

Consider two point masses  $M_1$  and  $M_2$  joined together by a rigid, massless rod [Fig. 8.18].

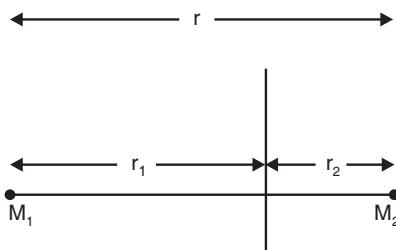


Fig. 8.18

The two masses are rotating like a dumbbell about an axis through their centre of mass  $C$  normal to the line joining the two masses. We have the relations,

$$M_1 r_1 = M_2 r_2, \text{ and } r_1 + r_2 = r.$$

Moment of inertia,  $I$ , of the system about the axis of rotation in terms of  $r$  is

$$I = M_1 r_1^2 + M_2 r_2^2 = \frac{M_1 M_2}{M_1 + M_2} r^2 = \mu r^2.$$

Here,  $\mu = M_1 M_2 / (M_1 + M_2)$  is the *reduced mass* of the system.

We can thus replace the rotating dumbbell by a single point mass  $\mu$  at a distance  $r$  from the axis of rotation.

The P.E. of the rotating dumbbell is zero, since the two point masses are rigidly joined together, with the distance between them remaining unaffected due to the rotation.

The angular momentum of this *simple rigid rotator*;  $p = I\omega$ .

Total energy of the rotator is

$$E = \frac{1}{2} I\omega^2 \quad (\because \text{P.E.} = 0)$$

Here,  $\omega$  is the uniform angular speed of rotation.

According to quantum theory, the angular momentum should be an integral multiple of  $\hbar$ .

$$\therefore p = I\omega = k\hbar$$

$$\text{and } E = E_k = \frac{1}{2} I\omega^2 = \frac{k^2\hbar^2}{2I}$$

Here,  $k$  is an integer.

**To derive the permitted energy values of the rigid rotator on the wave mechanical theory.**

The general Schrödinger equation in spherical polar coordinates is given by

$$\begin{aligned} & \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \psi}{\partial \theta} \right) \\ & + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} + \frac{2\mu}{\hbar^2} (E - V) \psi = 0 \end{aligned} \quad \dots(1)$$

As  $r$  is constant ( $r = 1$ ), the first factor  $\frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial \psi}{\partial r} \right) = 0$ ,  $V = 0$  and  $\mu$  may be replaced by  $I$ .

Schrödinger equation of rigid rotator reduces to

$$\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} + \frac{2I}{\hbar^2} E \psi = 0 \quad \dots(2)$$

The solution of Eq. (2) can be obtained by the method of separation of variables. Assume

$$\psi(\theta, \phi) = \psi_1(\theta) \psi_2(\phi) \quad \dots(3)$$

Substituting (3) in Eq. (2), and dividing both sides by  $\psi_1 \psi_2 / \sin^2 \theta$ , we have

$$\frac{\sin \theta}{\psi_1} \frac{d}{d\theta} \left( \sin \theta \frac{d\psi_1}{d\theta} \right) + \frac{2IE}{\hbar^2} \sin^2 \theta = - \frac{1}{\psi_2} \frac{d^2 \psi_2}{d\phi^2} \quad \dots(4)$$

Define a quantity  $m^2$  by the following equation:

$$\frac{1}{\psi_2} \frac{d^2 \psi_2}{d\phi^2} = -m^2 \text{ i.e., } \frac{d^2 \psi_2}{d\phi^2} + m^2 \psi_2 = 0 \quad \dots(5)$$

$$\therefore \frac{\sin \theta}{\psi_1} \frac{d}{d\theta} \left( \sin \theta \frac{d\psi_1}{d\theta} \right) + \frac{2IE}{\hbar^2} \sin^2 \theta = m^2$$

$$\text{or } \frac{1}{\sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\psi_1}{d\theta} \right) + \left( \frac{2IE}{\hbar^2} - \frac{m^2}{\sin^2 \theta} \right) \psi_1 = 0 \quad \dots(6)$$

The solution of Eq. (5) is

$$\psi_2 = \exp(\pm im\phi) \quad (m \text{ is any positive integer})$$

Putting  $\frac{2IE}{\hbar^2} = C$ , Eq. (6) becomes

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\psi_1}{d\theta} \right) + \left( C - \frac{m^2}{\sin^2 \theta} \right) \psi_1 = 0 \quad \dots(7)$$

This equation can be solved by substituting  $\cos \theta = x$

Equation (7) is then transformed to

$$\frac{d}{dx} \left[ (1-x^2) \frac{d\psi_1}{dx} \right] + \left[ C - \frac{m^2}{1-x^2} \right] \psi_1 = 0 \quad \dots(8)$$

Eq. (8) is called Legendre's equation.

Let us try the solution  $\psi_1 = (1-x^2)^{m/2} P(x)$  ...(9)

Here,  $P(x)$  is a function of  $x$  to be determined.

This type of solution enables us to put Eq. (8) in a standard form, namely

$$(1-x^2) \frac{d^2 P}{dx^2} - 2x(m+1) \frac{dP}{dx} + (C-m-m^2)P = 0.$$

We attempt a solution of this equation by expressing  $P$  as a power series in  $x$ , namely

$$P = \sum_{p=-\infty}^{+\infty} a_p x^p$$

Substituting this value of  $P$  in the differential equation, we get

$$\sum p(p-1) a_p x^{p-2} - \sum [p(p-1) + 2(m+1)p - (C-m-m^2)] a_p x^p = 0.$$

We find from the recurrence relation between the coefficients in the above equation that solutions in closed form are obtained when

$$C = (p+m)(p+m+1) = k(k+1).$$

Here, we have put  $k = p+m$ .

Substituting  $C = k(k+1)$  in  $C = \frac{2IE}{\hbar^2}$ , we get finally

$$E = k(k+1) \frac{\hbar^2}{2I} \quad \dots(10)$$

$k$  is called the *rotational quantum number*.

Eq. (10) gives allowed values for the energy (*i.e.*, eigen values) of a rigid rotator with free axis.

## 8.12 DERIVATION OF SCHRÖDINGER'S WAVE EQUATION IN MOMENTUM REPRESENTATION

In three dimensions, the time-dependent form of Schrödinger's equation is

$$\frac{-\hbar^2}{2m} \nabla^2 \psi + V\psi = i\hbar \frac{\partial \psi}{\partial t} \quad \dots(1)$$

The wave function for any plane wave is given by

$$\psi = \exp \left[ \frac{i}{\hbar} (\vec{p} \cdot \vec{r} - Et) \right]$$

Any superposition of such waves forms a wave packet.

The wave equation of the wave packet is

$$\psi = \frac{1}{(\sqrt{2\pi\hbar})^3} \int a(p) \exp\left[\frac{i}{\hbar}(\vec{p} \cdot \vec{r} - Et)\right] dp \quad \dots(2)$$

Differentiating Eq. (2) with respect to  $t$ , we get

$$\frac{\partial \psi}{\partial t} = \frac{1}{(\sqrt{2\pi\hbar})^3} \int \left(-\frac{iE}{\hbar}\right) a(p) \exp\left[\frac{i}{\hbar}(\vec{p} \cdot \vec{r} - Et)\right] dp \quad \dots(3)$$

Further  $\vec{p} \cdot \vec{r} = p_x x + p_y y + p_z z$

$$\therefore \frac{\partial \psi}{\partial x} = \frac{1}{(\sqrt{2\pi\hbar})^3} \int \frac{ip_x}{\hbar} a(p) \exp\left[\frac{i}{\hbar}(\vec{p} \cdot \vec{r} - Et)\right] dp$$

and

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{1}{(\sqrt{2\pi\hbar})^3} \int \left(-\frac{p_x^2}{\hbar^2}\right) a(p) \exp\left[\frac{i}{\hbar}(\vec{p} \cdot \vec{r} - Et)\right] dp$$

$$\frac{\partial^2 \psi}{\partial y^2} = \frac{1}{(\sqrt{2\pi\hbar})^3} \int \left(-\frac{p_y^2}{\hbar^2}\right) a(p) \exp\left[\frac{i}{\hbar}(\vec{p} \cdot \vec{r} - Et)\right] dp$$

and

$$\frac{\partial^2 \psi}{\partial z^2} = \frac{1}{(\sqrt{2\pi\hbar})^3} \int \left(-\frac{p_z^2}{\hbar^2}\right) a(p) \exp\left[\frac{i}{\hbar}(\vec{p} \cdot \vec{r} - Et)\right] dp$$

$\therefore$

$$\nabla^2 \psi = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2}$$

$$= \frac{1}{(\sqrt{2\pi\hbar})^3} \int \left(-\frac{p_x^2 + p_y^2 + p_z^2}{\hbar^2}\right) a(p) \exp\left[\frac{i}{\hbar}(\vec{p} \cdot \vec{r} - Et)\right] dp$$

$$= \frac{1}{(\sqrt{2\pi\hbar})^3} \int \left(-\frac{p^2}{\hbar^2}\right) a(p) \exp\left[\frac{i}{\hbar}(\vec{p} \cdot \vec{r} - Et)\right] dp \quad \dots(4)$$

Substituting the value of  $\nabla^2 \psi$  from Eq. (4) and  $\frac{\partial \psi}{\partial t}$  from Eq. (3) in Eq. (1), we get

$$\begin{aligned} -\frac{\hbar^2}{2m} \left(-\frac{p^2}{\hbar^2}\right) \frac{1}{(\sqrt{2\pi\hbar})^3} \int a(p) \exp\left[\frac{i}{\hbar}(\vec{p} \cdot \vec{r} - Et)\right] dp + V\psi \\ = \frac{1}{(\sqrt{2\pi\hbar})^3} i\hbar \left(-\frac{iE}{\hbar}\right) \int a(p) \exp\left[\frac{i}{\hbar}(\vec{p} \cdot \vec{r} - Et)\right] dp \end{aligned}$$

$$\text{or } \frac{1}{(\sqrt{2\pi\hbar})^3} \left(\frac{p^2}{2m} - E\right) \int a(p) \exp\left[\frac{i}{\hbar}(\vec{p} \cdot \vec{r} - Et)\right] dp + V\psi = 0 \quad \dots(5)$$

This is Schrödinger's three dimensional wave equation in momentum representation.

**One dimensional wave equation.** For a one dimensional case, Schrödinger's wave equation is

$$\frac{-\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} + V\psi = i\hbar \frac{\partial \psi}{\partial t} \quad \dots(6)$$

The equation for a wave packet in one dimension is

$$\psi = \frac{1}{\sqrt{2\pi\hbar}} \int a(p) \exp\left[\frac{i}{\hbar}(p_x x - Et)\right] dp$$

$$\therefore \frac{\partial\psi}{\partial t} = \frac{1}{\sqrt{2\pi\hbar}} \int \left(\frac{-iE}{\hbar}\right) a(p) \exp\left[\frac{i}{\hbar}(p_x x - Et)\right] dp \quad \dots(7)$$

and

$$\frac{\partial\psi}{\partial x} = \frac{1}{\sqrt{2\pi\hbar}} \int \left(\frac{i}{\hbar} p_x\right) a(p) \exp\left[\frac{i}{\hbar}(p_x x - Et)\right] dp$$

$$\frac{\partial^2\psi}{\partial x^2} = \frac{1}{(\sqrt{2\pi\hbar})^2} \int -\left(\frac{p_x}{\hbar}\right)^2 a(p) \exp\left[\frac{i}{\hbar}(p_x x - Et)\right] dp \quad \dots(8)$$

Substituting the value of  $\frac{\partial^2\psi}{\partial x^2}$  from Eq. (8), and  $\frac{\partial\psi}{\partial t}$  from Eq. (7) in Eq. (6), we get

$$\frac{1}{\sqrt{2\pi\hbar}} \left( \frac{p_x^2}{2m} - E \right) \int a(p) \exp\left[\frac{i}{\hbar}(p_x x - Et)\right] dp + V\psi = 0 \quad \dots(9)$$

This is Schrödinger's one dimensional wave equation in momentum representation.

### 8.12.1. Form of Wave Function in terms of Definite Momentum

Consider a free particle of total energy  $E$  and momentum  $p$  moving in the  $+x$  direction. The moving particle is associated with de Broglie waves which have wavelength  $\lambda$  and frequency  $v$ .

For de Broglie wave advancing in the positive  $x$  direction, the acceptable wave function is

$$\psi(x, t) = A e^{-i(\omega t - kx)} \quad \dots(1)$$

Here  $A$  is constant amplitude,  $\omega$  is constant angular frequency and  $k$  is propagation constant.

We have,

$$E = \hbar v = \hbar \omega \text{ and } p = \hbar k$$

$$\therefore \omega = \frac{E}{\hbar} \text{ and } k = \frac{p}{\hbar}$$

Substituting these values of  $\omega$  and  $k$  in Eq. (1),

$$\psi(x, t) = A e^{-i/\hbar(Et - px)} \quad \dots(2)$$

Eq. (2) describes the wave equivalent of an unrestricted particle of total energy  $E$  and momentum  $p$  moving in the  $+x$  direction.

In three dimensions, Eqs. (1) and (2) become

$$\psi(\mathbf{r}, t) = A e^{-i(\omega t - \mathbf{k.r})}$$

and

$$\psi(\mathbf{r}, t) = A e^{-i/\hbar(Et - \mathbf{p.r})}$$

Here,  $\mathbf{k}$  is the propagation vector.

$\mathbf{r}$  represents the position vector for the particle.

### EXERCISE

1. Arrive at the time-dependent Schrödinger's equation. (Madras, 2006)
2. Derive one dimensional time independent Schrödinger wave equation. (B.U. 2015)
3. What do you understand by Eigen value and Eigen function? (Madras, 2006)
4. Explain the physical interpretation of wave function. (B.U. 2013)
5. What is wave function? Discuss the properties of wave functions. (Garhwal 96)
6. Explain the normalization and orthogonality of wave function. (Agra 91, 92)
7. Calculate the values of energy of a particle in a one-dimensional box. Graphically indicate some of the wave functions for such a particle. (Madras, 2006)

8. Discuss quantum mechanically the problem of a particle in a finite square potential well. Draw diagrams showing the amplitude wave and probability density for the same. What will be the effect of increasing the width of the square well on energy levels? (Rajasthan 88)
9. Obtain the time independent Schrödinger's equation and solve it in the case of a particle in one dimensional rectangular potential well of finite depth. (Poorvanchal U. 2002)
10. (a) Write Schrödinger's equation for a particle in a rigid box and solve it. Find the eigen functions when a particle is kept in a rectangular box of dimensions  $l_x \times l_y \times l_z$ . Find the eigen values of momentum and energy.  
(b) Show that for a single value of energy or momentum, different quantum states are possible. Explain the term degeneracy. (G.N.D.U., 2001)
11. What is potential step? Find the reflection and transmission co-efficient for potential step of the form  $0 < E < V_0$ . Show that there is a finite probability of locating the particle in the region which is forbidden classically. What is the penetration distance? Give its relation to the mass of the incident particle. (G.N.D.U., 2007)
12. What is tunnel effect? Give any two examples. (Madras, 2006)
13. A particle of mass  $m$  and energy  $E < V_0$  travelling along  $x$ -axis has a potential barrier defined by
- $$V(x) = \begin{cases} 0 & \text{for } x < 0 \\ V_0 & \text{for } 0 < x < a \\ 0 & \text{for } x > a \end{cases}$$
- Derive the expression for reflection and transmission co-efficient of the particle. (G.N.D.U., 2006)
14. Formulate the Schrödinger's equation for a linear harmonic oscillator and solve it to obtain its eigen values and eigen function. (Madras, April 2006)
15. Find the solution of three dimensional harmonic oscillator in rectangular coordinates. (Garhwal, 1992)
16. Discuss the rigid rotator problem in wave mechanics and arrive at the eigen value of it. (Madras, Nov. 2006)

## QUANTUM THEORY OF THE HYDROGEN ATOM

**After reading this chapter, you should be able to**

- ◆ Apply the technique of separation of variables to solve hydrogen atom problem
- ◆ Explain the significance of the quantum numbers  $n$ ,  $l$ , and  $m_l$ .
- ◆ Discuss the difference between quantum mechanical model of the hydrogen atom and the Bohr model.

### 9.1 THE HYDROGEN ATOM

The hydrogen atom consists of a proton around which the electron revolves. The proton is assumed at rest at the origin of a rectangular coordinate system, and the electron is orbiting around the fixed proton at a radius  $r$  under the influence of the attractive coulomb field of the system. The potential energy function due to the coulomb field is  $V(r) = -e^2 / 4\pi\epsilon_0 r$ . From the viewpoint of quantum mechanics, the electron is represented by a wave system bounded by the potential well of the coulomb field. This circumstance results in a set of permitted standing wave systems, each corresponding to a particular possible value of the total energy.

#### Schroedinger's Equation for the Hydrogen Atom

Schrodinger's equation for the electron in three dimensions for the hydrogen atom is

$$\nabla^2 \psi + \frac{2m}{\hbar^2} (E - V) \psi = 0.$$

The geometrical symmetry of the problem suggests that it is more convenient to use the spherical polar coordinate system (Fig. 9.1).

The spherical polar coordinates are: (1)  $r$ , the radius vector, (2)  $\theta$ , the polar angle, and (3)  $\phi$ , the azimuth angle.

The rectangular coordinates are related to the spherical ones by

$$\begin{aligned} x &= r \sin \theta \cos \phi \\ y &= r \sin \theta \sin \phi \\ z &= r \cos \theta \end{aligned} \quad \dots(1)$$

The spherical polar coordinates,  $r, \theta, \phi$  of the point  $P$  have the following interpretations :

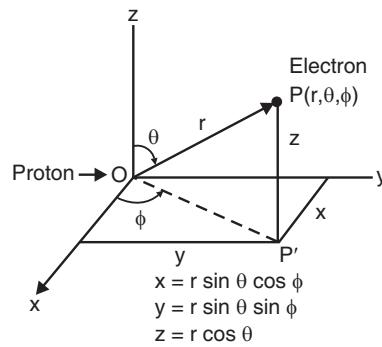


Fig. 9.1

$$r = \text{length of radius vector from origin } O \text{ to point } P = \sqrt{x^2 + y^2 + z^2}$$

$\theta = \text{angle between radius vector and } +z\text{-axis}$

$$= \text{polar or zenith angle} = \cos^{-1} z/r$$

$\phi = \text{angle between the projection } (OP') \text{ of the radius vector } OP \text{ on the } xy\text{-plane and the } +x\text{-axis}$

$$= \text{azimuth angle} = \tan^{-1} (y/x).$$

In spherical polar coordinates, Schrodinger's equation becomes

$$\begin{aligned} &\frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \psi}{\partial \theta} \right) \\ &+ \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} + \frac{2m}{\hbar^2} (E - V) \psi = 0 \end{aligned} \quad \dots(2)$$

Substituting  $V = -e^2 / 4\pi\epsilon_0 r$  and multiplying the entire equation by  $r^2 \sin^2 \theta$ , we obtain

$$\begin{aligned} &\sin^2 \theta \frac{\partial}{\partial r} \left( r^2 \frac{\partial \psi}{\partial r} \right) + \sin \theta \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{\partial^2 \psi}{\partial \phi^2} \\ &+ \frac{2mr^2 \sin^2 \theta}{\hbar^2} \left( E + \frac{e^2}{4\pi\epsilon_0 r} \right) \psi = 0 \end{aligned} \quad \dots(3)$$

Equation (3) is the partial differential equation for the wave function  $\psi$  of the electron in a hydrogen atom.

**Separation of Variables.** The wave function is now taken to depend upon  $r$ ,  $\theta$  and  $\phi$ . Here we look for solutions in which the wave function has the form of a product of three different functions.

$R(r)$ , which depends upon  $r$  alone;

$\Theta(\theta)$ , which depends upon  $\theta$  alone; and

$\Phi(\phi)$ , which depends upon  $\phi$  alone.

That is, we assume that

$$\psi(r, \theta, \phi) = R(r) \Theta(\theta) \Phi(\phi)$$

or

$$\psi = R \Theta \Phi \quad \dots(4)$$

We substitute this value of  $\psi$  in (3), divide throughout by  $R \Theta \Phi$ , and rearrange to get,

$$\begin{aligned} & \frac{\sin^2 \theta}{R} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + \frac{\sin \theta}{\Theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) \\ & + \frac{2mr^2 \sin^2 \theta}{\hbar^2} \left( E + \frac{e^2}{4\pi \epsilon_0 r} \right) = -\frac{1}{\Phi} \frac{d^2 \Phi}{d\phi^2} \end{aligned} \quad \dots(5)$$

Now the L.H.S. of this equation is a function of  $r$  and  $\theta$  alone while the R.H.S. is a function of  $\phi$  alone. Their equality implies that each side must equal a constant. It is convenient to call this constant  $m_l^2$ . Therefore, the differential equation for the function  $\Phi$  is

$$-\frac{1}{\Phi} \frac{d^2 \Phi}{d\phi^2} = m_l^2.$$

This gives the *azimuthal equation*  $\frac{d^2 \Phi}{d\phi^2} + m_l^2 \Phi = 0$   $\dots(6)$

When we substitute  $m_l^2$  for the R.H.S. of Eq. (5), divide the entire equation by  $\sin^2 \theta$ , and rearrange the various terms, we find that

$$\begin{aligned} & \frac{1}{R} \frac{d}{dr} \left[ r^2 \frac{dR}{dr} \right] + \frac{2mr^2}{\hbar^2} \left( E + \frac{e^2}{4\pi \epsilon_0 r} \right) \\ & = \frac{m_l^2}{\sin^2 \theta} - \frac{1}{\Theta \sin \theta} \frac{d}{d\theta} \left[ \sin \theta \frac{d\Theta}{d\theta} \right] \end{aligned}$$

Again we have an equation in which different variables appear on each side, requiring that both sides be equal to the same constant. We shall call this constant  $l(l+1)$ . Therefore, the equations for the function  $\Theta$  and  $R$  are

$$\frac{m_l^2}{\sin^2 \theta} - \frac{1}{\Theta \sin \theta} \frac{d}{d\theta} \left[ \sin \theta \frac{d\Theta}{d\theta} \right] = l(l+1) \quad \dots(7)$$

Eq. (7) is called the *polar equation*.

$$\frac{1}{R} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + \frac{2mr^2}{\hbar^2} \left( E + \frac{e^2}{4\pi \epsilon_0 r} \right) = l(l+1) \quad \dots(8)$$

Eq. (8) is called the *radial equation*.

The Schrodinger wave equation for the hydrogen atom has now been separated into three equations [(6), (7) and (8)] each dependent on only one of the coordinates.

**The Azimuthal equation.** The solution of the azimuthal wave equation [Eq. (6)] is

$$\Phi(\phi) = A \exp(im_l \phi) \quad \dots(9)$$

For this function to be single valued, its value must be the same for  $\phi = \phi$  and  $\phi = \phi + 2\pi$  or

$$A \exp(im_l \phi) = A \exp(im_l(\phi + 2\pi)).$$

For satisfying this condition,  $m_l$  is zero or a positive or negative integer. Therefore, the only permitted values of  $m_l$  are

$$m_l = 0, \pm 1, \pm 2, \pm 3, \dots \dots \dots \quad \dots(10)$$

The constant  $m_l$  is called the *magnetic quantum number* of the hydrogen atom.

**The polar equation.** The differential equation (7) has a rather complicated solution in terms of polynomials called the *associated Legendre polynomials* represented by  $P_l m_l(\cos \theta)$ . These polynomials exist only when the constant  $l$  is an integer equal to or greater than  $|m_l|$ . This requirement can be expressed as a condition on  $m_l$  in the form

$$m_l = 0, \pm 1, \pm 2, \pm 3, \dots \dots \pm l \quad \dots(11)$$

The constant  $l$  is called the *orbital quantum number*.

**The radial equation.** The solutions of the radial equation (8) are given by standard mathematical functions known as the *associated Laguerre polynomials*  $L_n, l(r)$ . Equation (8) can be solved only when  $E$  is positive or has one of the negative values  $E_n$  (signifying that the electron is bound to the atom) specified by

$$E_n = -\frac{me^4}{32\pi^2 \epsilon_0^2 \hbar^2} \left( \frac{1}{n^2} \right) = (-13.6) \frac{1}{n^2} eV \quad \dots(12)$$

This is in complete agreement with the predictions of the Bohr theory. Here  $n$  is again an integer whose value must be greater than or equal to  $(l+1)$  i.e.,  $n \geq (l+1)$ . We have now arrived at a system of three quantum numbers for the simple model of the hydrogen atom, which are related in the following way:

total quantum number	$n = 1, 2, 3, \dots \dots \dots$
orbital quantum number	$l = 0, 1, 2, \dots \dots (n-1)$
magnetic quantum number	$m_l = 0, \pm 1, \pm 2, \dots \dots \pm l$

### 9.1.1. The Solution of Polar Wave Equation

The polar equation is

$$\frac{m_l^2}{\sin^2 \theta} - \frac{1}{\Theta \sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) = l(l+1) \quad \dots(1)$$

$$\text{or } \frac{1}{\sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) + \left[ l(l+1) - \frac{m_l^2}{\sin^2 \theta} \right] \Theta = 0 \quad \dots(2)$$

$$\text{Put } x = \cos \theta \therefore \frac{dx}{d\theta} = -\sin \theta; \frac{d\Theta}{d\theta} = \frac{d\Theta}{dx} \frac{dx}{d\theta} = -\sin \theta \frac{d\Theta}{dx}$$

$$\therefore \frac{d}{d\theta} = -\sin \theta \frac{d}{dx}$$

Putting  $\frac{d}{d\theta}$  and  $\frac{d\Theta}{d\theta}$  in Eq. (2), we get

$$\frac{1}{\sin \theta} \times -\sin \theta \frac{d}{dx} \left( -\sin^2 \theta \frac{d\Theta}{dx} \right) + \left[ l(l+1) - \frac{m_l^2}{\sin^2 \theta} \right] \Theta = 0 \quad \dots(3)$$

$$\text{or } \frac{d}{dx} \left[ (1-x^2) \frac{d\Theta}{dx} \right] + \left[ l(l+1) - \frac{m_l^2}{(1-x^2)} \right] \Theta = 0 \quad \dots(4)$$

The differential equation (4) has a solution in terms of polynomials called the *associated Legendre polynomials* represented by  $P_l m_l(\cos \theta)$ .

$$\Theta = C P_{m_l} (\cos \theta) \text{ where } l = |m_l|, |m_l| + 1, |m_l| + 2, \dots \quad \dots(5)$$

The constant C can be calculated by normalising the function  $\Theta$ .

### 9.1.2. Solution of Radial Equation

The radial equation is

$$\frac{1}{R} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + \frac{2mr^2}{\hbar^2} \left( E + \frac{e^2}{4\pi \epsilon_0 r} \right) = l(l+1) \quad \dots(1)$$

$$\frac{d^2 R}{dr^2} + \frac{2}{r} \frac{dR}{dr} + \frac{2m}{\hbar^2} \left[ E + \frac{e^2}{4\pi \epsilon_0 r} \right] R = \frac{l(l+1)}{r^2} R$$

$$\text{or } \frac{d^2 R}{dr^2} + \frac{2}{r} \frac{dR}{dr} + \left[ \frac{2mE}{\hbar^2} + \frac{2me^2}{4\pi \epsilon_0 \hbar^2 r} - \frac{l(l+1)}{r^2} \right] R = 0 \quad \dots(2)$$

Putting  $\frac{-2mE}{\hbar^2} = \beta^2$  in Eq. (2), we get

[Note. Negative sign with energy  $E$  indicates that the electron is bound to the nucleus]

$$\frac{d^2 R}{dr^2} + \frac{2}{r} \frac{dR}{dr} + \left[ -\beta^2 + \frac{2me^2}{4\pi \epsilon_0 \hbar^2 r} - \frac{l(l+1)}{r^2} \right] R = 0$$

$$\text{or } \frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + \left[ -\beta^2 + \frac{2me^2}{4\pi \epsilon_0 \hbar^2 r} - \frac{l(l+1)}{r^2} \right] R = 0 \quad \dots(3)$$

Now put  $\rho = 2\beta r$  (where  $\rho$  is a dimension-less variable.) ...(4)

$$\therefore \frac{d\rho}{dr} = 2\beta$$

$$\text{Also } \frac{dR}{dr} = \frac{dR}{d\rho} \cdot \frac{d\rho}{dr} = 2\beta \frac{dR}{d\rho}$$

$$\text{or } r^2 \frac{dR}{dr} = 2\beta r^2 \frac{dR}{d\rho} = 2\beta \frac{\rho^2}{4\beta^2} \frac{dR}{d\rho} = \frac{\rho^2}{2\beta} \frac{dR}{d\rho}$$

Differentiating both sides of this equation w.r.t.  $r$ , we get

$$\frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) = \frac{d}{dr} \left( \frac{\rho^2}{2\beta} \frac{dR}{d\rho} \right) = \frac{d}{d\rho} \left( \frac{\rho^2}{2\beta} \frac{dR}{d\rho} \right) \frac{d\rho}{dr} = \frac{d}{d\rho} \left( \rho^2 \frac{dR}{d\rho} \right) \quad \dots(5)$$

Substituting the value  $r = \rho/2\beta$  from Eq. (4) and  $\frac{d}{dr} \left( r^2 \frac{dR}{dr} \right)$  from Eq. (5) in Eq. (3), we get

$$\frac{4\beta^2}{\rho^2} \frac{d}{d\rho} \left( \rho^2 \frac{dR}{d\rho} \right) + \left[ -\beta^2 + \frac{me^2 \beta}{\pi \epsilon_0 \hbar^2 \rho} - \frac{4l(l+1)\beta^2}{\rho^2} \right] R = 0$$

$$\text{or } \frac{d^2 R}{d\rho^2} + \frac{2}{\rho} \frac{dR}{d\rho} + \left[ -\frac{1}{4} + \frac{me^2}{4\pi \epsilon_0 \hbar^2 \rho \beta} - \frac{l(l+1)}{\rho^2} \right] R = 0$$

$$\text{or } \frac{d^2 R}{d\rho^2} + \frac{2}{\rho} \frac{dR}{d\rho} + \left[ \frac{n}{\rho} - \frac{l(l+1)}{\rho^2} - \frac{1}{4} \right] R = 0 \quad \dots(6)$$

where

$$n = \frac{me^2}{4\pi\epsilon_0\hbar^2\beta} \quad \dots(6)(a)$$

**Asymptotic solution.** For large  $\rho$ , the only terms that remain in Eq. (6) are

$$\frac{d^2R}{d\rho^2} - \frac{R}{4} = 0$$

The solution, which behaves properly at infinity, is  $R = Ae^{-\rho/2}$

With the asymptotic solution, the possible solution of Eq. (6) is

$$R = Ae^{-\rho/2} F(\rho) \quad \dots(7)$$

where  $F(\rho)$  is a polynomial of finite order in  $\rho$ .

Substitution of Eq. (7) into Eq. (6) gives

$$\frac{d^2F}{d\rho^2} + \left(\frac{2}{\rho} - 1\right) \frac{dF}{d\rho} + \left[\frac{n-1}{\rho} - \frac{l(l+1)}{\rho^2}\right] F = 0 \quad \dots(8)$$

In order that the solution of Eq. (8) remains finite for all values of  $\rho$ , the condition is

$$n-1 = l, l+1, l+2, l+3 \dots \text{etc.}$$

or

$$n = l+1, l+2, l+3 \dots \text{etc.}$$

$n$  is called the *principal quantum number*.

$$l = 0, 1, 2, \dots (n-1),$$

$$\text{i.e., } n = 1, 2, 3, 4, \dots$$

**The energy eigenvalues.** From Eq. (6) (a),  $\beta = \frac{me^2}{4\pi\epsilon_0\hbar^2} \cdot \frac{1}{n}$

$$\text{or } \beta^2 = \frac{m^2 e^4}{16\pi^2 \epsilon_0^2 \hbar^4} \cdot \frac{1}{n^2}$$

$$\text{But } \beta^2 = \frac{-2mE}{\hbar^2}$$

$$\therefore \frac{-2mE}{\hbar^2} = \frac{m^2 e^4}{16\pi^2 \epsilon_0^2 \hbar^4} \cdot \frac{1}{n^2}$$

$$E_n = -\frac{me^4}{32\pi^2 \epsilon_0^2 \hbar^2} \cdot \frac{1}{n^2} \quad \dots(9)$$

Eq. (9) is in agreement with the old quantum theory and with experiment.

**Radial wave function.** The radial wave function of the hydrogen atom is given by Eq. (7).

$$R = Ae^{-\rho/2} F(\rho)$$

The solution of Eq. (8) will depend upon two quantum numbers  $n$  and  $l$  and  $\rho = 2\beta r$ .

$$\therefore R_{nl}(\rho) = A_{nl} e^{-\rho/2} F_{nl}(\rho) = A_{nl} e^{-\beta r} F_{nl}(2\beta r) \quad \dots(10)$$

$$\text{Now } \beta = \frac{me^2}{4\pi\epsilon_0\hbar^2} \cdot \frac{1}{n} = \frac{1}{a_0 n}$$

Here,  $a_0 = \frac{4\pi\epsilon_0\hbar^2}{me^2}$  is the radius of the first Bohr orbit.

$$\therefore R_{nl} = A_{nl} e^{-r/a_0 n} F_{nl}\left(\frac{2r}{a_0 n}\right) \quad \dots(11)$$

It can be shown that the polynomial  $F_{nl}\left(\frac{2r}{a_0 n}\right)$  is the product of  $\left(\frac{2r}{a_0 n}\right)^l$  and the *associated Laguerre polynomial*  $L_{n+l}^{2l+1}\left(\frac{2r}{na_0}\right)$ . Thus

$$F_{nl}\left(\frac{2r}{a_0 n}\right) = \left(\frac{2r}{a_0 n}\right)^l L_{n+l}^{2l+1}\left(\frac{2r}{na_0}\right)$$

Thus the acceptable radial wave functions are

$$R_{nl} = A_{nl} e^{-r/na_0} \left(\frac{2r}{na_0}\right)^l L_{n+l}^{2l+1}\left(\frac{2r}{na_0}\right) \quad \dots(12)$$

The first three normalised radial wave functions are given below.

$$\begin{aligned} R_{10}(r) &= \frac{2}{a_0^{3/2}} e^{-r/a_0} \\ R_{20}(r) &= \frac{1}{(2a_0)^{3/2}} \left(2 - \frac{r}{a_0}\right) e^{-r/2a_0} \\ R_{21}(r) &= \frac{1}{(2a_0)^{3/2}} \frac{1}{\sqrt{3}} \frac{r}{a_0} e^{-r/2a_0} \end{aligned}$$

#### **Complete wave function for the hydrogen atom.**

The complete hydrogen atom wave functions are given by

$$\begin{aligned} \psi(r, \theta, \phi) &= R(r) \Theta(\theta) \Phi(\phi) \\ &= A_{nl} e^{-r/na_0} \left(\frac{2r}{na_0}\right)^l L_{n+l}^{2l+1}\left(\frac{2r}{na_0}\right) C_{lm_l} P_{lm_l}(\cos \theta) \frac{1}{\sqrt{2\pi}} \exp(im_l \phi) \\ \psi(r, \theta, \phi) &= N_{nlm_l} e^{-r/na_0} \left(\frac{2r}{na_0}\right)^l L_{n+l}^{2l+1}\left(\frac{2r}{na_0}\right) P_{lm_l}(\cos \theta) \exp(im_l \phi) \end{aligned}$$

Here,  $N_{nlm_l} = A_{nl} C_{lm_l} \frac{1}{\sqrt{2\pi}}$  is called the normalisation constant.

#### **Ground state wave function.**

$$\psi_{100} = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2}} \frac{2}{a_0^{3/2}} e^{-r/a_0} = \frac{1}{\sqrt{\pi}} \frac{2}{a_0^{3/2}} e^{-r/a_0}$$

#### **9.1.3. Expression for the Energy of the Electron of the Hydrogen Atom in the Ground State**

The ground state normalized wave function is

$$R(r) = \frac{2}{a_0^{3/2}} e^{-r/a_0} \quad \dots(1)$$

The radial equation is

$$\frac{1}{\psi_1} \frac{d}{dr} \left( r^2 \frac{d\psi_1}{dr} \right) + \frac{8\pi^2 m r^2}{h^2} \left( E + \frac{e^2}{4\pi\epsilon_0 r} \right) = l(l+1)$$

$$\text{or} \quad \frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{d\Psi_1}{dr} \right) + \left[ \frac{8\pi^2 m}{h^2} \left( E + \frac{e^2}{4\pi\epsilon_0 r} \right) - \frac{l(l+1)}{r^2} \right] \Psi_1 = 0 \quad \dots(2)$$

$$\frac{d\Psi_1}{dr} = - \left( \frac{2}{a_0^{5/2}} \right) e^{-r/a_0}$$

$$\frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{d\Psi_1}{dr} \right) = \left[ \frac{2}{a_0^{5/2}} - \frac{4}{ra_0^{5/2}} \right] e^{-r/a_0}$$

For the ground state,  $l = 0$  and  $E = E_1$ .

Substituting these values in Eq. (2), we get

$$\left[ \left( \frac{2}{a_0^{5/2}} + \frac{16\pi^2 m E_1}{h^2 a_0^{5/2}} \right) + \left( \frac{4\pi m e^2}{\epsilon_0 h^2 a_0^{5/2}} - \frac{4}{a_0^{5/2}} \right) r \right] e^{-r/a_0} = 0$$

Each parenthesis must equal 0 for the entire equation to equal 0.

$$\begin{aligned} \frac{4\pi m e^2}{\epsilon_0 h^2 a_0^{5/2}} - \frac{4}{a_0^{5/2}} &= 0 \\ \therefore a_0 &= \frac{\epsilon_0 h^2}{\pi m e^2} \\ \frac{2}{a_0^{5/2}} + \frac{16\pi^2 m E_1}{h^2 a_0^{5/2}} &= 0 \end{aligned} \quad \dots(3)$$

$$\text{or} \quad E_1 = - \frac{h^2}{8\pi^2 m a_0^2} = - \frac{m e^4}{8\epsilon_0^2 h^2} \quad \dots(4)$$

This is the expression for the lowest energy state for the electron of the hydrogen atom.

#### 9.1.4. Normalized Wave Functions of the Hydrogen Atom

The following Table gives the normalized wave functions of the hydrogen atom for  $n = 1$ .

$n$	$l$	$m_l$	$R(r)$	$\Theta(\theta)$	$\Phi(\phi)$	$\Psi(r, \theta, \phi)$
1	0	0	$\frac{2}{a_0^{5/2}} e^{-r/a_0}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2\pi}}$	$\frac{1}{\sqrt{\pi}} \frac{1}{a_0^{5/2}} e^{-r/a_0}$

The ground state wave function is given by

$$\Psi_{1,0,0}(r, \theta, \phi) = \left( \frac{1}{\pi a_0^3} \right)^{1/2} \exp\left(-\frac{r}{a_0}\right) \quad \dots(1)$$

where  $a_0$  is the Bohr radius. This wave function depends only on  $r$  and not on  $\theta$  and  $\phi$ .

The four wave functions corresponding to the first excited state, as specified by the  $n, l, m_l$  values, are

$$\Psi_{2,0,0}(r, \theta, \phi) = \frac{1}{4(2\pi a_0^3)^{1/2}} \left( 2 - \frac{r}{a_0} \right) \exp\left(-\frac{r}{2a_0}\right) \quad \dots(2)$$

$$\Psi_{2,1,0}(r, \theta, \phi) = \frac{1}{8(\pi a_0^3)^{1/2}} \frac{r}{a_0} \exp\left(-\frac{r}{2a_0}\right) \sin \theta \exp(i\phi) \quad \dots(3)$$

$$\Psi_{2,1,1}(r, \theta, \phi) = \frac{1}{4(2\pi a_0^3)^{1/2}} \frac{r}{a_0} \exp\left(-\frac{r}{2a_0}\right) \cos \theta \quad \dots(4)$$

$$\Psi_{2,1,-1}(r, \theta, \phi) = \frac{1}{8(\pi a_0^3)^{1/2}} \frac{r}{a_0} \exp\left(-\frac{r}{2a_0}\right) \sin \theta \exp(-i\phi) \quad \dots(5)$$

We note that  $\Psi_{2,0,0}(r, \theta, \phi)$  is spherically symmetric, that is, depends only on  $r$ . These wave functions are normalized.

### 9.1.5 Significance of Various Quantum Numbers

The wave functions of the hydrogen atom are characterized by three quantum numbers. Let us consider the quantum mechanical interpretation of  $n$ ,  $l$ , and  $m_l$ .

#### 1. The principal quantum number ( $n$ )

The radial wave equation for the hydrogen atom can be solved only when  $E$  has a continuous range of positive values, or a discrete set of *negative* values (which correspond to *bound states*), given by

$$E = -\frac{\mu e^4}{8\epsilon_0^2 h^2} \left( \frac{1}{n^2} \right), \quad n = 1, 2, 3, \dots$$

Thus  $n$  describes the quantisation of electron energy in the hydrogen atom.

#### 2. The orbital quantum number ( $l$ )

The total energy  $E$  of the electron includes the kinetic energy because of the radial motion, kinetic energy because of the orbital motion and also the potential energy.

The total energy of the electron is

$$E = KE_{radial} + KE_{orbital} - \frac{e^2}{4\pi\epsilon_0 r} \quad \dots(1)$$

The radial equation is

$$\frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{d\Psi_1}{dr} \right) + \left[ \frac{8\pi^2 m}{h^2} \left( E + \frac{e^2}{4\pi\epsilon_0 r} \right) - \frac{l(l+1)}{r^2} \right] \Psi_1 = 0 \quad \dots(2)$$

$$\frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{d\Psi_1}{dr} \right) + \frac{8\pi^2 m}{h^2} \left[ K_{radial} + K_{orbital} - \frac{h^2 l(l+1)}{8\pi^2 m r^2} \right] \Psi_1 = 0 \quad \dots(3)$$

The equation is brought completely dependent on the radial motion only if we put

$$\begin{aligned} KE_{orbital} - \frac{h^2 l(l+1)}{8\pi^2 m r^2} &= 0 \\ KE_{orbital} &= \frac{h^2 l(l+1)}{8\pi^2 m r^2} = \frac{1}{2} m v_{orbital}^2 \end{aligned}$$

Here,  $v_{orbital}$  is the orbital velocity of the electron.

$$\therefore l(l+1) \left( \frac{h}{2\pi} \right)^2 = (mv_{orbital} r)^2 = L^2$$

Here,  $L = mv_{orbital} r$  = angular momentum of the electron in its orbital motion.

This means that the electron in the hydrogen atom has an angular momentum given by

$$L = \sqrt{l(l+1)} \frac{h}{2\pi} \quad \dots(4)$$

Here,  $l = 0, 1, 2, \dots (n - 1)$ . Thus, like energy, the angular momentum is also quantised. This quantisation is described by the orbital quantum number  $l$ . The orbital quantum number  $l$  determines the magnitude  $L$  of the electron's angular momentum  $\mathbf{L}$ .

### 3. Magnetic orbital quantum number ( $m_l$ )

An electron revolving about a nucleus (Fig. 9.2) is a current loop and has a magnetic field. Hence an atomic electron interacts with an external magnetic field  $\mathbf{B}$ . The electron orbital angular momentum vector is  $\mathbf{L}$ . The magnetic quantum number  $m_l$  specifies the direction of  $\mathbf{L}$  by determining the component of  $\mathbf{L}$  in the field direction. The  $z$ -axis is taken parallel to  $\mathbf{B}$ .  $\theta$  is the angle between  $\mathbf{L}$  and the  $z$ -axis.

The component of  $\mathbf{L}$  in the  $z$ -direction is

$$L_z = m_l \hbar .$$

Since, for a given  $l$ , there are  $(2l + 1)$  possible values of  $m_l (= 0, \pm 1, \pm 2, \dots, \pm l)$ , the angle  $\theta$  can assume  $(2l + 1)$  discrete values. In other words, the angular momentum vector  $\mathbf{L}$  can have  $(2l + 1)$  discrete orientations with respect to the magnetic field. This quantisation of the orientation of atoms in space is known as 'space quantisation'.

The space quantisation of the orbital angular momentum vector  $\mathbf{L}$  corresponding to  $l = 2$  is shown in Fig. 9.3.

The uncertainty principle prohibits the angular-momentum vector  $\mathbf{L}$  from having a definite direction in space. The simultaneous quantization of  $|\mathbf{L}|$  and  $L_z$  can be described graphically by means of the vector model. According to the vector model, the angular-momentum vector  $\mathbf{L}$  precesses around the  $z$ -axis, tracing out a cone. The angle between the vector  $\mathbf{L}$  and the  $z$ -axis remains constant.

From Fig. 9.4,

$$\cos \theta = \frac{L_z}{|\mathbf{L}|} = \frac{m_l \hbar}{\sqrt{l(l+1)\hbar^2}} = \frac{m_l}{\sqrt{l(l+1)}} .$$

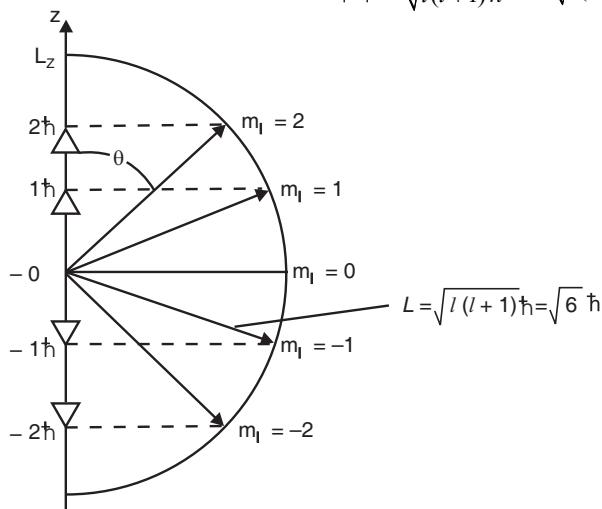


Fig. 9.3

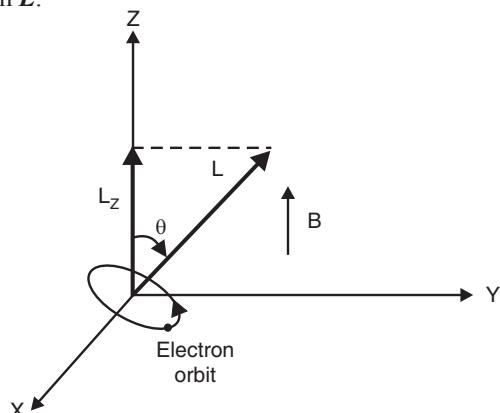


Fig. 9.2

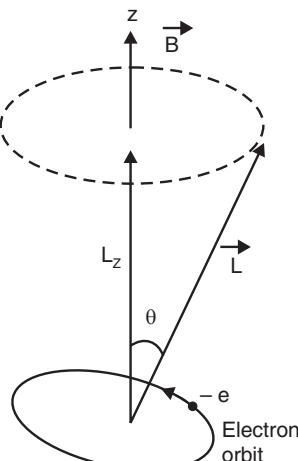


Fig. 9.4

The smallest possible angle occurs for  $m_l = l$ , which gives  $\cos \theta = l/\sqrt{l(l+1)}$ ; the largest angle occurs for  $m_l = 0$ , which gives  $\cos \theta = 0$  and  $\theta = 90^\circ$ . The average values of  $L_x$  and  $L_y$  are zero, but the instantaneous values of  $L_x$  and  $L_y$  oscillate around zero. Because of the precession, the instantaneous values of  $L_x$  and  $L_y$  change quickly—successive measurements give different results,

and values of  $L_x$  and  $L_y$  are not well defined. Only the average values of  $L_x$  and  $L_y$  are well defined—both averages are zero. But  $L_z$  always has the specific value  $m_p\hbar$ .

## 9.2 ELECTRON PROBABILITY DENSITY

The probability density  $P(r, \theta, \phi)$  of the electron at the point  $(r, \theta, \phi)$  is proportional to  $|\psi|^2$ .

The probability of finding the electron in a small volume  $dV$  is

$$P(r, \theta, \phi) dV = |\psi(r, \theta, \phi)|^2 dV$$

In spherical polar coordinates

$$dV = r^2 \sin \theta dr d\theta d\phi$$

Thus, we have

$$P(r, \theta, \phi) dV = |\psi(r, \theta, \phi)|^2 r^2 \sin \theta dr d\theta d\phi \quad \dots(1)$$

The probability  $P(r) dr$  of finding the electron between  $r$  and  $(r + dr)$ , irrespective of the coordinates  $\theta, \phi$ , is obtained by integrating Eq. (1) over the coordinates  $\theta, \phi$ . Thus,

$$P(r) dr = \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin \theta |\psi(r, \theta, \phi)|^2 r^2 dr \quad \dots(2)$$

For the ground state of the hydrogen atom, we have

$$|\psi(r, \theta, \phi)|^2 = \frac{1}{\pi a_0^3} \exp\left(-\frac{2r}{a_0}\right) \quad \dots(3)$$

Substituting this in Eq. (2), we get

$$P(r) dr = \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin \theta \frac{1}{\pi a_0^3} \exp\left(-\frac{2r}{a_0}\right) r^2 dr \quad \dots(4)$$

$$\begin{aligned} &= 2\pi \int_0^\pi \sin \theta d\theta \frac{1}{\pi a_0^3} \exp\left(-\frac{2r}{a_0}\right) r^2 dr \\ &= [2\pi(-\cos \theta)] \int_0^\pi \frac{1}{\pi a_0^3} \exp\left(-\frac{2r}{a_0}\right) r^2 dr \\ &= \left[ \frac{1}{\pi a_0^3} \exp\left(-\frac{2r}{a_0}\right) \right] 4\pi r^2 dr \end{aligned} \quad \dots(5)$$

Here,  $4\pi r^2 dr$  is the volume of a spherical shell of radius  $r$  and thickness  $dr$ . Hence, Eq. (5) is the probability density times the volume element, and can be written as

$$P(r) dr = |\psi_{1,0,0}|^2 4\pi r^2 dr. \quad \dots(6)$$

In Fig. 9.5,  $P(r) dr$  has been plotted against  $r$  for two states of hydrogen atom corresponding to  $n = 1$  and  $n = 2$  ( $l = 0$  in both cases). It is seen that the principal maxima of the curves closely agree with the radii of the Bohr orbits ( $r = a_0, 4a_0, \dots$ ). Thus, whereas

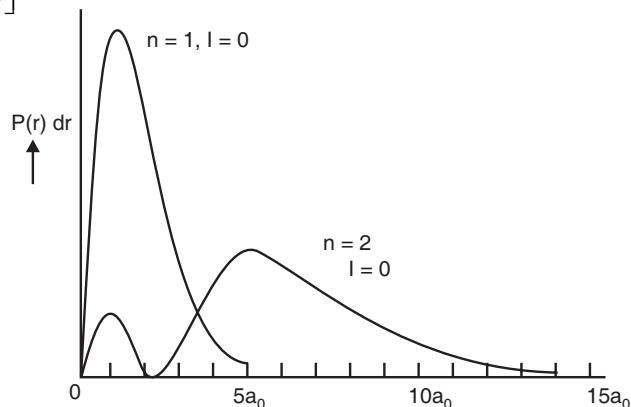


Fig. 9.5

the electron is most likely to be found at the locations of Bohr orbits, it has a finite probability of being found elsewhere.

The radial probability density  $P(r)$ , from Eq. (5), is

$$P(r) = \frac{4}{a_0^3} \exp\left(-\frac{2r}{a_0}\right) r^2 \quad \dots(7)$$

A plot of radial distribution function  $P(r)$  against  $r$  is shown in Fig. 9.6

The most probable distance of the electron from the nucleus, i.e., the value of  $r$  at which  $P(r)$  is maximum, is obtained by equating  $\frac{dP(r)}{dr}$  to zero. That is

$$\frac{d}{dr} \left[ \frac{4}{a_0^3} (e^{-2r/a_0}) r^2 \right] = 0$$

$$\text{or } \frac{4}{a_0^3} \left[ e^{-2r/a_0} (2r) + r^2 e^{-2r/a_0} (-2/a_0) \right] = 0$$

$$\text{or } 2r + r^2 (-2/a_0) = 0$$

$$\text{or } r = a_0.$$

Thus the most probable distance of the electron from the nucleus in the normal state of hydrogen atom is equal to the Bohr's radius.

**EXAMPLE 1.** Calculate the mean value of  $r$  for an electron in the ground state of the hydrogen atom.

**SOL.** The probability for finding the electron in a small volume  $dV$  is  $|\psi_{100}|^2 dV$ . Consider a volume in the shape of a thin spherical shell of radius  $r$  and thickness  $dr$  (Fig. 9.7). For this volume,  $dV = 4\pi r^2 dr$ . Hence

$$\begin{pmatrix} \text{probability for} \\ \text{finding electron} \\ \text{in interval from} \\ r \text{ to } r + dr \end{pmatrix} = 4\pi |\psi_{100}|^2 r^2 dr \quad \dots(1)$$

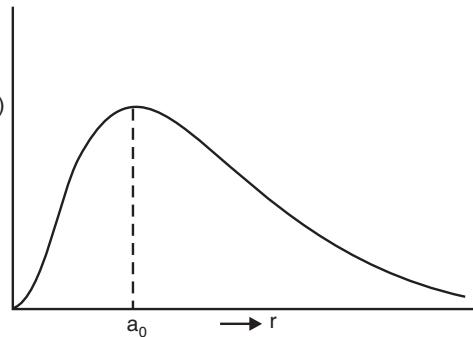


Fig. 9.6

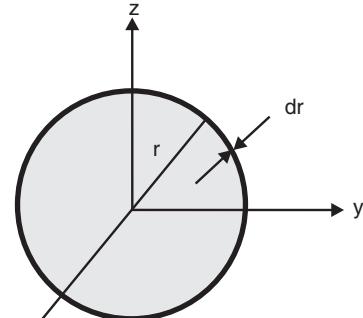


Fig. 9.7

This can also be interpreted as the probability for finding the value of  $r$  in the interval  $dr$ . To obtain the mean value of  $r$ , we must multiply this probability by  $r$  and integrate over all values of  $r$ :

$$\langle r \rangle = \int_0^\infty 4\pi |\psi_{100}|^2 r^3 dr \quad \dots(2)$$

$$= \int_0^\infty \frac{4}{a_0^3} e^{-2r/a_0} r^3 dr \quad \dots(3)$$

We have the standard integral,

$$\int_0^\infty e^{-\beta r} r^n dr = \frac{n!}{\beta^{n+1}} \text{ for } n = 0, 1, 2, \dots \quad \dots(4)$$

$$\therefore \langle r \rangle = \frac{4}{a_0^3} \frac{3!}{(2/a_0)^4} = \frac{3}{2} a_0$$

Thus the average value of  $r$  for a 1s electron is  $1.5 a_0$ .

**EXAMPLE 2.** Calculate the average value of  $1/r$  for a 1s electron in the hydrogen atom.

**SOL.** The expectation value of  $1/r$  is

$$\langle \frac{1}{r} \rangle = \int_0^\infty \left( \frac{1}{r} \right) |\psi|^2 dV \quad \dots(1)$$

The wave function of a 1s electron is

$$\begin{aligned} \psi_{1,0,0} &= \frac{1}{\sqrt{(\pi a_0^3)}} e^{-r/a_0} \\ \therefore |\psi|^2 &= \frac{1}{\pi a_0^3} e^{-2r/a_0} \\ dV &= r^2 \sin \theta dr d\theta d\phi. \end{aligned}$$

Substituting these values in Eq. (1),

$$\begin{aligned} \langle \frac{1}{r} \rangle &= \frac{1}{(\pi a_0^3)} \int_0^\infty r (e^{-2r/a_0}) dr \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi \\ &= \frac{1}{\pi a_0^3} \left( \frac{a_0^2}{4} \right) (2)(2\pi) \\ \therefore \langle \frac{1}{r} \rangle &= \frac{1}{a_0} \end{aligned}$$

### 9.3 RADIATIVE TRANSITIONS

We start with the assumption that an atomic electron whose average position relative to the nucleus of the atom remains constant in time does not radiate. But if this position oscillates with time, electromagnetic waves are emitted. The frequency of these waves is the same as that of the oscillation.

The time-dependent wave function  $\psi$  of an electron is the product of a time-independent wave function  $\psi_0$  and a time-dependent function  $e^{-(iE/\hbar)t}$ . The frequency of the time-varying function is  $v = E/\hbar$ .

$$\therefore \psi = \psi_0 e^{-(iE/\hbar)t}. \quad \dots(1)$$

The expectation value  $\langle x \rangle$  of the position of the electron is

$$\langle x \rangle = \int_{-\infty}^{\infty} x \psi^* \psi dx = \int_{-\infty}^{\infty} x \psi_0 * \psi_0 dx \quad \dots(2)$$

The expectation value  $\langle x \rangle$  is constant in time since  $\psi_0$  and  $\psi_0^*$  are functions of position only. No radiation results in this case when the atom is in the specific quantum state  $\psi$ , since the electron does not oscillate.

Thus, Bohr's hypothesis of non-radiating energy levels is automatically realized in the behaviour of the stationary states.

Let  $\psi_i$  and  $\psi_f$  be two states in which the atomic electron can exist. Let  $E_i$  and  $E_f$  be the respective energies of these states, with  $E_i > E_f$ . Let the electron be in the state  $\psi_f$  initially when it gets excited to the state  $\psi_i$  from where it then undergoes a transition to the lower state  $\psi_f$ . The wave function  $\psi$  of an electron capable of existing in both states  $\psi_i$  and  $\psi_f$  may be written as

$$\psi = a \psi_f + b \psi_i$$

Here  $a^*a$  and  $b^*b$  are the relative probabilities of the electron being in the states  $\psi_f$  and  $\psi_i$  respectively. Just before the excitation and immediately after the transition,  $a = 1$ ,  $b = 0$ . When the electron is in the excited state,  $a = 0$  and  $b = 1$ . In between,  $a$  and  $b$  have nonvanishing values. Always,  $a^*a + b^*b = 1$ .

Then,  $\langle x \rangle$  is given by

$$\begin{aligned} \langle x \rangle &= \int_{-\infty}^{\infty} x(a^* \psi_f^* + b^* \psi_i^*)(a\psi_f + b\psi_i) dx \\ &= \int_{-\infty}^{\infty} [x(a^2 \psi_f^* \psi_f + b^* a \psi_i^* \psi_f + a^* b \psi_f^* \psi_i + b^2 \psi_i^* \psi_i)] dx \end{aligned} \quad \dots(3)$$

Here,  $a^*a = a^2$  and  $b^*b = b^2$ .

The integrals involving the first and the last terms in parenthesis are constant in time. The integrals involving the second and third terms are periodic in time with the frequency  $(E_i - E_f)/\hbar$ . These result in electromagnetic radiation of the same frequency when the electron undergoes a transition between the states specified by  $\psi_i$  and  $\psi_f$ .

Let us expand Eq. (3) using Eq. (1).

$$\begin{aligned} \langle x \rangle &= a^2 \int_{-\infty}^{\infty} x \psi_{of}^* \psi_{of} dx + b^* a \int_{-\infty}^{\infty} x \psi_{oi}^* e^{+(iE_i/\hbar)t} \psi_{of} e^{-(iE_f/\hbar)t} dx \\ &\quad + a^* b \int_{-\infty}^{\infty} x \psi_{of}^* e^{+(iE_f/\hbar)t} \psi_{oi} e^{-(iE_i/\hbar)t} dx \\ &\quad + b^2 \int_{-\infty}^{\infty} x \psi_{oi}^* \psi_{oi} dx \end{aligned} \quad \dots(4)$$

Here,  $\psi_{of}^* \psi_{oi} = \psi_{oi}^* \psi_{of}$  and  $a^*b = b^*a$ .

The integrals involving the second and third terms are the time-varying terms. They can be combined into the single term

$$\begin{aligned} a^* b \int_{-\infty}^{\infty} x \psi_{of}^* \psi_{oi} \left[ e^{(i/\hbar)(E_i - E_f)t} + e^{-(i/\hbar)(E_i - E_f)t} \right] dx \\ = 2a^* b \cos\left(\frac{E_i - E_f}{\hbar}\right) t \int_{-\infty}^{\infty} x \psi_{of}^* \psi_{oi} dx \end{aligned} \quad \dots(5)$$

Eq. (5) contains the time-varying factor

$$\cos\left(\frac{E_i - E_f}{\hbar}\right) t = \cos 2\pi\nu t.$$

The electron's position therefore oscillates sinusoidally at the frequency

$$\nu = \frac{E_i - E_f}{h} \quad \dots(6)$$

Hence electromagnetic waves are emitted and the frequency of these waves is the same as that of the oscillation. Thus Bohr's postulate  $\nu = (E_i - E_f)/h$  follows naturally from the quantum theory of the atom.

#### 9.4 SELECTION RULES

The integral

$$\int_{-\infty}^{\infty} x \psi_{of}^* \psi_{oi} dx \quad \dots(1)$$

carries all the essential quantum mechanical information about the two atomic states involved in the transition. The intensity of the radiation is proportional to the integral. Hence the general condition necessary for an atom in an excited state to radiate is that the value of the integral is not zero. Transitions for which this integral is finite are called *allowed transitions*. Transitions for which this integral is zero are called *forbidden transitions*.

Consider the transition between the initial and final atomic states of the hydrogen atom, labeled respectively by the quantum numbers  $(n', l', m'')$  and  $(n, l, m_l)$ . Let  $u$  represent either the  $x$ ,  $y$  or  $z$  coordinate. Then the condition for an allowed transition is

$$\int_{-\infty}^{\infty} u \psi_{n,l,m_l}^* \psi_{n',l',m''} \neq 0 \quad \dots(2)$$

The wave functions  $\Psi_{n, l, m_l}$  for the hydrogen atom are known. Hence Eq. (2) can be evaluated for  $u = x$ ,  $u = y$ , and  $u = z$  for all pairs of states differing in one or more quantum numbers. The selection rules for allowed transitions are

$$\Delta l = \pm 1 \quad \dots(3)$$

$$\Delta m_l = 0, \pm 1 \quad \dots(4)$$

The change in  $n$  is not restricted.

Fig. 9.8 shows Lyman transitions allowed by the  $\Delta l = \pm 1$  selection rule. All the wavelengths of the Lyman series arise from transitions of the type  $np \rightarrow 1s$ .

Fig. 9.9 shows that all the wavelengths of the Balmer series are due to transitions connecting  $l' = 0$  to  $l = 1$  or  $l' = 1$  to  $l = 0$  or  $l' = 2$  to  $l = 1$ .

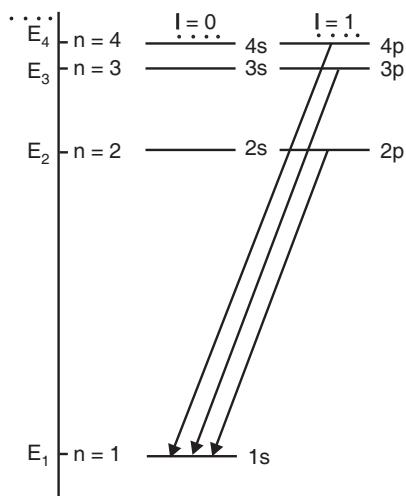


Fig. 9.8

The wave functions  $\Psi_{n, l, m_l}$  for the hydrogen atom are

known. Hence Eq. (2) can be evaluated for  $u = x$ ,  $u = y$ , and  $u = z$  for all pairs of states differing in one or more quantum numbers. The selection rules for allowed transitions are

$$\Delta l = \pm 1 \quad \dots(3)$$

$$\Delta m_l = 0, \pm 1 \quad \dots(4)$$

The change in  $n$  is not restricted.

Fig. 9.8 shows Lyman transitions allowed by the  $\Delta l = \pm 1$  selection rule. All the wavelengths of the Lyman series arise from transitions of the type  $np \rightarrow 1s$ .

Fig. 9.9 shows that all the wavelengths of the Balmer series are due to transitions connecting  $l' = 0$  to  $l = 1$  or  $l' = 1$  to  $l = 0$  or  $l' = 2$  to  $l = 1$ .

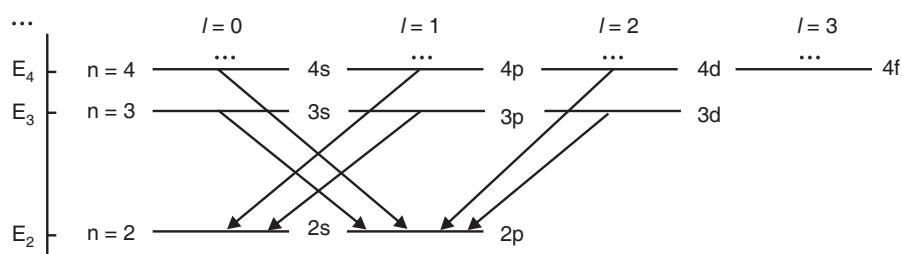


Fig. 9.9

**EXERCISE**

1. Solve the Schrodinger's equation for the hydrogen atom problem. *(Madras, 2006)*
2. Solve the azimuthal wave equation of hydrogen atom. *(B.U. 2012)*
3. Discuss the polar wave equation and its solution of hydrogen atom. *(B.U. 2013)*
4. Solve the radial part of the Schrodinger equation for the hydrogen atom. Obtain the energy eigen values and radial wave function. *(Garhwal 1991, Poorvanchal 2002, P.U. 2008)*
5. Explain the ground state energy of hydrogen atom. *(B.U. 2015)*
6. Obtain ground state wave functions for hydrogen atom using Schroedinger equation. Also calculate the most probable distance of electron from nucleus. *(Meerut 1991)*
7. State Schroedinger's equation for hydrogen atom in spherical co-ordinates. Explain the significance of various quantum numbers defining a state of the system. *(Vikram 1987)*

## OPERATOR FORMALISM OF QUANTUM MECHANICS

**After reading this chapter, you should be able to**

- ◆ Explain the formal structure of quantum mechanics as a set of five postulates
- ◆ List in general properties of operators and their eigenvalues and eigenfunctions
- ◆ Utilize the commutation relations of operators to determine whether or not two physical properties can be simultaneously measured
- ◆ Prove Ehrenfest's theorem and discuss how it is related to the correspondence principle
- ◆ Explain the concept of wave packet.

## 10.1 POSTULATES OF QUANTUM MECHANICS

### Postulate 1 – The State

*The state of a system is described by a wave function  $\psi(\mathbf{r}, t)$ .*

**Explanation.** For a single particle, the *wave-function* is a function only of position  $\mathbf{r}$  and time  $t$ , and is written  $\psi(\mathbf{r}, t)$ . The wave function  $\psi(\mathbf{r}, t)$  gives the complete knowledge of the behaviour of the particle.

Similarly,  $\psi(\mathbf{r})$  gives the stationary state which is independent of time.

- The wave function is a complex function of co-ordinates and time.  
The probability of finding the particle in the space within  $r$  and  $r + dr$  is  $\psi^* \psi dr$ .
- In order to conform to the physically realisable states, it must be well-behaved, *i.e.*, it must be singled-valued, continuous and finite.

### Postulate 2 – Operators

*Every physical observable is associated with a linear Hermitian operator.*

**Explanation.** The result of measurements is a number. The eigenvalues of a *Hermitian operator* are real, which justifies their use.

- The *linearity* condition stems from the superposition principle.

### Statement of Superposition Principle

If any system (a particle or an assembly of particles) can exist in a state represented by the wave function  $\Psi_1$  and also in a state represented by the wave function  $\Psi_2$ , then it can be also in a state represented by a wave function  $\Psi$  such that

$$\Psi = c_1 \Psi_1 + c_2 \Psi_2.$$

Here,  $c_1$  and  $c_2$  are arbitrary, in general, complex numbers.

- In quantum mechanics, dynamical variables like position, momentum, angular momentum, energy, etc., are called ‘Observables’. The observables are represented by **hermitian operators**.

Hermitian Operators naturally arise in quantum mechanics because their expectation values are real. Table 10.1 lists the operators that correspond to various observable quantities.

TABLE 10.1. Operators Associated with Various Observable Quantities

Quantity	Operator
Position, $x$	$x$
Linear momentum, $p$	$-i\hbar \frac{\partial}{\partial x}$ (or) $-i\hbar \nabla$
Angular momentum, $\mathbf{r} \times \mathbf{p}$	$-i\hbar \mathbf{r} \times \nabla$
Kinetic energy, KE = $\frac{p^2}{2m}$	$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}$ (or) $-\frac{\hbar^2}{2m} \nabla^2$
Total energy, $E$	$i\hbar \frac{\partial}{\partial t}$
Total energy (Hamiltonian form), $H$	$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x)$ (or) $-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r})$

**Linear Operator.** An operator transforms one function into another. A well-known example is the differential operator  $D = d/dx$ , e.g.,  $Dx^3 = 3x^2$ .

In quantum mechanics, we are concerned with linear operators.

An operator  $\alpha$  is said to be linear if

$$\alpha(\psi_1 + \psi_2) = \alpha \psi_1 + \alpha \psi_2 \text{ and } \alpha(a \psi) = a \alpha \psi.$$

$\psi_1$  and  $\psi_2$  are arbitrary functions.  $a$  is a constant which may or may not be complex.

The operator  $D$  is linear.

**Definition of Operator.** An operator is defined as a rule by which a given function is transformed into another function.

An operator tells us what operation to carry out on the quantity that follows it. Thus the operator  $i\hbar \frac{\partial}{\partial t}$  instructs us to take the partial derivative of what comes after it with respect to  $t$  and multiply the result by  $i\hbar$ .

### Postulate 3 – Eigenvalues

The only possible values which a measurement of the observable (whose operator is  $\alpha$ ) can yield are the eigen values  $a_n$  of the equation

$$\alpha \psi_n = a_n \psi_n, \quad n = 1, 2, 3, \dots$$

This is true provided  $\int \psi_n^* \psi_n d\tau < \infty$  and  $\psi_n$  is single-valued.

- The eigenfunctions  $\psi_n$  form a complete set of  $n$  independent functions.
- The eigenvalues  $a_n$  are just numbers.

An experiment always gives a real number. Hence, the eigenvalues have to be real.

Since Hermitian operators have real eigenvalues, the operators associated with physical quantities must be Hermitian.

**EXAMPLE.** An eigenfunction of the operator  $\frac{d^2}{dx^2}$  is  $\psi = e^{2x}$ . Find the corresponding eigenvalue.

**SOL.** The eigenvalue equation is  $\mathbf{G} \psi_n = G_n \psi_n$ .

Here,  $\mathbf{G}$  is the operator that corresponds to a certain dynamical variable  $G$  and each  $G_n$  is a real number.

$$\mathbf{G}\psi = \frac{d^2}{dx^2}(e^{2x}) = 4e^{2x} = 4\psi \quad (\because e^{2x} = \psi)$$

$$\therefore \mathbf{G}\psi = 4\psi.$$

The eigenvalue is  $G = 4$ .

### Postulate 4 – Expectation Values

The average value of an observable  $a$ , corresponding to the operator  $\alpha$ , for a system described by the wave function  $\psi$  is given by

$$\langle a \rangle = \frac{\int \psi^* \alpha \psi d\tau}{\int \psi^* \psi d\tau}.$$

The quantity  $\langle a \rangle$  is called the expectation value.

- If the wave function is normalised, the denominator is unity. We can write

$$\langle a \rangle = \int \psi^* \alpha \psi d\tau$$

- The first four postulates deal with the properties of the quantum system at any given instant of time.

### Postulate 5 – Time development of a quantum system

The development in time of the wave function  $\psi$  of a system is given by the equation.

$$\hat{H}\psi = i\hbar \frac{\partial}{\partial t} \psi.$$

Here,  $\hat{H}$  is the *Hamiltonian operator*.

This postulate is simply a statement of the time dependent Schrodinger equation. If the wave function is known at some initial time, the above equation determines  $\psi$  at any other time.

- The time development of state, *i.e.*, how the state changes as time progresses, is given by the Schrodinger equation.

## OPERATORS IN QUANTUM MECHANICS

### 10.1.1. Operator for Momentum

The wave function for a free particle moving in the positive  $x$ -direction is

$$\psi(x, t) = A e^{i/\hbar(p_x x - Et)} \quad \dots(1)$$

Differentiating Eq. (1) with respect to  $x$ , we get

$$\begin{aligned} \frac{\partial \psi}{\partial x} &= A \left( \frac{i}{\hbar} \right) p_x e^{i/\hbar(p_x x - Et)} = \frac{i}{\hbar} p_x \psi \\ \frac{\hbar \frac{\partial \psi}{\partial x}}{i} &= p_x \psi \\ -i\hbar \frac{\partial \psi}{\partial x} &= p_x \psi \end{aligned} \quad \dots(2)$$

Eq. (2) is the eigen value equation for the  $x$ -component of the momentum.

- $-i\hbar \frac{\partial}{\partial x}$  is the operator.
  - $\psi(x, t)$  is the *eigen function* of the operator.
  - $p_x$  is the *eigen value* of the operator.
- $\therefore -i\hbar \frac{\partial}{\partial x}$  is the operator for the  $x$ -component of the momentum. Thus

$$\hat{p}_x = -i\hbar \frac{\partial}{\partial x} \quad \dots(3)$$

Similarly, for the  $y$  and  $z$ -components of the momentum, the operators are

$$\hat{p}_y = -i\hbar \frac{\partial}{\partial y} \quad \dots(4)$$

$$\hat{p}_z = -i\hbar \frac{\partial}{\partial z} \quad \dots(5)$$

In three dimensions, the operator for the momentum  $\mathbf{p}$  is

$$\hat{p} = -i\hbar \nabla \quad \dots(6)$$

**EXAMPLE 1.** Find the eigen values of the linear momentum operator.

**Sol.** The eigenvalue equation of linear momentum is

$$\begin{aligned} -i\hbar \frac{d\psi}{dx} &= a\psi; (a = \text{eigenvalue}) \\ \frac{d\psi}{\psi} &= \frac{ia}{\hbar} dx \end{aligned}$$

Integrating,  $\ln \psi = \frac{ia}{\hbar} x + \ln C; C = \text{constant}$

$$\therefore \psi = C \exp\left(\frac{ia}{\hbar} x\right)$$

For  $\psi$  to be finite in the region  $-\infty$  to  $+\infty$ , the eigenvalue  $a$  has to be real. Hence, all real values of  $a$  are the eigenvalues of linear momentum. In other words, linear momentum is not quantized.

### 10.1.2. Operator for Kinetic Energy

From the momentum operator, we have

$$-i\hbar \frac{\partial \psi}{\partial x} = p_x \psi \quad \dots(1)$$

Differentiating Eq. (1) with respect to  $x$ , we get

$$-i\hbar \frac{\partial^2 \psi}{\partial x^2} = p_x \frac{\partial \psi}{\partial x}$$

But,  $\frac{\partial \psi}{\partial x} = \frac{1}{-i\hbar} p_x \psi$

$$\therefore -i\hbar \frac{\partial^2 \psi}{\partial x^2} = p_x \left( \frac{1}{-i\hbar} \right) p_x \psi$$

or  $-\hbar^2 \frac{\partial^2 \psi}{\partial x^2} = p_x^2 \psi \quad \dots(2)$

Dividing Eq. (2) by  $2m$ , we get

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} = \frac{p_x^2}{2m} \psi$$

or  $-\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} = K\psi \quad \dots(3)$

Here,  $K = \frac{p_x^2}{2m}$  = kinetic energy of the particle.

Eq. (3) is the eigen value equation for the kinetic energy of the particle moving in the  $x$ -direction.

- $-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}$  is the operator.
- $\psi(x, t)$  is the eigen function of the operator.
- $K = \frac{p_x^2}{2m}$  is the eigen value of the operator.

$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}$  is the kinetic energy operator for the motion of the particle in the  $x$ -direction.

$$\therefore \hat{K} = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \quad \dots(4)$$

In three dimensions, the operator for kinetic energy is

$$\hat{K} = -\frac{\hbar^2}{2m} \nabla^2 \quad \dots(5)$$

### 10.1.3. Operator for Total Energy

The total energy of a particle moving in the  $x$ -direction is given by

$$E = \frac{p_x^2}{2m} + V(x) \quad \dots(1)$$

Here,  $V(x)$  is the potential energy.

$$\text{or} \quad \frac{p_x^2}{2m} + V = E$$

Multiplying through by  $\psi(x, t)$ , we get

$$\frac{p_x^2}{2m} \psi + V \psi = E \psi$$

$$\text{But} \quad \frac{p_x^2}{2m} \psi = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2}$$

$$\therefore -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} + V \psi = E \psi$$

$$\left[ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V \right] \psi = E \psi \quad \dots(2)$$

Eq. (2) is the eigen value equation for the total energy  $E$  of the particle moving in the  $x$ -direction.

- $-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V$  is the operator.
- $\psi(x, t)$  is the eigen function of the operator.
- $E$  is the eigen value of the operator.

$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V$  is the total energy operator for the motion of the particle along the  $x$ -direction.

$$\hat{H}_x = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \quad \dots(3)$$

$\hat{H}_x$  is called the Hamiltonian operator for one-dimensional motion.

For three-dimensional motion, the total energy operator is

$$\hat{H} = -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \quad \dots(4)$$

Here,

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

$\hat{H}$  is called the *Hamiltonian operator*.

#### 10.1.4. Operator for Total Energy in terms of Partial Derivative with respect to Time

The wave function for a free particle moving in the positive  $x$ -direction is

$$\psi(x, t) = A e^{i/\hbar(p_x x - Et)} \quad \dots(1)$$

Differentiating Eq. (1) with respect to  $t$ , we get

$$\begin{aligned} \frac{\partial \psi}{\partial t} &= A \left( \frac{i}{\hbar} \right) (-E) e^{i/\hbar(p_x x - Et)} \\ \frac{\partial \psi}{\partial t} &= -\frac{i}{\hbar} E \psi \\ \frac{\partial \psi}{\partial t} &= \frac{1}{i\hbar} E \psi \\ \text{or} \quad i\hbar \frac{\partial}{\partial t} \psi &= E \psi \end{aligned} \quad \dots(2)$$

Evidently the dynamical quantity  $E$  in some sense corresponds to the differential operator  $i\hbar \frac{\partial}{\partial t}$ .

The total – energy operator is

$$\hat{E} = i\hbar \frac{\partial}{\partial t} \quad \dots(3)$$

#### 10.1.5. Orbital Angular Momentum Operator

The classical orbital angular momentum of a particle is

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} \quad \dots(1)$$

where  $\mathbf{r}$  and  $\mathbf{p}$  are the position and momentum vectors of the particle, respectively.  $\mathbf{p}$  is represented by the vector operator  $-i\hbar\nabla$ . So  $\mathbf{L}$  is represented by the vector operator  $-i\hbar(\mathbf{r} \times \nabla)$ .

$$\mathbf{L} = \mathbf{r} \times (-i\hbar \nabla) = -i\hbar \mathbf{r} \times \nabla \quad \dots(2)$$

Let  $L_x, L_y, L_z$  be cartesian components of  $\mathbf{L}$ .

$$\mathbf{i} L_x + \mathbf{j} L_y + \mathbf{k} L_z = -i\hbar \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x & y & z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \end{vmatrix}$$

$$\therefore \hat{L}_x = -i\hbar \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right)$$

$$\hat{L}_y = -i\hbar \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right)$$

$$\hat{L}_z = -i\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) \quad \dots(3)$$

Let us now represent Cartesian components  $L$  in terms of spherical polar coordinates. The Cartesian coordinates  $(x, y, z)$  and spherical polar coordinates  $(r, \theta, \phi)$ , are related by Fig. 10.1.

$$\begin{aligned} x &= r \sin \theta \cos \phi, \\ y &= r \sin \theta \sin \phi, \\ z &= r \cos \theta. \end{aligned} \quad \dots(1)$$

These equations yield

$$r^2 = x^2 + y^2 + z^2, \cos \theta = \frac{z}{r}. \quad \dots(2)$$

$$\tan \phi = \frac{y}{x}$$

Using Eqs. (1) and (2), we obtain

$$\begin{aligned} \frac{\partial r}{\partial x} &= \sin \theta \cos \phi, \frac{\partial r}{\partial y} = \sin \theta \sin \phi, \frac{\partial r}{\partial z} = \cos \theta \\ \frac{\partial \theta}{\partial x} &= \frac{1}{r} \cos \theta \cos \phi, \frac{\partial \theta}{\partial y} = \frac{1}{r} \cos \theta \sin \phi, \frac{\partial \theta}{\partial z} = -\frac{1}{r} \sin \theta \\ \frac{\partial \phi}{\partial x} &= -\frac{\sin \phi}{r \sin \theta}, \frac{\partial \phi}{\partial y} = \frac{\cos \phi}{r \sin \theta}, \frac{\partial \phi}{\partial z} = 0. \end{aligned} \quad \dots(3)$$

Using these relations, we obtain

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{\partial f}{\partial r} \frac{\partial r}{\partial x} + \frac{\partial f}{\partial \theta} \frac{\partial \theta}{\partial x} + \frac{\partial f}{\partial \phi} \frac{\partial \phi}{\partial x} \\ &= \sin \theta \cos \phi \frac{\partial f}{\partial r} + \frac{1}{r} \cos \theta \cos \phi \frac{\partial f}{\partial \theta} - \left( -\frac{1}{r} \frac{\sin \phi}{\sin \theta} \frac{\partial f}{\partial \phi} \right) \end{aligned} \quad \dots(4)$$

$$\begin{aligned} \frac{\partial f}{\partial y} &= \frac{\partial f}{\partial r} \frac{\partial r}{\partial y} + \frac{\partial f}{\partial \theta} \frac{\partial \theta}{\partial y} + \frac{\partial f}{\partial \phi} \frac{\partial \phi}{\partial y} \\ &= \sin \theta \sin \phi \frac{\partial f}{\partial r} + \frac{1}{r} \cos \theta \sin \phi \frac{\partial f}{\partial \theta} + \frac{1}{r} \frac{\cos \phi}{\sin \theta} \frac{\partial f}{\partial \phi}. \end{aligned} \quad \dots(5)$$

$$\frac{\partial f}{\partial z} = \frac{\partial f}{\partial r} \frac{\partial r}{\partial z} + \frac{\partial f}{\partial \theta} \frac{\partial \theta}{\partial z} + \frac{\partial f}{\partial \phi} \frac{\partial \phi}{\partial z} = \cos \theta \frac{\partial f}{\partial r} - \frac{1}{r} \sin \theta \frac{\partial f}{\partial \theta}. \quad \dots(6)$$

$$L_x = i\hbar \left( \sin \phi \frac{\partial}{\partial \theta} + \cot \theta \cos \phi \frac{\partial}{\partial \phi} \right) \quad \dots(7)$$

$$L_y = i\hbar \left( -\cos \phi \frac{\partial}{\partial \theta} + \cot \theta \sin \phi \frac{\partial}{\partial \phi} \right) \quad \dots(8)$$

$$L_z = -i\hbar \frac{\partial}{\partial \phi}. \quad \dots(9)$$

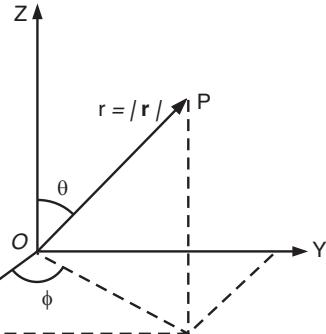


Fig. 10.1

$L^2$  in spherical polar coordinates is given by

$$\begin{aligned} L^2 &= L_x^2 + L_y^2 + L_z^2 \\ &= -\hbar^2 \left[ \left\{ \sin^2 \phi \frac{\partial^2}{\partial \theta^2} + \cot^2 \theta \cos^2 \phi \frac{\partial^2}{\partial \phi^2} + \sin \phi \frac{\partial}{\partial \theta} \left( \cot \theta \cos \phi \frac{\partial}{\partial \phi} \right) \right\} \right] \end{aligned}$$

$$\begin{aligned}
& + \cot \theta \cos \phi \frac{\partial}{\partial \phi} \left( \sin \phi \frac{\partial}{\partial \theta} \right) \Bigg) + \left\{ \cos^2 \phi \frac{\partial^2}{\partial \theta^2} + \cot^2 \theta \sin^2 \phi \frac{\partial^2}{\partial \phi^2} \right. \\
& \left. - \cos \phi \frac{\partial}{\partial \theta} \left( \cot \theta \sin \phi \frac{\partial}{\partial \phi} \right) + \cot \theta \sin \phi \frac{\partial}{\partial \phi} \left( -\cos \phi \frac{\partial}{\partial \theta} \right) \right\} + \left. \left\{ \frac{\partial^2}{\partial \phi^2} \right\} \right] \\
= & -\hbar^2 \left[ \frac{\partial^2}{\partial \theta^2} + (\cot^2 \theta + 1) \frac{\partial^2}{\partial \phi^2} + \sin \phi (-\operatorname{cosec}^2 \theta) \cos \phi \frac{\partial}{\partial \phi} \right. \\
& \left. + \cot \theta \cos^2 \phi \frac{\partial}{\partial \theta} - \cos \phi (-\operatorname{cosec}^2 \theta) \sin \phi \frac{\partial}{\partial \phi} + \cot \theta \sin^2 \phi \frac{\partial}{\partial \theta} \right] \\
= & -\hbar^2 \left[ \frac{\partial^2}{\partial \theta^2} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} + \cot \theta \frac{\partial}{\partial \theta} \right] \\
= & -\hbar^2 \left[ \frac{\cos \theta}{\sin \theta} \frac{\partial}{\partial \theta} + \frac{\partial^2}{\partial \theta^2} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right] \\
\therefore L^2 = & -\hbar^2 \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right] \quad \dots(10)
\end{aligned}$$

This is the expression for the operator of the square of the angular momentum.

#### Expressions for the Eigen values of the square of the total angular momentum ( $L^2$ ) and its z-component ( $L_z$ ).

The angular momentum of atoms and molecules is almost as important as the energy as it helps to give physical significance to our mathematical functions. For a one-electron atom, the angular momentum  $\mathbf{L}$  is quantized both in magnitude and direction, with relations

$$|\mathbf{L}|^2 = l(l+1)\hbar^2$$

and

$$L_z = m_l \hbar.$$

Here,  $|\mathbf{L}|^2$  is the square of the magnitude of the orbital angular momentum about the centre of the atom and  $L_z$  is the component of the angular momentum along the z-axis.

$l$  and  $m_l$  are quantum numbers which restrict  $|\mathbf{L}|^2$  and  $L_z$  to certain specific values.

To prove the above relations for an atom we must show that the wave functions of the atom are eigenfunctions of the angular momentum operators  $\hat{\mathbf{L}}^2$  and  $\hat{L}_z$  with eigenvalues  $l(l+1)\hbar^2$  and  $m_l\hbar$  respectively.

Let us now first apply the operator  $\hat{L}_z$  to the one-electron atom wave function

$$\psi(r, \theta, \phi) = R(r) \Theta(\theta) \Phi(\phi)$$

$$\hat{L}_z \psi = -i\hbar \frac{\partial \psi}{\partial \phi}$$

$$\text{or} \quad L_z \psi = -i\hbar R \Theta \frac{d\Phi}{d\phi}$$

The function  $\Phi(\phi)$  for the atom is given by

$$\Phi = A e^{im_l \phi}$$

$$\frac{d\Phi}{d\phi} = A i m_l e^{im_l \phi} = i m_l \Phi.$$

Substituting it in the above expression for  $\hat{L}_z \psi$ , we get

$$\hat{L}_z \psi = -i\hbar im_l R \Theta \Phi$$

or  $\hat{L}_z \psi = m_l \hbar \psi$

This result means that the wave functions  $\psi$  of the one-electron atom are the eigen functions of  $\hat{L}_z$  having eigen values given by

$$L_z = m_l \hbar. \quad \dots(4)$$

This is the expression for the quantized values of the  $z$ -component of the angular momentum of the atom. ( $L_x$  and  $L_y$ , however, do *not* obey quantization relations).

We next apply the operator  $\hat{\mathbf{L}}^2$  to the wave function  $\psi = R \Theta \Phi$ .

$$\begin{aligned} \hat{\mathbf{L}}^2 \psi &= -\hbar^2 \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right] R \Theta \Phi \quad \dots(5) \\ &= -\hbar^2 R \left[ \frac{\Phi}{\sin \theta} \frac{d}{d \theta} \left( \sin \theta \frac{d \Theta}{d \theta} \right) + \frac{\Theta}{\sin^2 \theta} \frac{d^2 \Phi}{d \phi^2} \right]. \end{aligned}$$

The  $\Phi$ -equation is  $\Phi = A e^{im_l \Phi}$

$$\frac{d^2 \Phi}{d \phi^2} = A (im_l)^2 e^{im_l \Phi} = -m_l^2 \Phi.$$

∴  $\hat{\mathbf{L}}^2 \psi = -\hbar^2 R \Phi \left[ \frac{1}{\sin \theta} \frac{d}{d \theta} \left( \sin \theta \frac{d \Theta}{d \theta} \right) - \frac{m_l^2}{\sin^2 \theta} \Theta \right].$

The function  $\Theta(\theta)$  is a solution of the equation

$$\frac{1}{\sin \theta} \frac{d}{d \theta} \left( \sin \theta \frac{d \Theta}{d \theta} \right) + \left[ l(l+1) - \frac{m_l^2}{\sin^2 \theta} \right] \Theta = 0.$$

This means that the quantity in brackets in the above expression for  $\hat{\mathbf{L}}^2 \psi$  is equal to  $-l(l+1)\Theta$ .

Hence we get

$$\hat{\mathbf{L}}^2 \psi = -\hbar^2 R \Phi [-l(l+1)\Theta]$$

or  $\hat{\mathbf{L}}^2 \psi = l(l+1)\hbar^2 R \Theta \Phi$

or  $\hat{\mathbf{L}}^2 \psi = l(l+1)\hbar^2 \psi$

It means that the wave functions  $\psi$  of the one-electron atom are the eigenfunctions of  $\hat{\mathbf{L}}^2$  having eigenvalues given by

$$|\mathbf{L}|^2 = l(l+1)\hbar^2$$

This is the expression for the quantized values of the square of the angular momentum of the atom.

## 10.2 COMMUTING OPERATORS

- If  $\hat{A}\hat{B} = \hat{B}\hat{A}$ ,  $\hat{A}$  and  $\hat{B}$  are said to commute.
- If  $\hat{A}\hat{B} \neq \hat{B}\hat{A}$ ,  $\hat{A}$  and  $\hat{B}$  are said not to commute.

**Example.** If  $\hat{A} = x$  and  $\hat{B} = \frac{d}{dx}$  then,

$$\begin{aligned}\hat{A}\hat{B}\psi &= x \frac{d\psi}{dx} \\ \hat{B}\hat{A}\psi &= \frac{d}{dx}(x\psi) \\ &= \psi + x \frac{d\psi}{dx}\end{aligned}$$

Hence

$$\hat{A}\hat{B}\psi \neq \hat{B}\hat{A}\psi$$

or

$$\hat{A}\hat{B} \neq \hat{B}\hat{A}$$

In the special case if  $\hat{A}\hat{B} = \hat{B}\hat{A}$ , then the operators  $A$  and  $B$  are said to commute.

- The difference  $AB - BA$  is called the commutator of  $A$  and  $B$ .
- A bracket notation is used to denote the commutator

$$[A, B] \equiv AB - BA.$$

- When two operators commute, it is possible to measure both the variables associated with them simultaneously and accurately. Further, there will be no uncertainty in either of them.
- If two operators do not commute, it is not possible to measure the variables associated with them simultaneously and accurately. If one of them is measured accurately, it is not possible to measure the other simultaneously with accuracy.

### 10.2.1. Simultaneous eigenfunctions

If  $A$  and  $B$  are linear operators, and  $\psi$  is a function satisfying both the equations

$$A\psi = \alpha\psi, \quad B\psi = \beta\psi, \quad \dots(1)$$

then  $\psi$  is a *simultaneous* eigenfunction of  $A$  and  $B$ , belonging to the eigenvalues  $\alpha$  and  $\beta$ , respectively.

For example, consider a nonrelativistic free particle, of energy  $E = \frac{1}{2}mv^2$  and momentum  $p = mv$ . The wave function

$$\psi_p = \frac{1}{\sqrt{2\pi\hbar}} e^{(i/\hbar)px}, \quad \dots(2)$$

satisfies simultaneously the equations

$$\hat{p}\psi = p\psi \quad \text{and} \quad \hat{H}\psi = E\psi.$$

Here,  $\hat{H} = \frac{\hat{p}^2}{2m}$  is the Hamiltonian operator for a free particle.

Equations (1) imply that

$$B A\psi = B(\alpha\psi) = \alpha B\psi = \alpha\beta\psi,$$

and

$$A B\psi = A(\beta\psi) = \beta A\psi = \beta\alpha\psi.$$

By subtraction,

$$(AB - BA)\psi = 0. \quad \dots(3)$$

This equation shows that  $\psi$  is also an eigenfunction of the operator  $(AB - BA)$ , belonging to the eigenvalue zero. The condition (3) is necessary in order that  $\psi$  be a simultaneous eigenfunction of  $A$  and  $B$ .

The operator in Eq. (3) is called the *commutator* of  $A$  and  $B$  and is written,

$$[A, B] = AB - BA. \quad \dots(4)$$

Two operators satisfying the equation

$$[A, B] = 0 \quad \dots(5)$$

are said to *commute*. This equation means that Eq. (3) is true for every  $\psi$  which is a member of the class of functions under consideration.

*The eigenfunctions of commuting operators can always be constructed in such a way that they are simultaneous eigenfunctions.*

### 10.2.2. Simultaneous Measurability of Observables

If two observables are simultaneously measurable in a particular state of a given system, then the state function is an eigenfunction of both the operators. Two observables are said to be *compatible*, if their operators have a common set of eigenfunctions. The following two theorems indicate the connection between compatible observables and commuting operators.

**Theorem 1.** *Operators having common set of eigenfunctions commute.*

**Proof.** Consider operators  $A$  and  $B$  with the common set of eigenfunctions  $\psi_i, i = 1, 2, \dots$  as

$$A\psi_i = a_i\psi_i, \quad \text{and} \quad B\psi_i = b_i\psi_i \quad \dots(1)$$

$$\text{Then} \quad AB\psi_i = A(b_i\psi_i) = b_iA\psi_i = a_i b_i \psi_i \quad \dots(2)$$

$$\text{and} \quad BA\psi_i = B(a_i\psi_i) = a_iB\psi_i = a_i b_i \psi_i \quad \dots(3)$$

Since  $AB\psi_i = BA\psi_i$ ,  $A$  commutes with  $B$ . Hence the result.

**Theorem 2.** *Commuting operators have common set of eigenfunctions.*

**Proof.** Consider two commuting operators  $A$  and  $B$ . The eigenvalue equation for  $A$  is

$$A\psi_i = a_i\psi_i, \quad i = 1, 2, \dots \quad \dots(1)$$

Operating both sides from left by  $B$ , we get

$$BA\psi_i = a_iB\psi_i$$

Since  $B$  commutes with  $A$ ,

$$A(B\psi_i) = a_i(B\psi_i) \quad \dots(2)$$

That is,  $B\psi_i$  is an eigenfunction of  $A$  with the same eigenvalue  $a_i$ . If  $A$  has only nondegenerate eigenvalues,  $B\psi_i$  can differ from  $\psi_i$  only by a multiplicative constant, say  $b_i$

$$B\psi_i = b_i\psi_i \quad \dots(3)$$

In other words,  $\psi_i$  is a simultaneous eigenfunction of both  $A$  and  $B$ .

### Explanaton of the Two Theorems (Theorem of commutativity and simultaneity)

**Theorem.** *If two operators  $A$  and  $B$  possess a complete set of simultaneous eigen functions, then the operators  $A$  and  $B$  commute, i.e.,*

$$AB = BA \quad \text{or} \quad AB - BA = 0 \quad \text{or} \quad [A, B] = 0.$$

### Converse of theorem

The converse of the theorem is stated as follows:

**Theorem.** *If two operators commute, they possess a set of simultaneous eigen functions.*

### 10.2.3. Commutator Algebra

We list here some elementary rules for the calculation of commutators.

If  $A$ ,  $B$ , and  $C$  are three linear operators

$$[A, B] = -[B, A] \quad \dots(1)$$

$$[A, B + C] = [A, B] + [A, C] \quad \dots(2)$$

$$[A, BC] = [A, B] C + B [A, C] \quad \dots(3)$$

$$[AB, C] = [A, C] B + A [B, C] \quad \dots(4)$$

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0 \quad \dots(5)$$

Now, let us simplify  $[A, [B, C]]$ .

$$\begin{aligned} [A, [B, C]] &= A [B, C] - [B, C] A \\ &= A (BC - CB) - (BC - CB) A \\ &= ABC - ACB - BCA + CBA. \end{aligned}$$

### (a) Commutation relation between position and momentum

(i) Let  $x$  and  $p_x$  be represented as

$$x \rightarrow \hat{x}, p_x \rightarrow \hat{p}_x = -i\hbar \frac{\partial}{\partial x}.$$

Consider a function  $\psi(x)$ . Then

$$p_x \psi = -i\hbar \frac{\partial \psi}{\partial x} \quad \dots(1)$$

and

$$xp_x \psi = x \left( -i\hbar \frac{\partial \psi}{\partial x} \right) \quad \dots(1)$$

$$\text{Also } xp_x \psi = -i\hbar \frac{\partial}{\partial x} (x\psi) = -i\hbar \psi + x \left( -i\hbar \frac{\partial \psi}{\partial x} \right) \quad \dots(2)$$

Eq. (1) – Eq. (2) gives,

$$xp_x \psi - p_x x\psi = x \left( -i\hbar \frac{\partial \psi}{\partial x} \right) - \left[ -i\hbar \psi + x \left( -i\hbar \frac{\partial \psi}{\partial x} \right) \right] = i\hbar \psi \quad \dots(3)$$

$$\therefore [x, p_x] = i\hbar \quad \dots(3)$$

$$\text{Now, } p_y = -i\hbar \frac{\partial}{\partial y} \text{ and } p_z = -i\hbar \frac{\partial}{\partial z}.$$

We have proved the relation  $[x, p_x] = i\hbar$ .

More generally, we have

$$[x, p_x] = [y, p_y] = [z, p_z] = i\hbar. \quad \dots(4)$$

while all other commutators such as  $[x, p_y]$  – vanish. It follows that, for example,  $x$  and  $p_y$  have common eigenfunctions and can be measured simultaneously with arbitrary accuracy. In contrast, since  $x$  and  $p_x$  do not commute with each other, a precise simultaneous measurement of both of these observables is impossible.

$$(ii) [x^2, p_x] = [xx, p_x] = [x, p_x] x + x [x, p_x] = i\hbar x + xi\hbar = 2i\hbar x.$$

### (b) Commutation relation between Hamiltonian H and momentum p

The Hamiltonian operator for free particle is

$$\hat{H}_x = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \text{ since potential energy } V = 0$$

$$\text{and } \hat{p}_x = -i\hbar \frac{d}{dx}.$$

Consider a function  $\psi(x)$ . Then,

$$p \psi = -i\hbar \frac{d \psi}{dx}$$

$$\therefore H_p \psi = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \left( -i\hbar \frac{d\psi}{dx} \right).$$

Similarly,  $pH \psi = -i\hbar \frac{d}{dx} \left[ -\frac{\hbar^2}{2m} \frac{d^2 \psi}{dx^2} \right]$

Solving,  $[H, p] = Hp - pH = 0$ ,

i.e., momentum of a free particle commutes with the Hamiltonian operator.

**(c) The commutation rules for the components of orbital angular momentum ( $\mathbf{L}$ )**

Let  $L_x, L_y, L_z$  be cartesian components of  $\mathbf{L}$ .

$$\begin{aligned} \hat{L}_x &= -i\hbar \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) \\ \hat{L}_y &= -i\hbar \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right) \\ \hat{L}_z &= -i\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) \end{aligned} \quad \dots(1)$$

Now,

$$[L_x, L_y] = L_x L_y - L_y L_x.$$

Substituting operator values

$$\begin{aligned} L_x L_y &= (-i\hbar)^2 \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right) \\ &= -\hbar^2 \left\{ y \frac{\partial}{\partial z} \left( z \frac{\partial}{\partial x} \right) - y \frac{\partial}{\partial z} \left( x \frac{\partial}{\partial z} \right) - z \frac{\partial}{\partial y} \left( z \frac{\partial}{\partial x} \right) + z \frac{\partial}{\partial y} \left( x \frac{\partial}{\partial z} \right) \right\} \\ &= -\hbar^2 \left\{ y \frac{\partial}{\partial x} + yz \frac{\partial^2}{\partial z \partial x} - yx \frac{\partial^2}{\partial z^2} - z^2 \frac{\partial^2}{\partial y \partial x} + zx \frac{\partial^2}{\partial y \partial z} \right\} \end{aligned}$$

Similarly,

$$\begin{aligned} L_y L_x &= -\hbar^2 \left\{ \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right) \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) \right\} \\ &= -\hbar^2 \left\{ z \frac{\partial}{\partial x} \left( y \frac{\partial}{\partial z} \right) - z \frac{\partial}{\partial x} \left( z \frac{\partial}{\partial y} \right) - x \frac{\partial}{\partial z} \left( y \frac{\partial}{\partial z} \right) + x \frac{\partial}{\partial z} \left( z \frac{\partial}{\partial y} \right) \right\} \\ &= -\hbar^2 \left\{ zy \frac{\partial^2}{\partial x \partial z} - z^2 \frac{\partial^2}{\partial x \partial y} - xy \frac{\partial^2}{\partial z^2} + x \frac{\partial}{\partial y} + xz \frac{\partial^2}{\partial z \partial y} \right\} \end{aligned}$$

But  $\frac{\partial^2}{\partial x \partial z} = \frac{\partial^2}{\partial z \partial x}$  and  $\frac{\partial^2}{\partial y \partial z} = \frac{\partial^2}{\partial z \partial y}$  and so on.

$$\therefore L_x L_y - L_y L_x = [L_x, L_y] = -\hbar^2 \left\{ y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \right\}$$

Thus

$$\begin{aligned} [L_x, L_y] &= -i\hbar \left\{ -i\hbar \left( y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \right) \right\} \\ &= i\hbar \left\{ -i\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) \right\} \end{aligned}$$

$$[L_x, L_y] = i\hbar L_z \quad \dots(2)$$

$\hat{L}_x, \hat{L}_y, \hat{L}_z$  do not commute with one another.

$$[\hat{L}_x, \hat{L}_y] = i\hbar \hat{L}_z; [\hat{L}_y, \hat{L}_z] = i\hbar \hat{L}_x; [\hat{L}_z, \hat{L}_x] = i\hbar \hat{L}_y$$

**(d) Commutation relation of  $L^2$  with components  $L_x, L_y$  and  $L_z$**

We have,

$$\begin{aligned} L^2 &= L_x^2 + L_y^2 + L_z^2 \\ [L^2, L_x] &= [L_x^2 + L_y^2 + L_z^2, L_x] \\ &= (L_x^2 + L_y^2 + L_z^2) L_x - L_x (L_x^2 + L_y^2 + L_z^2) \\ &= L_x^2 L_x + L_y^2 L_x + L_z^2 L_x - (L_x L_x^2 + L_x L_y^2 + L_x L_z^2) \\ &= (L_y^2 L_x - L_x L_y^2) + (L_z^2 L_x - L_x L_z^2) \\ &= [L_y^2, L_x] + [L_z^2, L_x] \end{aligned}$$

We know that

So,

$$\begin{aligned} [AB, C] &= [A, C]B + A[B, C] \\ [L^2, L_x] &= [L_y L_y, L_x] + [L_z L_z, L_x] \\ &= [L_y, L_x] L_y + L_y [L_y, L_x] + [L_z, L_x] L_z + L_z [L_z, L_x] \end{aligned}$$

Now,

$$[L_x, L_y] = i\hbar L_z; [L_y, L_x] = -i\hbar L_z; [L_z, L_x] = i\hbar L_y$$

∴

$$[L^2, L_x] = (-i\hbar L_z) L_y + L_y (-i\hbar L_z) + (i\hbar L_y) L_z + L_z (i\hbar L_y) = 0.$$

∴

$$[L^2, L_x] = 0 \quad \dots(1\ a)$$

Similarly,

$$[L^2, L_y] = 0 \quad \dots(1\ b)$$

$$[L^2, L_z] = 0 \quad \dots(1\ c)$$

Hence  $L^2$  commutes with any of the three components of the angular momentum operator.

**(e) Ladder operators  $L_+$  and  $L_-$ :** Let us define the operators

$$L_+ = L_x + iL_y \text{ and } L_- = L_x - iL_y \quad \dots(2)$$

**Commutation Relations of  $L_z$  with  $L_+$  and  $L_-$**

$$\begin{aligned} [L_z, L_+] &= [L_z, L_x + iL_y] = [L_z, L_x] + i[L_z, L_y] \\ &= i\hbar L_y + i(-i\hbar L_x) \\ &= i\hbar L_y + \hbar L_x \\ &= \hbar [L_x + iL_y] = \hbar L_+ \end{aligned} \quad \dots(3)$$

and

$$\begin{aligned} [L_z, L_-] &= [L_z, L_x - iL_y] \\ &= [L_z, L_x] - i[L_z, L_y] \\ &= i\hbar L_y - i(-i\hbar L_x) \\ &= i\hbar L_y - \hbar L_x = -\hbar (L_x - iL_y) \\ &= -\hbar L_- \end{aligned} \quad \dots(4)$$

∴

$$[L_z, L_\pm] = \pm \hbar L_\pm \quad \dots(5)$$

**Commutation Relation of  $L_+$  and  $L_-$  mutually**

$$\begin{aligned} [L_+, L_-] &= [(L_x + iL_y), (L_x - iL_y)] \\ &= [L_x, L_x] - i[L_x, L_y] + i[L_y, L_x] + [L_y, L_y] \\ &= 0 - i(i\hbar L_z) + i(-i\hbar L_z) + 0 \\ &= \hbar L_z + \hbar L_z \\ &= 2\hbar L_z. \end{aligned} \quad \dots(6)$$

$$[L_+, L_-] = 2\hbar L_z.$$

(f) Commutation relation of orbital angular momentum with position

$$\begin{aligned}[L_x, x] &= [L_x x - x L_x] \\ &= \left\{ y \left( -i\hbar \frac{\partial}{\partial z} \right) - z \left( -i\hbar \frac{\partial}{\partial y} \right) \right\} x - x \left\{ y \left( -i\hbar \frac{\partial}{\partial z} \right) - z \left( -i\hbar \frac{\partial}{\partial y} \right) \right\}\end{aligned}$$

If  $\psi(x)$  is a function of  $x$ , we have

$$\begin{aligned}(L_x x - x L_x) \psi &= \left\{ y \left( -i\hbar \frac{\partial}{\partial z} (x\psi) \right) - z \left( -i\hbar \frac{\partial}{\partial y} (x\psi) \right) \right\} - x \left\{ y \left( -i\hbar \frac{\partial \psi}{\partial z} \right) - z \left( -i\hbar \frac{\partial \psi}{\partial y} \right) \right\} \\ &= 0\end{aligned}$$

or  $[L_x, x] = 0$

Similarly,

$$\begin{aligned}[L_x, y] &= i\hbar z; & [L_x, z] &= -i\hbar y; \\ [L_y, y] &= 0; & [L_y, z] &= i\hbar x; & [L_y, x] &= -i\hbar z, \\ [L_z, y] &= -i\hbar x; & [L_z, z] &= 0; & [L_z, x] &= i\hbar y\end{aligned} \quad \dots(7)$$

Similarly, relations between  $\mathbf{L}$  and  $\mathbf{p}$  vectors by keeping in mind the cyclic order of the indices are

$$x \rightarrow y, y \rightarrow z, z \rightarrow x.$$

Thus  $[L_x, p_x] = 0, [L_x, p_y] = i\hbar p_z, [L_x, p_z] = -i\hbar p_y$  etc.

#### 10.2.4. Hermitian operator

**Definition.** An operator  $A$  operating on any two functions  $\psi_m(x)$  and  $\psi_n(x)$  is said to be *Hermitian*, if it satisfies the relation

$$\int \psi_m^* A \psi_n dx = \int (A^* \psi_m^*) \psi_n dx \quad \dots(1)$$

The integration here is over the domain in which the functions  $\psi_m(x)$  and  $\psi_n(x)$  are defined.

If an operator satisfies Eq. (1) whenever  $\psi_m(x)$  and  $\psi_n(x)$  are normalisable, we call it *Hermitian*, or *self-adjoint*.

#### Properties of Hermitian Operators

1. The eigenvalues of Hermitian operators are real.

**Proof.** Consider the eigenvalue equation of a Hermitian operator  $A$  such that

$$A \psi_n(x) = a_n \psi_n(x) \quad \dots(1)$$

Multiplying Eq. (1) from left by  $\psi_n^*$  and integrating with respect to  $x$  from  $-\infty$  to  $+\infty$ , we have

$$\int_{-\infty}^{\infty} \psi_n^* A \psi_n dx = a_n \int_{-\infty}^{\infty} \psi_n^* \psi_n dx = a_n \quad \dots(2)$$

When the Hermitian property of  $A$  is used, Eq. (2) reduces to

$$\int_{-\infty}^{\infty} (A^* \psi_n^*) \psi_n dx = a_n$$

or  $a_n^* \int_{-\infty}^{\infty} \psi_n^* \psi_n dx = a_n$

$\therefore a_n^* = a_n \quad \dots(3)$

Eq. (3) is possible only if  $a_n$  is real.

This proves that the eigen values of a Hermitian operator are real.

2. Eigenfunctions of a Hermitian operator that belong to distinct eigenvalues are orthogonal.

**Proof:** Let the Hermitian operator be  $A$ .

Let the eigen functions be  $\psi_n(x)$ ,  $\psi_m(x)$  with corresponding eigen values  $a_n$ ,  $a_m$ . That is

$$A \psi_n(x) = a_n \psi_n(x) \quad \dots(1)$$

and

$$A \psi_m(x) = a_m \psi_m(x) \quad \dots(2)$$

Multiplying Eq. (1) from left by  $\psi_m^*$  and integrating in the limits  $-\infty$  to  $+\infty$ , we get

$$\int_{-\infty}^{\infty} \psi_m^* A \psi_n dx = a_n \int_{-\infty}^{\infty} \psi_m^* \psi_n dx$$

Since operator  $A$  is Hermitian,

$$\int_{-\infty}^{\infty} A^* \psi_m^* \psi_n dx = a_n \int_{-\infty}^{\infty} \psi_m^* \psi_n dx$$

or

$$a_m \int_{-\infty}^{\infty} \psi_m^* \psi_n dx = a_n \int_{-\infty}^{\infty} \psi_m^* \psi_n dx$$

or

$$(a_m - a_n) \int_{-\infty}^{\infty} \psi_m^* \psi_n dx = 0$$

Since the eigenvalues are distinct,  $(a_m - a_n) \neq 0$ .

$$\therefore \int_{-\infty}^{\infty} \psi_m^* \psi_n dx = 0 \quad \dots(3)$$

That is,  $\psi_m$  and  $\psi_n$  are orthogonal.

**EXAMPLE.** Show that the momentum operator  $\frac{\hbar}{i} \cdot \frac{\partial}{\partial x}$  is Hermitian. (Purvanchal 2005)

**SOL.** Momentum operator  $\hat{p} = \frac{\hbar}{i} \cdot \frac{\partial}{\partial x}$

Complex conjugate of  $\hat{p}$ ,  $\hat{p}^* = -\frac{\hbar}{i} \cdot \frac{\partial}{\partial x}$

If  $\hat{p}$  is hermitian operator, then in a given state  $\psi$ , its average value should be real.

i.e.,  $\langle \hat{p} \rangle = \int_{-\infty}^{\infty} \psi^* \frac{\hbar}{i} \cdot \frac{\partial \psi}{\partial x} dx$  should be real.

Integrating the equation by parts,

$$\langle \hat{p} \rangle = \frac{\hbar}{i} [\psi^* \psi]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{\hbar}{i} \cdot \frac{\partial \psi^*}{\partial x} \psi dx = \int_{-\infty}^{\infty} \psi \left( -\frac{\hbar}{i} \cdot \frac{\partial \psi^*}{\partial x} \right) dx = \langle p^* \rangle$$

Since the expectation value of  $\hat{p}$  is equal to the expectation value of its complex conjugate  $\hat{p}^*$ , it is obvious that  $\langle \hat{p} \rangle$  is real. Hence the momentum operator  $\hat{p}$  is hermitian.

**II Method.** Prove that the operator  $-i\hbar \frac{d}{dx}$  is Hermitian.

**SOL.** Consider the integral

$$\int_{-\infty}^{\infty} \psi_a^* \left( -i\hbar \frac{d}{dx} \right) \psi_b dx \quad \dots(1)$$

I use integration by parts.

$$\int_{-\infty}^{\infty} \psi_a^* \left( -i\hbar \frac{d}{dx} \right) \psi_b dx = -i\hbar \left[ \psi_a^* \psi_b \right]_{-\infty}^{\infty} + i\hbar \int_{-\infty}^{\infty} \psi_b \frac{d}{dx} \psi_a^* dx$$

I throw away the boundary term (first term on the R.H.S.) for the usual reason:  $\psi_a$  and  $\psi_b$  must go to zero at  $\pm\infty$ .

$$\therefore \int_{-\infty}^{\infty} \psi_a^* \left( -i\hbar \frac{d}{dx} \right) \psi_b dx = \int_{-\infty}^{\infty} \left( -i\hbar \frac{d\psi_a}{dx} \right)^* \psi_b dx \quad \dots(2)$$

Eq. (2) is the condition for the operator  $-i\hbar \frac{d}{dx}$  to be Hermitian. Hence, the operator  $-i\hbar \frac{d}{dx}$  is Hermitian.

#### 10.2.5. Parity Operator

For functions of the variable  $x$ , the parity operator  $\hat{\pi}$  is defined as

$$\hat{\pi}\psi(x) = \psi(-x)$$

i.e., when the wavefunction  $\psi(x)$  is operated by the parity operator, it gets reflected in its coordinates.

##### Properties of the Parity Operator

**1.  $\pi$ -operator is a linear operator.** Let the parity operator operate on wavefunctions  $\psi_1$  and  $\psi_2$ . Then

$$\hat{\pi}[\psi_1(x) + \psi_2(x)] = \psi_1(-x) + \psi_2(-x) = \hat{\pi}\psi_1(x) + \hat{\pi}\psi_2(x)$$

$$\text{Also, } \hat{\pi}C\psi_1(x) = C\psi(-x) = C\hat{\pi}\psi(x).$$

Hence  $\hat{\pi}$  operator is a linear operator.

**2.  $\hat{\pi}$  operator is Hermitian.** The scalar product of two functions  $\psi$  and  $\phi$  is defined as

$$(\phi, \psi) = (\psi, \phi)^* = \int \phi^* \psi dx \quad \dots(1)$$

Consider the scalar product of  $\hat{\pi}\psi$  and a function  $\phi$  i.e.,

$$(\hat{\pi}\psi, \phi) = \int_{-\infty}^{+\infty} \psi^*(-x) \phi(x) dx = \int_{-\infty}^{+\infty} \psi^*(x') \phi(-x') dx'$$

The second integral is obtained from the first by the substitution  $x' = -x$ .

Since the value of the integral is unaffected by renaming the variable of integration, we have

$$(\hat{\pi}\psi, \phi) = \int_{-\infty}^{\infty} \psi^*(x) \phi(-x) dx = (\psi, \hat{\pi}\phi) \quad \dots(2)$$

Eq. (2) shows that  $\hat{\pi}$  is a Hermitian operator.

The eigenvalues of  $\hat{\pi}$  are therefore real.

**3. Eigenvalues of parity operator.** The eigenvalue of the  $\hat{\pi}$  operator is given by

$$\hat{\pi}\psi = \lambda\psi. \quad \dots(1)$$

Here,  $\lambda$  is the eigenvalue of the operator  $\hat{\pi}$ .

Operating Eq. (1) again by  $\hat{\pi}$  operator, we have

$$\hat{\pi}^2\psi = \hat{\pi}(\lambda\psi) = \lambda(\hat{\pi}\psi) = \lambda(\lambda\psi) = \lambda^2\psi \quad \dots(2)$$

From definition of parity operator

$$\hat{\pi}\psi(x) = \psi(-x)$$

$$\begin{aligned}\hat{\pi}[\hat{\pi}\psi(x)] &= \hat{\pi}\psi(-x) = \psi(x) \\ \hat{\pi}^2\psi(x) &= \psi(x)\end{aligned}\quad \dots(3)$$

Comparing Eqs. (2) and (3), we note that the eigen values of  $\pi$ -operator satisfy the equation

$$\lambda^2 = 1$$

or  $\lambda = \pm 1.$  \dots(4)

The eigenfunctions corresponding to eigenvalue  $\lambda = + 1$  of the parity operator are the even functions  $\psi_e$ , which satisfy

$$\psi_e(x) = \psi_e(-x). \quad \dots(5)$$

The odd functions  $\psi_o$  belong to the eigenvalue  $\lambda = -1$  and satisfy the relation

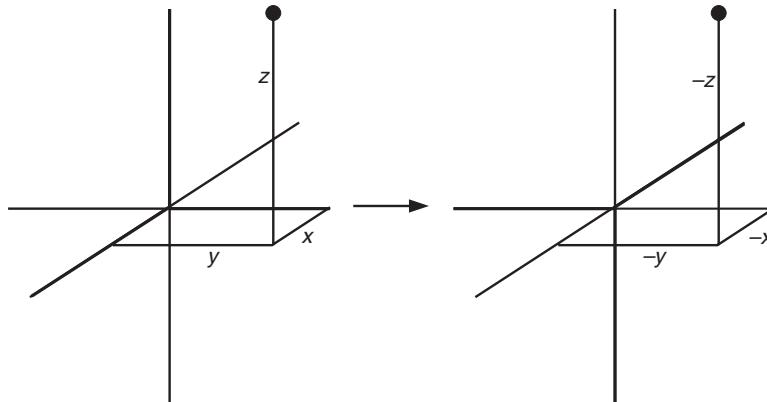
$$\psi_o(x) = -\psi_o(-x). \quad \dots(6)$$

The eigenfunctions of  $\hat{\pi}$  are a complete set with respect to the class of functions of  $x$ . This is shown by the identity

$$\psi(x) = \frac{1}{2}[\psi(x) + \psi(-x)] + \frac{1}{2}[\psi(x) - \psi(-x)] \quad \dots(7)$$

Thus any  $\psi(x)$  is a linear combination of an *even* function of  $x$ ,  $[\psi(x) + \psi(-x)]$ , and an *odd* function of  $x$ ,  $[\psi(x) - \psi(-x)].$

- The parity operation in three dimensions is defined by reflecting all three axes according to the coordinate transformation  $(x, y, z) \rightarrow (-x, -y, -z)$  [Fig. 10.2].



**Fig. 10.2**

The operation is also called *space inversion* as the transformation changes a right-handed system of coordinates into a left handed system.

The Parity operator  $\hat{\pi}$  is defined by the relation

$$\hat{\pi}\psi(\mathbf{r}) = \psi(-\mathbf{r}).$$

Thus the parity operator corresponds to an inversion of the position coordinate  $\mathbf{r}$  through the origin.

#### 10.2.6. Probability Density

Let  $\psi(x, t)$  be a normalised wave function of a particle in a given state.  $\psi(x, t)$  by itself is sometimes called *Probability amplitude* for the position of the particle.

The product  $\psi^*(x, t)\psi(x, t)$  is called *probability density* and is real and hence observable. It

is the probability of finding the particle in unit interval of space centered about  $x$  at time  $t$ .

$$\text{Probability density} = \psi^*(x, t) \psi(x, t) = |\psi(x, t)|^2 \quad \dots(1)$$

A large value of  $|\psi|^2$  means the strong possibility of the particle's presence. A small value of  $|\psi|^2$  means the slight possibility of the particle's presence. This interpretation was made by Max Born.

The probability that the particle may be found in the region between  $x$  and  $(x + dx)$  at time  $t$  is

$$P = \psi^*(x, t) \psi(x, t) dx \quad \dots(2)$$

The probability of finding the particle in a volume element  $d\tau = dx dy dz$  about any point  $\mathbf{r}$  at time  $t$  is expressed as

$$P(\mathbf{r}, t) d\tau = |\psi(\mathbf{r}, t)|^2 d\tau. \quad \dots(3)$$

#### 10.2.7. Probability Current Density

Suppose  $\psi(\mathbf{r}, t)$  is the state function representing a one-particle system.

The probability that the particle is in the finite volume  $\tau$  is

$$P = \int_{\tau} \psi^* \psi d\tau \quad \dots(1)$$

The rate of change of this probability is

$$\frac{dP}{dt} = \int_{\tau} \left( \psi \frac{\partial \psi^*}{\partial t} + \psi^* \frac{\partial \psi}{\partial t} \right) d\tau \quad \dots(2)$$

Time-dependent Schrodinger equation is

$$-\frac{\hbar^2}{2m} \nabla^2 \psi + V \psi = i\hbar \frac{\partial \psi}{\partial t} \quad \dots(3)$$

Assuming  $V$  to be real, complex conjugate of Eq. (3) is

$$-\frac{\hbar^2}{2m} \nabla^2 \psi^* + V \psi^* = -i\hbar \frac{\partial \psi^*}{\partial t} \quad \dots(4)$$

$$\frac{\partial \psi}{\partial t} = \frac{1}{i\hbar} \left[ -\frac{\hbar^2}{2m} \nabla^2 \psi + V \psi \right]; \frac{\partial \psi^*}{\partial t} = -\frac{1}{i\hbar} \left[ -\frac{\hbar^2}{2m} \nabla^2 \psi^* + V \psi^* \right]$$

$$\begin{aligned} \therefore \psi \frac{\partial \psi^*}{\partial t} + \psi^* \frac{\partial \psi}{\partial t} &= -\frac{i\hbar}{2m} [\psi \nabla^2 \psi^* - \psi^* \nabla^2 \psi] \\ &= -\frac{i\hbar}{2m} \nabla \cdot (\psi \nabla \psi^* - \psi^* \nabla \psi) \end{aligned}$$

We define probability current density as

$$\mathbf{j} = \frac{i\hbar}{2m} [\psi \nabla \psi^* - \psi^* \nabla \psi] \quad \dots(5)$$

$$\therefore \int_{\tau} \left( \psi \frac{\partial \psi^*}{\partial t} + \psi^* \frac{\partial \psi}{\partial t} \right) d\tau = - \int_{\tau} \nabla \cdot \mathbf{j} d\tau$$

$$\text{Eq. (2) becomes } \frac{dP}{dt} = - \int_{\tau} \nabla \cdot \mathbf{j} d\tau \quad \dots(6)$$

The volume integral can be transformed to a surface integral by Gauss's theorem, *i.e.*,

$$\frac{dP}{dt} = - \iint_S \mathbf{j} \cdot d\mathbf{S} \quad \dots(7)$$

Here  $S$  is the surface enclosing the volume  $\tau$ . The direction of  $dS$  is along the outward normal. Clearly from Eq. (7), the integral of  $j$  over the surface  $S$  is the probability that the particle will cross the surface going outwards in unit time.

$$\text{From Eqs. (2) and (6), } \frac{\partial}{\partial t}(\psi^* \psi) + \nabla \cdot j = 0 \quad \dots(8)$$

This equation expresses the *conservation of probability density*. It is analogous to the equations of continuity of hydrodynamics and electrodynamics.

### MOTION OF WAVE PACKETS

#### 10.3 WAVE PACKET

A pure sine wave is completely unlocalized—it extends from  $-\infty$  to  $+\infty$ . But a classical particle is completely localized. Our quantum description mixes particles and waves. The particles are approximately, but not completely, localized. An electron, for example, is bound to a specific atom. We know its position to within an uncertainty of the order of the diameter of the atom ( $10^{-10}$  m). But we don't know exactly where it is within that atom. The method used in physics to describe such a situation is that of a *wave packet*.

A particle in motion is represented by a wave that is confined to a small region of space around the particle. Such a wave is called a **wave packet**.

A wave packet is the superposition of a large number of waves. The waves interfere constructively in the vicinity of the particle, giving the resultant wave a large amplitude. The waves interfere destructively far from the particle. So the resultant wave has a small amplitude in regions where we don't expect to find the particle.

Figure 10.3 shows a wave packet propagating along the  $x$ -axis. It consists of “ripples” contained within an “envelope”.

The “envelope” travels at the group velocity ( $v_g$ ).

The individual ripples travel at the phase velocity ( $v_p$ ).

- The *phase velocity* of the wave is the speed with which a single component wave moves through the medium.
- The envelope moves at a speed different from that of the individual waves.

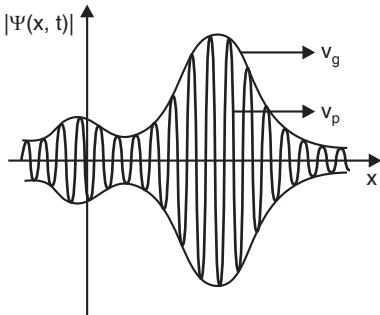


Fig. 10.3

##### 10.3.1. Ehrenfest's Theorem

According to Ehrenfest's theorem, the classical equations

$$m \frac{d\mathbf{r}}{dt} = \mathbf{p} \text{ and } \frac{d\mathbf{p}}{dt} = -\nabla V$$

[ $V(r)$  is the potential energy function] are also valid in quantum mechanics, provided we replace all the classical quantities by the expectation values of their corresponding quantum mechanical operators.

**Statement.** *Ehrenfest theorem states that the average motion of a wave-packet agrees with the motion of the corresponding classical particle.*



Paul Ehrenfest  
(1880 – 1933)

In other words, it states that the Schrodinger wave equation leads to the satisfaction of classical Newton's laws of motion on the average.

### Explanation

- **Ehrenfest's theorem** tells us that *expectation values obey classical laws*.
- Ehrenfest's theorem shows that a wave packet moves like a classical particle whenever the expectation value gives a good representation of the classical variable.

This is usually the macroscopic limit in which the finite size and the internal structure of the packet can be ignored. So Ehrenfest's theorem provides an example of the correspondence principle.

### Proof.

Let  $x$  be the position co-ordinate of a particle of mass  $m$  at time  $t$ .

$$(a) \quad \begin{aligned} \frac{d}{dt} \langle x \rangle &= \frac{d}{dt} \int \psi^* x \psi d^3 r \\ &= \int \psi^* x \frac{\partial \psi}{\partial t} d^3 r + \int \frac{\partial \psi^*}{\partial t} x \psi d^3 r \end{aligned} \quad \dots(1)$$

Time dependent Schrodinger equation is

$$\begin{aligned} i\hbar \frac{\partial \psi}{\partial t} &= -\frac{\hbar^2}{2m} \nabla^2 \psi + V \psi \\ \text{or} \quad \frac{\partial \psi}{\partial t} &= \frac{1}{i\hbar} \left\{ -\frac{\hbar^2}{2m} \nabla^2 \psi + V \psi \right\} \end{aligned}$$

Complex conjugate of time-dependent Schrodinger equation is

$$\begin{aligned} -i\hbar \frac{\partial \psi^*}{\partial t} &= -\frac{\hbar^2}{2m} \nabla^2 \psi^* + V \psi^* \\ \text{or} \quad \frac{\partial \psi^*}{\partial t} &= -\frac{1}{i\hbar} \left\{ -\frac{\hbar^2}{2m} \nabla^2 \psi^* + V \psi^* \right\} \end{aligned}$$

Substituting values of  $\partial \psi / \partial t$  and  $\partial \psi^* / \partial t$  in Eq. (1), we get

$$\begin{aligned} \frac{d}{dt} \langle x \rangle &= \int \psi^* x \left\{ \frac{1}{i\hbar} \left( -\frac{\hbar^2}{2m} \nabla^2 \psi + V \psi \right) \right\} d^3 r \\ &\quad + \int \left\{ -\frac{1}{i\hbar} \left( -\frac{\hbar^2}{2m} \nabla^2 \psi^* + V \psi^* \right) \right\} x \psi d^3 r \\ &= -\frac{\hbar}{2im} \int [\psi^* x (\nabla^2 \psi) - (\nabla^2 \psi^*) x \psi] d^3 r \\ \frac{d}{dt} \langle x \rangle &= -\frac{\hbar}{2im} \int \psi^* x (\nabla^2 \psi) d^3 r + \frac{\hbar}{2im} \int (\nabla^2 \psi^*) x \psi d^3 r \quad \dots(2) \end{aligned}$$

The second integral can be integrated by parts :

$$\int (\nabla^2 \psi^*) x \psi d^3 r = - \int (\nabla \psi^*) \cdot \nabla (x \psi) d^3 r + \int_A (x \psi \nabla \psi^*)_n dA$$

Here the integral of the normal component of  $x \psi \nabla \psi^*$  over the infinite bounding surface  $A$  is zero because a wave packet  $\psi$  vanishes at great distances.

$$\int (\nabla^2 \psi^*) x \psi d^3 r = - \int (\nabla \psi^*) \cdot \nabla (x \psi) d^3 r$$

A second partial integration, in which the surface integral again vanishes, results in

$$\int (\nabla^2 \psi^*) x \psi d^3 r = \int \psi^* \nabla^2 (x \psi) d^3 r \quad \dots(3)$$

Using Eq. (3), Eq. (2) can be written as

$$\begin{aligned} \frac{d}{dt} \langle x \rangle &= -\frac{\hbar}{2im} \int \psi^* x (\nabla^2 \psi) d^3 r + \frac{\hbar}{2im} \int \psi^* \nabla^2 (x \psi) d^3 r \\ &= -\frac{\hbar}{2im} \int \psi^* [x \nabla^2 \psi - \nabla^2 (x \psi)] d^3 r \end{aligned}$$

But  $x \nabla^2 \psi - \nabla^2 (x \psi) = -2 \frac{\partial \psi}{\partial x}$

$$\therefore \frac{d}{dt} \langle x \rangle = \frac{\hbar}{im} \int \psi^* \frac{\partial \psi}{\partial x} d^3 r = \frac{1}{m} \int \psi^* \left( -i \hbar \frac{\partial \psi}{\partial x} \right) d^3 r$$

or  $\frac{d}{dt} \langle x \rangle = \frac{1}{m} \langle p_x \rangle$

Similarly,  $\frac{d}{dt} \langle y \rangle = \frac{1}{m} \langle p_y \rangle$  and  $\frac{d}{dt} \langle z \rangle = \frac{1}{m} \langle p_z \rangle$ .

Combining all components,

$$\frac{d}{dt} \langle \mathbf{r} \rangle = \frac{\langle \mathbf{p} \rangle}{m}. \quad \dots(4)$$

(b) The time rate of change of  $X$ -component of momentum is given by,

$$\begin{aligned} \frac{d}{dt} \langle p_x \rangle &= \frac{d}{dt} \int \psi^* \left( -i \hbar \frac{\partial \psi}{\partial x} d^3 r \right) \\ &= -i \hbar \left( \int \psi^* \frac{\partial}{\partial x} \frac{\partial \psi}{\partial t} d^3 r + \int \frac{\partial \psi^*}{\partial t} \frac{\partial \psi}{\partial x} d^3 r \right) \\ &= - \int \psi^* \frac{\partial}{\partial x} \left( -\frac{\hbar^2}{2m} \nabla^2 \psi + V \psi \right) d^3 r + \int \left( -\frac{\hbar^2}{2m} \nabla^2 \psi^* + V \psi^* \right) \frac{\partial \psi}{\partial x} d^3 r \\ &= - \int \psi^* \left[ \frac{\partial}{\partial x} (V \psi) - V \frac{\partial \psi}{\partial x} \right] d^3 r = - \int \psi^* \frac{\partial V}{\partial x} \psi d^3 r = \left\langle -\frac{\partial V}{\partial x} \right\rangle \end{aligned}$$

Similarly,  $\frac{d}{dt} \langle p_y \rangle = \left\langle -\frac{\partial V}{\partial y} \right\rangle$  and  $\frac{d}{dt} \langle p_z \rangle = \left\langle -\frac{\partial V}{\partial z} \right\rangle$

Combining all components,

$$\frac{d}{dt} \langle \mathbf{p} \rangle = \langle -\nabla V \rangle \quad \dots(5)$$

Equations (4) and (5) are analogous to the classical equations of motion :

$$\frac{d\mathbf{r}}{dt} = \frac{\mathbf{p}}{m} \text{ and } \frac{d\mathbf{p}}{dt} = -\nabla V$$

This proves Ehrenfest's theorem.

### MOTION OF A FREE WAVE PACKET IN ONE DIMENSION

- We find the minimum value of the uncertainty product  $\Delta x \cdot \Delta p$  and the possible forms of the one-dimensional wave packet that correspond to it, all at a definite instant of time.

## 10.4 THE MINIMUM UNCERTAINTY PRODUCT

Consider two observables, position  $x$  and momentum  $p$ .

$\langle x \rangle$  is the expectation value of  $x$ .

$\langle p \rangle$  is the expectation value of  $p$ .

The uncertainty  $\Delta x$  is

$$\Delta x = \langle (x - \langle x \rangle)^2 \rangle^{1/2}$$

The uncertainty  $\Delta p$  is

$$\begin{aligned} \Delta p &= \langle (p - \langle p \rangle)^2 \rangle^{1/2} \\ (\Delta x)^2 &= \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - 2\langle x \rangle \langle x \rangle + \langle \langle x \rangle^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 \\ (\Delta p)^2 &= \langle (p - \langle p \rangle)^2 \rangle = \langle p^2 \rangle - \langle p \rangle^2 \end{aligned} \quad \dots(1)$$

We define two new operators as

$$\alpha \equiv x - \langle x \rangle; \quad \beta = p - \langle p \rangle = -i\hbar \left( \frac{d}{dx} - \left\langle \frac{d}{dx} \right\rangle \right) \quad \dots(2)$$

Then

$$\begin{aligned} (\Delta x)^2 (\Delta p)^2 &= \int_{-\infty}^{\infty} \psi^* \alpha^2 \psi dx \int_{-\infty}^{\infty} \psi^* \beta^2 \psi dx \\ (\Delta x)^2 (\Delta p)^2 &= \int_{-\infty}^{\infty} (\alpha^* \psi^*) (\alpha \psi) dx \int_{-\infty}^{\infty} (\beta^* \psi^*) (\beta \psi) dx \end{aligned} \quad \dots(3)$$

We shall use the *Schwarz inequality*

$$\int |f|^2 dx \int |g|^2 dx \geq \left| \int f^* g dx \right|^2$$

Here,  $f$  and  $g$  are arbitrary functions.

The equality is applicable only if  $f = \gamma g$ , where  $\gamma$  is a constant.

If now we replace  $f$  by  $\alpha \psi$  and  $g$  by  $\beta \psi$ , Eq. (3) becomes

$$(\Delta x)^2 (\Delta p)^2 \geq \left| \int (\alpha^* \psi^*) (\beta \psi) dx \right|^2 = \left| \int \psi^* \alpha \beta \psi dx \right|^2 \quad \dots(4)$$

The last term in Eq. (4) can be written as

$$\begin{aligned} &\left| \int \psi^* \left[ \frac{1}{2}(\alpha \beta - \beta \alpha) + \frac{1}{2}(\alpha \beta + \beta \alpha) \right] \psi dx \right|^2 \\ &= \frac{1}{4} \left| \int \psi^* (\alpha \beta - \beta \alpha) \psi dx \right|^2 + \frac{1}{4} \left| \int \psi^* (\alpha \beta + \beta \alpha) \psi dx \right|^2 \end{aligned} \quad \dots(5)$$

The cross term in the product that is omitted on the right side of Eq. (5) can be seen to vanish when use is made of the relation

$$(\int \psi^* \alpha \beta \psi dx)^* = \int \psi \alpha^* \beta^* \psi^* dx = \int (\beta^* \psi^*) (\alpha \psi) dx = \int \psi^* \beta \alpha \psi dx.$$

Now from Eq. (2),

$$(\alpha \beta - \beta \alpha) \psi = -i\hbar \left[ x \frac{d\psi}{dx} - \frac{d}{dx} (x\psi) \right] = i\hbar \psi \quad \dots(6)$$

We thus obtain from Eqs. (4) to (6),

$$(\Delta x)^2 (\Delta p)^2 \geq \frac{1}{4} \hbar^2$$

or  $\Delta x \cdot \Delta p \geq \frac{1}{2}\hbar$  ... (7)

- The equality can hold only if the second term on the right side of Eq. (5) is zero.
- Eq. (7) is the precise expression of the Heisenberg position-momentum uncertainty relation.

### Form of the Minimum Packet

The minimum uncertainty product is attained only when two conditions are fulfilled:

$$(i) \quad \alpha\psi = \gamma\beta\psi \quad \dots(8)$$

$$(ii) \quad \int \psi^* (\alpha\beta + \beta\alpha) \psi dx = 0 \quad \dots(9)$$

Equations (8) and (2) give us a differential equation for  $\psi$

$$\frac{d\psi}{dx} = \left[ \frac{i}{\gamma\hbar} (x - \langle x \rangle) + \frac{i\langle p \rangle}{\hbar} \right] \psi \quad \dots(10)$$

Integrating Eq. (10), we get

$$\psi(x) = N \exp \left[ \frac{i}{2\gamma\hbar} (x - \langle x \rangle)^2 + \frac{i\langle p \rangle x}{\hbar} \right] \quad \dots(11)$$

Here,  $N$  is an arbitrary constant.

Equation (9), with the help of Eq. (8), becomes

$$\left( \frac{1}{\gamma} + \frac{1}{\gamma^*} \right) \int \psi^* \alpha^2 \psi dx = 0$$

Since the integral is not zero, this yields

$$\left( \frac{1}{\gamma} + \frac{1}{\gamma^*} \right) = 0$$

which requires that  $\gamma$  be pure imaginary.

We want Eq. (11) to represent a wave packet for which the integral of  $|\psi|^2$  should converge.

So  $\gamma$  must be negative imaginary.

- The magnitude of the constant  $N$  is fixed by normalizing  $\psi$ .

$$\int |\psi|^2 dx = 1$$

- $\gamma$  is determined by requiring that

$$\int (x - \langle x \rangle)^2 |\psi|^2 dx = (\Delta x)^2.$$

The integrals are evaluated and lead to the normalized minimum wave packet.

$$\psi(x) = [2\pi(\Delta x)^2]^{-1/4} \exp \left[ \frac{(x - \langle x \rangle)^2}{4(\Delta x)^2} + \frac{i\langle p \rangle x}{\hbar} \right] \quad \dots(12)$$

Eq. (12) is a gaussian function.

- *The wave packet having the minimum uncertainty product has a gaussian shape.*
- Width of the wave-packet  $\Delta x$  is given in terms of standard deviation  $\sigma$  as

$$(2\Delta x)^2 = 2\sigma^2$$

$$\therefore \sigma = \sqrt{2}\Delta x.$$

Figure 10.4 is a graph of the function  $|\psi(x)|$ .

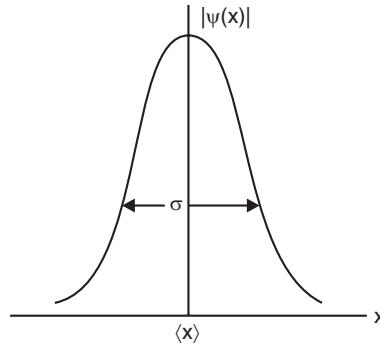


Fig. 10.4

- The quantity  $\sigma$  measures the width of the packet.
- This wave packet moves with uniform average momentum  $\langle p \rangle$ .
- At time  $t = 0$ , the center of the packet is located at  $x = \langle x \rangle$ ,

### ENERGY EIGENFUNCTIONS

## 10.5 ORTHOGONALITY OF EIGENFUNCTIONS

We shall first prove that *all energy eigenvalues  $E_n$  are real and that if  $E_n \neq E_m$  the corresponding eigenfunctions are necessarily orthogonal.*

### (i) To show that the Energy Eigen Values are Real

Consider a particle of mass  $m$  moving in one dimension in a potential  $V(x)$ . The one dimensional Schrodinger wave equation for the state ( $n$ ) corresponding to the energy eigenvalue  $E_n$  is

$$\frac{-\hbar^2}{2m} \frac{d^2 \psi_n}{dx^2} + V(x) \psi_n = E_n \psi_n \quad \dots(1)$$

Similarly, for the state ( $m$ ) corresponding to the energy eigen value  $E_m$ ,

$$\frac{-\hbar^2}{2m} \frac{d^2 \psi_m}{dx^2} + V(x) \psi_m = E_m \psi_m \quad \dots(2)$$

The complex conjugate of Eq. (2) is

$$\frac{-\hbar^2}{2m} \frac{d^2 \psi_m^*}{dx^2} + V(x) \psi_m^* = E_m^* \psi_m^* \quad \dots(3)$$

Multiplying Eq. (1) by  $\psi_m^*$  and Eq. (3) by  $\psi_n$ , we have

$$\frac{-\hbar^2}{2m} \psi_m^* \frac{d^2 \psi_n}{dx^2} + \psi_m^* V(x) \psi_n = E_n \psi_m^* \psi_n \quad \dots(4)$$

$$\frac{-\hbar^2}{2m} \psi_n \frac{d^2 \psi_m^*}{dx^2} + \psi_n V(x) \psi_m^* = E_m^* \psi_m^* \psi_n \quad \dots(5)$$

Subtracting Eq. (5) from Eq. (4), we have

$$\frac{-\hbar^2}{2m} \left[ \psi_m^* \frac{d^2 \psi_n}{dx^2} - \psi_n \frac{d^2 \psi_m^*}{dx^2} \right] = (E_n - E_m^*) \psi_m^* \psi_n$$

$$\frac{-\hbar^2}{2m} \left[ \frac{d}{dx} \left( \psi_m^* \frac{d\psi_n}{dx} - \psi_n \frac{d\psi_m^*}{dx} \right) \right] = (E_n - E_m^*) \psi_m^* \psi_n \quad \dots(6)$$

Integrating Eq. (6) between the limits  $-\infty$  to  $+\infty$ , we get

$$\begin{aligned} \frac{-\hbar^2}{2m} \int_{-\infty}^{+\infty} \frac{d}{dx} \left( \psi_m^* \frac{d\psi_n}{dx} - \psi_n \frac{d\psi_m^*}{dx} \right) dx &= (E_n - E_m^*) \int_{-\infty}^{+\infty} \psi_m^* \psi_n dx \\ \frac{-\hbar^2}{2m} \left[ \left( \psi_m^* \frac{d\psi_n}{dx} - \psi_n \frac{d\psi_m^*}{dx} \right) \right]_{-\infty}^{+\infty} &= (E_n - E_m^*) \int_{-\infty}^{+\infty} \psi_m^* \psi_n dx \end{aligned} \quad \dots(7)$$

Since  $\psi_m$  and  $\psi_n$  vanish at infinity, the R.H.S. vanishes.

$$\therefore (E_n - E_m^*) \int_{-\infty}^{\infty} \psi_m^* \psi_n dx = 0 \quad \dots(8)$$

$$\text{Put } m = n. \text{ Then, } (E_n - E_n^*) \int_{-\infty}^{\infty} \psi_n^* \psi_n dx = 0$$

$$(E_n - E_n^*) \int_{-\infty}^{\infty} |\psi_n|^2 dx = 0$$

The integral is a positive quantity.

$$\begin{aligned} \therefore (E_n - E_n^*) &= 0 \\ \therefore E_n &= E_n^* \end{aligned}$$

This proves that all energy eigen values are real.

- (ii) To show that any two eigen functions corresponding to different energy eigen values are orthogonal to each other.**

Let  $\psi_m$  and  $\psi_n$  correspond to the energies  $E_m$  and  $E_n$ . Then, from Eq. (8), we have

$$\begin{aligned} (E_n - E_m^*) \int_{-\infty}^{\infty} \psi_m^* \psi_n dx &= 0 \\ \text{If } E_m \neq E_n, \quad \int_{-\infty}^{\infty} \psi_m^* \psi_n dx &= 0 \end{aligned} \quad \dots(9)$$

This represents the *orthogonality condition*. If two functions  $\psi_n$  and  $\psi_m$  have this property, they are *orthogonal* to each other.

### Orthonormality condition

Since the Schrodinger equation is a linear equation, if  $\psi$  is a solution,  $c\psi$  is also a solution where  $c$  is an arbitrary complex number.

We can always choose the constant such that

$$\int_{-\infty}^{\infty} |\psi_n|^2 dx = 1 \quad \dots(10)$$

This is called the normalization condition.

Equations (9) and (10) can be combined in the form

$$\int_{-\infty}^{\infty} \psi_m^* \psi_n dx = \delta_{mn} \quad \dots(11)$$

Here,  $\delta_{mn}$  is the Kronecker delta function defined through the equation

$$\delta_{mn} = \begin{cases} 0 & m \neq n \\ 1 & m = n \end{cases}$$

Equation (11) is called the **orthonormality condition**.

**Orthonormal function.** The functions which are *orthogonal* and also *normalised* are called orthonormal functions.

## 10.6 DIRAC DELTA FUNCTION

**Definition.** The Dirac  $\delta$ -function is defined such that

$$\int_{-\infty}^{\infty} \delta(x-x') dx' = 1 \quad \dots(1) \quad \text{and} \quad \delta(x-x') = \begin{cases} 0 & (\text{for } x-x' \neq 0) \\ \infty & (\text{for } x-x' = 0) \end{cases} \quad \dots(2)$$

The properties of the  $\delta$ -function listed in Eqs. (1) and (2) make it clear that the Dirac  $\delta$ -function is not a function in a rigorous mathematical sense since the integral (if it exists) of a function that is zero everywhere except at one point must vanish.

The *Dirac delta-function* is used in mathematical physics wherever exist functions with non-zero values in very short interval.

**EXAMPLE.** Consider the position observable  $x$ . The quantum mechanical operator for position is  $\hat{x}$ . Consider the eigen value equation with this operator  $\hat{x}$ , operating on the eigen function  $\psi_0$

$$\hat{x}\psi_0 = x'\psi_0$$

$\psi_0$  is the eigen function and it corresponds to a state, in which the particle is located precisely at the point  $x = x'$ . In other words,  $\psi_0$  must be *infinitely concentrated* at  $x = x'$  (Fig. 10.5).

### Properties of $\delta$ -Function

1.  $\delta(x) = \delta(-x)$
2.  $f(x)\delta(x-a) = f(a)\delta(x-a)$
3.  $\int f(x)\delta(x-a) dx = f(a)$
4.  $\delta(ax) = \frac{1}{|a|}\delta(x)$  for  $a > 0$
5.  $x\delta(x) = 0$

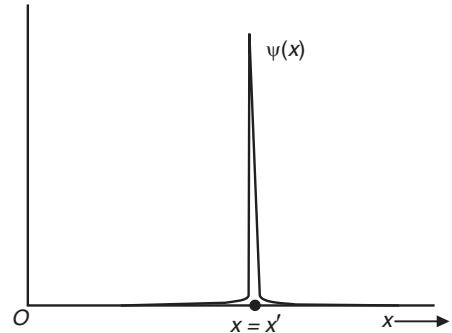


Fig. 10.5

### 10.6.1. The completeness condition in terms of the Dirac delta function

Let  $\psi_n(x)$  form a complete set of orthonormal functions in the domain  $-\infty < x < \infty$ .

The orthonormality condition would require

$$\int_{-\infty}^{\infty} \psi_m^*(x) \psi_n(x) dx = \delta_{mn} \quad \dots(1)$$

Here,  $\delta_{mn}$  is the Kronecker delta function defined through the following equation:

$$\delta_{mn} = \begin{cases} 0 & \text{for } m \neq n \\ 1 & \text{for } m = n \end{cases} \quad \dots(2)$$

Since  $\psi_n(x)$  form a complete set of functions, we can expand an arbitrary “well-behaved” function  $\phi(x)$  in terms of  $\psi_n(x)$ :

$$\phi(x) = \sum_{n=0,1,2,\dots}^{\infty} c_n \psi_n(x) \quad \dots(3)$$

In order to determine  $c_n$ , we multiply Eq. (3) by  $\psi_m^*(x)$  and integrate

$$\begin{aligned} \int_{-\infty}^{\infty} \psi_m^*(x) \phi(x) dx &= \sum_{n=0,1,2,\dots}^{\infty} c_n \int_{-\infty}^{\infty} \psi_m^*(x) \psi_n(x) dx = \sum_{n=0,1,2,\dots}^{\infty} c_n \delta_{mn} \\ &= c_m \\ \therefore c_n &= \int_{-\infty}^{\infty} \psi_n^*(x) \phi(x) dx \end{aligned} \quad \dots(4)$$

Substituting in Eq. (3), we get

$$\phi(x) = \sum_{n=0,1,2,\dots}^{\infty} \left[ \int_{-\infty}^{\infty} \psi_n^*(x') \phi(x') dx' \right] \psi_n(x)$$

Here, we have put a prime on the integration variable so that it does not get confused with the variable of  $\psi_n(x)$ .

Carrying out the summation first, we get

$$\phi(x) = \int_{-\infty}^{\infty} \phi(x') dx' \sum_{n=0,1,2,\dots}^{\infty} \psi_n^*(x') \psi_n(x)$$

Comparing with Eq. (3), we may write

$$\sum_{n=0,1,2,\dots}^{\infty} \psi_n^*(x') \psi_n(x) = \delta(x - x') \quad \dots(5)$$

Eq. (5) represents the **completeness condition**.

## 10.7 SYSTEM OF TWO IDENTICAL PARTICLES

*Identical particles are those particles in a system for which there will be no change in system by interchanging the particles.*

For a system of 2 particles,

$$|\psi|^2(1, 2) = |\psi|^2(2, 1) \quad \dots(1)$$

Wave functions that are unaffected by an exchange of particles are said to be *symmetric*.

$$\psi(2, 1) = \psi(1, 2) \quad \dots(2)$$

Wave functions that reverse sign by an exchange of particles are said to be *antisymmetric*.

$$\psi(2, 1) = -\psi(1, 2) \quad \dots(3)$$

Let particle 1 be in state  $a$  and particle 2 be in state  $b$ . Then, the wave function of the system is,

$$\psi_I = \psi_a(1) \psi_b(2) \quad \dots(4)$$

Let particle 2 be in state  $a$  and particle 1 be in state  $b$ . Then, the wave function is

$$\psi_{II} = \psi_a(2) \psi_b(1) \quad \dots(5)$$

A linear combination of  $\psi_I$  and  $\psi_{II}$  is the proper description of the system.

The symmetric normalized wave function can be written as

$$\psi_s = \frac{1}{\sqrt{2}}[\psi_a(1)\psi_b(2) + \psi_a(2)\psi_b(1)] \quad \dots(6)$$

If the wave function of the particles is symmetric, then the particles are said to obey the *Bose Einstein statistics*. Particles of 0 or integral spin have wave functions that are symmetric to an exchange of any pair of them. These particles, which include photons, alpha particles, and helium atoms, do not obey the exclusion principle. This type of particles are known as *bosons*.

The antisymmetric normalized wave function of the system is

$$\psi_A = \frac{1}{\sqrt{2}}[\psi_a(1)\psi_b(2) - \psi_a(2)\psi_b(1)] \quad \dots(7)$$

Now, the particles are said to obey the *Fermi-Dirac statistics*. They are called *fermions*. Protons, neutrons and electrons are fermions. They have odd half-integral spin. They obey the Pauli exclusion principle.

In Eq. (7) put  $a = b$ . Then,

$$\psi_A = \frac{1}{\sqrt{2}}[\psi_a(1)\psi_a(2) - \psi_a(2)\psi_a(1)] = 0 \quad \dots(8)$$

Hence the two particles *cannot* be in the same quantum state. Pauli found that no two electrons in an atom can be in the same quantum state.

## 10.8 LINEAR HARMONIC OSCILLATOR (OPERATOR METHOD)

### (a) The Ladder (or Raising and Lowering) Operators:

The Hamiltonian of a linear harmonic oscillator is

$$\hat{H} = \frac{1}{2m}\hat{p}^2 + \frac{1}{2}m\omega_c^2\hat{x}^2 \quad \dots(1)$$

$\omega_c$  is the angular frequency of the harmonic oscillator.

$\hat{x}$  and  $\hat{p}$  are the coordinate and momentum operators. The quantum condition is

$$[\hat{x}, \hat{p}] = i\hbar \quad \dots(2)$$

Let us now introduce a pair of operators,

$$a = \left(\frac{m\omega_c}{2\hbar}\right)^{1/2}x + i\left(\frac{1}{2m\hbar\omega_c}\right)^{1/2}p \quad \dots(3)$$

$$a^\dagger = \left(\frac{m\omega_c}{2\hbar}\right)^{1/2}x - i\left(\frac{1}{2m\hbar\omega_c}\right)^{1/2}p \quad \dots(4)$$

$$aa^\dagger = \frac{H}{\hbar\omega_c} + \frac{1}{2} \text{ and } a^\dagger a = \frac{H}{\hbar\omega_c} - \frac{1}{2} \quad \dots(5)$$

$a$  and  $a^\dagger$  satisfy the commutation relation

$$[a, a^\dagger] = aa^\dagger - a^\dagger a = 1 \quad \dots(6)$$

$$H = \left(a^\dagger a + \frac{1}{2}\right)\hbar\omega_c \quad \dots(7)$$

Since the relation between  $H$  and  $a^\dagger a$  involves only ordinary numbers (no operators), their eigenvalues will also bear the same relation, and their eigenfunctions coincide. We shall now see that the eigenvalues of  $a^\dagger a$  are integers  $n(n = 0, 1, 2, \dots)$ . So the eigenvalues of the Hamiltonian are

$$\left(a^\dagger a + \frac{1}{2}\right)\hbar\omega_c.$$

$$[a^\dagger a, a] = (a^\dagger a)a - a(a^\dagger a) = (a^\dagger a - aa^\dagger)a = -a \quad \dots(8)$$

Similarly  $[a^\dagger a, a^\dagger] = a^\dagger$  ... (9)

Let  $u$  be an eigenfunction of  $(a^\dagger a)$ , belonging to the eigenvalue  $\lambda$ , i.e.,

$$(a^\dagger a)u = \lambda u \quad \dots(10)$$

Using Eqs. (6) and (10), we have

$$(a^\dagger a)au = (aa^\dagger - 1)au = a(a^\dagger a - 1)u = a(\lambda - 1)u = (\lambda - 1)au \quad \dots(11)$$

Similarly,  $(a^\dagger a)a^\dagger u = (\lambda + 1)a^\dagger u \quad \dots(12)$

Eqs. (11) and (12) state that  $au$  and  $a^\dagger u$  are also eigenfunctions of  $(a^\dagger a)$ , belonging to the eigenvalues  $(\lambda - 1)$  and  $(\lambda + 1)$  respectively.

Starting with  $u$ , we can now construct wave functions  $a^\dagger u$ ,  $(a^\dagger)^2 u$ , ... belonging to eigenvalues  $(\lambda + 1)$ ,  $(\lambda + 2)$ , ... and also  $au$ ,  $a^2 u$ , ... belonging to  $(\lambda - 1)$ ,  $(\lambda - 2)$ , ... The eigenvalues thus form a 'ladder' with unit spacing between steps.

$a^\dagger$  has the effect of raising the eigenvalue by one step.

$a$  has the effect of lowering the eigenvalue by one step.

For this reason,  $a^\dagger$  and  $a$  are called ladder operators, or *raising* and *lowering operators*.

(b) *The Eigenvalue Spectrum:*  $a^\dagger a$  is a positive operator. Therefore, it can have no negative eigenvalues. So the sequence of eigenvalues  $(\lambda - 1)$ ,  $(\lambda - 2)$ , ... must terminate before it reaches negative values, and correspondingly, the sequence  $au$ ,  $a^2 u$ , ... must terminate. Let us call the last eigenfunction in the sequence as  $u_0$ . Then, by definition, there does not exist a state  $au_0$ . In other words,

$$au_0 = 0 \quad \dots(13)$$

This implies obviously that

$$a^\dagger a u_0 = 0 \quad \dots(14)$$

i.e.,  $u_0$  is an eigenstate of  $(a^\dagger a)$ , belonging to the eigenvalue 0. The other eigenstates can now be constructed by repeated application of  $a^\dagger$  to  $u_0$ . Since the effect of  $a^\dagger$  applied once is to increase the eigenvalue by unity, we have that the eigenstates  $u_0$ ,  $(a^\dagger)u_0$ ,  $(a^\dagger)^2 u_0$ , ... belong to the eigenvalues, 0, 1, 2, ... of  $(a^\dagger a)$ .

Thus the eigenvalue spectrum of  $(a^\dagger a)$  consists of the set of nonnegative integers  $n$ .

Hence  $a^\dagger a$  is called the *number operator*.

The eigenvalues of  $H$  are  $\left(n + \frac{1}{2}\right)\hbar\omega_c \quad n = 0, 1, 2, \dots$

(c) *The Energy Eigenfunctions:* The eigenstates which we have constructed above are mutually orthogonal. The norm of  $(a^\dagger)^n u_0$  is  $n!$

$$((a^\dagger)^m u_0, (a^\dagger)^n u_0) = n! \delta_{mn} \quad \dots(15)$$

Therefore, the normalized eigenfunctions are given by

$$u_n = \frac{(a^\dagger)^n}{\sqrt{n!}} u_0, \quad n = 0, 1, 2, \dots \quad \dots(16)$$

By construction, they satisfy the relation

$$(a^\dagger a)u_n = nu_n \quad \dots(17)$$

Observe that Eq. (16) can be rewritten as

$$u_n = \frac{n^\dagger}{\sqrt{n}} \frac{(a^\dagger)^{n-1}}{\sqrt{(n-1)!}} u_0 = \frac{a^\dagger}{\sqrt{n}} u_{n-1}$$

$$\therefore a^\dagger u_{n-1} = \sqrt{n} u_n \quad \dots(18)$$

Further, by operating on this equation with  $a$ , we get

$$\begin{aligned}\sqrt{n} au_n &= aa^\dagger u_{n-1} = (a^\dagger a + 1)u_{n-1} = nu_{n-1} \\ \therefore au_n &= \sqrt{n} u_{n-1}\end{aligned}\quad \dots(19)$$

From Eqs. (18) and (19), we get

$$(u_m, a^\dagger u_n) = \sqrt{n+1} \delta_{m,n+1}; (u_m, au_n) = \sqrt{n} \delta_{m,n-1} \quad \dots(20)$$

Eqs. (18), (19) and (20) enable us to calculate the effect of any function of  $x$  and  $p$  on the energy eigenfunctions by expressing them first in terms of  $a$  and  $a^\dagger$  through Eqs. (3) and (4).

#### **Determination of the actual form of the wave functions (coordinate-representation)**

Consider Eq. (13). On substituting Eq. (3) in it and writing  $p = -i\hbar d/dx$ , we get the differential equation

$$\left( \frac{d}{dx} + \frac{m\omega_c}{\hbar} x \right) u_0 = 0 \quad \dots(21)$$

The solution of this equation, with proper choice of normalization constant, is

$$u_0(x) = (\alpha^2/\pi)^{1/4} \cdot e^{-\frac{1}{2}\alpha^2 x^2}, \text{ with } \alpha = (m\omega_c/\hbar)^{1/2} \quad \dots(22)$$

All other eigenfunctions are obtained from this by using Eq. (16).

Observing that  $a^\dagger$  can be written in the form

$$a^\dagger = -\frac{1}{\sqrt{2}} \left( \frac{d}{d\rho} - \rho \right) = -\frac{1}{\sqrt{2}} e^{\frac{1}{2}\rho^2} \frac{d}{d\rho} e^{-\frac{1}{2}\rho^2}, \text{ with } \rho = \alpha x,$$

and using the fact that the Hermite polynomials can be expressed as

$$H_n(\rho) = (-)^n e^{\rho^2} \frac{d^n}{d\rho^n} e^{-\rho^2} = (-)^n e^{\frac{1}{2}\rho^2} \left( e^{\frac{1}{2}\rho^2} \frac{d}{d\rho} e^{-\frac{1}{2}\rho^2} \right)^n e^{-\frac{1}{2}\rho^2} \quad \dots(23)$$

We get

$$u_n(x) = N_n e^{-\frac{1}{2}\rho^2} H_n(\rho) = N_n e^{-\frac{1}{2}\alpha^2 x^2} \cdot H_n(\alpha x), n = 0, 1, 2, \dots, \quad \dots(24)$$

$$\text{The normalization constant } N_n = \left( \frac{\alpha}{\sqrt{\pi} 2^n n!} \right)^{1/2} \quad \dots(25)$$

## 10.9 HILBERT SPACE

The vector space, suitable for quantum mechanics is called the *Hilbert* space.

It is a linear vector space of '*infinite*' dimensions.

It is, in general, complex, such that all infinite series occurring in it are *convergent*.

Consider a wavefunction  $\psi$  in Hilber space.

In an  $N$ -D space, we may choose a set of  $N$  linearly independent vectors (functions), say,  $\psi_1, \psi_2, \dots, \psi_N$ . Then,  $\psi$  in this space can be expressed in terms of these functions.

$$\begin{aligned}\psi &= c_1 \psi_1 + c_2 \psi_2 + \dots + c_n \psi_n \\ \psi &= \sum_{i=1}^N c_i \psi_i\end{aligned}$$

The vectors  $\psi_i$ 's, are called basis vectors (functions).

When  $N \rightarrow \infty$ , the definition is in Hilbert space.

The coefficients  $c_i$ 's are, in general, complex and are called the components of the vector  $\psi$ . The  $c_i$ 's are given by

$$c_i = (\vec{\psi}_i \cdot \vec{\psi}) = \int \psi_i^*(x) \psi(x) dx$$

The wavefunction  $\psi(x)$  is completely specified if all the  $c_i$ 's are known. We can write these in a column vector as

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_i \\ \vdots \end{bmatrix}$$

Thus a state function can be represented by a column vector in an infinite dimensional space by imagining an axis for each function  $\psi_i$ , and  $c_i$  corresponds to the component of the vector in the direction of that axis. This infinite dimensional linear vector space spanned by the basis vectors  $\vec{\psi}_i$ 's ( $i = 1, 2, 3, \dots$ ) is called a *Hilbert space*. As the state functions in the Hilbert space are represented by column vectors, we may specifically call them as *state vectors*.

## 10.10 DIRAC'S BRA AND KET NOTATION

In this section, we introduce the *Dirac notation*. It is a very convenient and compact notation for the *scalar product* of two functions.

The scalar product of two square integrable functions  $\psi_1(\mathbf{r})$  and  $\psi_2(\mathbf{r})$  is denoted by the symbol  $\langle \psi_1 | \psi_2 \rangle$ .

$$\langle \psi_1 | \psi_2 \rangle \equiv \int \psi_1^*(\mathbf{r}) \psi_2(\mathbf{r}) d\mathbf{r} \quad \dots(1)$$

The symbol  $|\psi_2\rangle$  is known as a *ket*. The symbol  $\langle\psi_1|$  is known as a *bra*.

From the definition (1), we have

$$\langle \psi_1 | \psi_2 \rangle = \langle \psi_2 | \psi_1 \rangle^*. \quad \dots(2a)$$

If  $c$  is a complex number and  $\psi_3$  a third function, we also have

$$\langle \psi_1 | c\psi_2 \rangle = c \langle \psi_1 | \psi_2 \rangle \quad \dots(2b)$$

$$\langle c\psi_1 | \psi_2 \rangle = c^* \langle \psi_1 | \psi_2 \rangle \quad \dots(2c)$$

$$\langle \psi_3 | \psi_1 + \psi_2 \rangle = \langle \psi_3 | \psi_1 \rangle + \langle \psi_3 | \psi_2 \rangle. \quad \dots(2d)$$

- Two functions  $\psi_1$  and  $\psi_2$  are said to be *orthogonal* if their scalar product vanishes:

$$\langle \psi_1 | \psi_2 \rangle = 0. \quad \dots(3)$$

- The *normalisation condition* is written compactly as

$$\langle \psi | \psi \rangle = 1. \quad \dots(4)$$

- The *orthonormality relation* is expressed as  $\langle \psi_n | \psi_m \rangle = \delta_{nm} \quad n, m = 1, 2, 3, \dots$

- In the Schrodinger representation  $|\psi\rangle \rightarrow \psi(x)$ ,  $\langle\psi| \rightarrow \psi^*(x)$ .

**State Vectors.** Dirac introduced the symbol  $|\psi\rangle$  to denote an abstract state vector as distinct from its representation, the wave function.

Thus, the state vector corresponding to  $\psi_a(\mathbf{r})$  is denoted by the *ket*  $|a\rangle$ .

The state vectors constitute a complex vector space, *i.e.*,

If  $|\psi\rangle, |\chi\rangle$  are state vectors, so is  $c|\psi\rangle + c'|\chi\rangle$  for any complex numbers  $c$  and  $c'$ .

All vectors  $c|\psi\rangle, c \neq 0$ , which have the same ‘direction’ as  $|\psi\rangle$  correspond to the *same* physical state. The set of all such vectors constitute what is called a *ray* in the space. Therefore, we say that any physical state is represented by a ray (rather than a single vector) in the vector space.

### Hilbert Space

- **Norm and Scalar Product:** The squared length or norm of the vector  $|\psi\rangle$  is denoted by  $\langle\psi|\psi\rangle$ . By definition,  $\langle\psi|\psi\rangle$  is a real non-negative number :

$$\langle\psi|\psi\rangle \geq 0, \langle\psi|\psi\rangle = 0 \text{ if and only if } |\psi\rangle = 0. \quad \dots(1a)$$

- $\langle\phi|\psi\rangle = \langle\psi|\phi\rangle^*$   $\dots(1b)$

- The scalar product of two vectors is linear in the second vector :

$$\langle\phi|(c|\psi\rangle + c'|\chi\rangle) = c\langle\phi|\psi\rangle + c'\langle\phi|\chi\rangle \quad \dots(1c)$$

Any vector space (in general infinite dimensional) within which a scalar product with the properties (1) is defined between any two vectors is called a *Hilbert Space*.

The space of ket vectors (or bra vectors) representing quantum mechanical states is thus a Hilbert space.

**Basis in Hilbert Space:** Let  $|\phi_1\rangle, |\phi_2\rangle, \dots$  be a linearly independent set of states in the Hilbert space, *i.e.*, such that no linear combination of these vectors vanishes except when all the coefficients vanish. Let the set be also a *complete* one, in the sense that any arbitrary vector  $|\psi\rangle$  of the space can be expressed in the form

$$|\psi\rangle = c_1|\phi_1\rangle + c_2|\phi_2\rangle + \dots = \sum_i c_i|\phi_i\rangle \quad \dots(3)$$

Then the set  $|\phi_1\rangle, |\phi_2\rangle, \dots$  can be chosen as a *basis* or ‘coordinate system’ for the Hilbert space. The set of numbers  $c_1, c_2, \dots$  in Eq. (3) are the ‘components’ of  $|\psi\rangle$  with respect to this basis, and it provides a *representation of*  $|\psi\rangle$ .

### Properties of Bra and Ket Notation

- (1) Operation on a ket vector from the left with an operator  $\hat{A}$  produces another ket vector:

$$\hat{A}|\psi\rangle = |\psi'\rangle \quad \dots(1)$$

The operation on a bra vector from the right with an operator  $\hat{A}$  produces another bra vector:

$$\langle\psi|\hat{A} = \langle\psi'|. \quad \dots(2)$$

- (2) The kets may be multiplied by complex numbers and may be added to give other kets *i.e.*,

$$a_1|Q\rangle + a_2|R\rangle = |S\rangle \quad \dots(3)$$

Here,  $a_1$  and  $a_2$  are two complex numbers.

- (3) The expectation value of an operator  $\hat{A}$  in the state  $\psi$  is

$$\begin{aligned}\langle\hat{A}\rangle &= \int \psi^* \hat{A} \psi d\tau = \int \psi^* \psi' d\tau = \langle\psi|\psi'\rangle = \langle\psi|A|\psi\rangle \\ \langle\hat{A}\rangle &= \langle\psi|A|\psi\rangle\end{aligned}$$

- (4) Now we consider the eigenvalue problem for quantum mechanical state.

If operator  $\hat{A}$  operates on a ket  $|\psi\rangle$  from left,

$$\hat{A}|\psi\rangle = \lambda|\psi\rangle.$$

Then, the ket  $|\psi\rangle$  is called an eigenket of the operator  $\hat{A}$  and  $\lambda$  is called its eigenvalue.

- (5) If we have a complete set of eigenkets  $|\psi_i\rangle$ 's, ( $i = 1, 2, 3, \dots$ ), then in analogy with the complete set of eigen-functions, we can express any arbitrary ket  $|\psi\rangle$  as

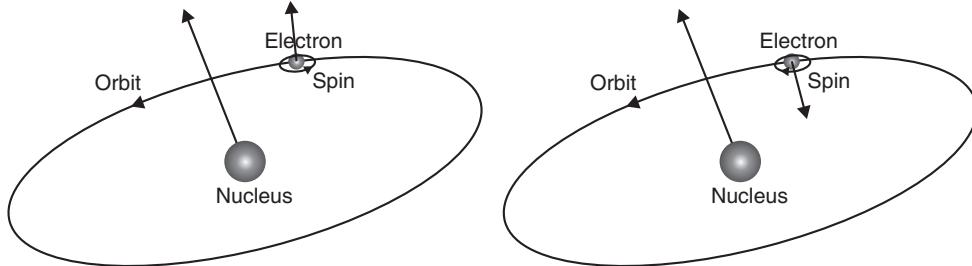
$$|\psi\rangle = \sum_{i=1}^{\infty} c_i |\psi_i\rangle \quad \text{where } c_i = \langle\psi_i|\psi\rangle.$$

## 10.11 TOTAL ANGULAR MOMENTUM

The total angular momentum of a particle is the vector sum of the orbital angular momentum  $\mathbf{L}$  and spin angular momentum  $\mathbf{S}$  i.e.,

Total angular momentum,  $\mathbf{J} = \mathbf{L} + \mathbf{S}$ .

[Fig. 10.6]



**Fig. 10.6**

The total angular momentum is denoted by  $\mathbf{J} = (J_x, J_y, J_z)$ .

**Definition.** The total angular momentum is defined as the generalized angular momentum operator  $J$  as any Hermitian operator whose components satisfy the commutation rules,

$$[J_x, J_y] = i\hbar J_z; [J_y, J_z] = i\hbar J_x; [J_z, J_x] = i\hbar J_y \quad \dots(1)$$

The symbol  $\mathbf{J}$  is introduced here for a general angular momentum vector.

The above three equations can equivalently be written as

$$\mathbf{J} \times \mathbf{J} = i\hbar J \quad \dots(2)$$

$$[J_x, J_x] = [J_y, J_y] = [J_z, J_z] = 0.$$

### 10.11.1. Commutation Relations of the Square of the Total Angular Momentum Operator with components

The square of the angular momentum operator,

$$\mathbf{J}^2 = J_x^2 + J_y^2 + J_z^2$$

commutes with  $J_x, J_y$  and  $J_z$ , namely

$$[\mathbf{J}^2, J_x] = [\mathbf{J}^2, J_y] = [\mathbf{J}^2, J_z] = 0.$$

**Proof:** The total angular momentum is defined by the relation

$$J^2 = J_x^2 + J_y^2 + J_z^2 \quad \dots(1)$$

Let us derive the commutation relation of  $J^2$  with components  $J_x, J_y, J_z$ . Let us take

$$[J^2, J_x] = [J_x^2 + J_y^2 + J_z^2, J_x] = [J_x^2, J_x] + [J_y^2, J_x] + [J_z^2, J_x] \quad \dots(2)$$

We know that

$$[ab, c] = a[b, c] + [a, c]b,$$

So

$$\begin{aligned} [J_x^2, J_x] &= [J_x J_x, J_x] \\ &= J_x [J_x, J_x] + [J_x, J_x] J_x = 0 \\ \therefore [J^2, J_x] &= [J_x^2, J_x] + [J_y^2, J_x] \\ &= [J_y J_y, J_x] + [J_z J_z, J_x] \end{aligned}$$

$$\begin{aligned}
&= J_y[J_y, J_x] + [J_y, J_x]J_y + J_z[J_z, J_x] + [J_z, J_x]J_z \\
&= J_y(-i\hbar J_z) + (-i\hbar J_z)J_y + J_z(i\hbar J_y) + (i\hbar J_y)J_z \\
&= 0.
\end{aligned}$$

∴  $[J^2, J_x] = 0$  ... (3)

Similarly,

$$[J^2, J_y] = 0. \quad \dots(4)$$

$$[J^2, J_z] = 0. \quad \dots(5)$$

Thus the operator  $J^2 = J_x^2 + J_y^2 + J_z^2$  commutes with  $J_x, J_y$  and  $J_z$ .

### 10.11.2. Commutation rule for the Components of Generalised Angular Momentum Operator

We know that

$$\begin{aligned}
[J_x, J_y] &= J_x J_y - J_y J_x \\
\mathbf{J} &= \mathbf{r} \times \mathbf{p} = \mathbf{r} \times (-i\hbar \nabla) = -i\hbar \mathbf{r} \times \nabla
\end{aligned}$$

$$\mathbf{J} = -i\hbar \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x & y & z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \end{vmatrix}$$

$$\begin{aligned}
(i) \quad J_x J_y &= (-i\hbar) \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) (-i\hbar) \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right) \\
&= -\hbar^2 \left\{ y \frac{\partial}{\partial z} \left( z \frac{\partial}{\partial x} \right) - y \frac{\partial}{\partial z} \left( x \frac{\partial}{\partial z} \right) - z \frac{\partial}{\partial y} \left( z \frac{\partial}{\partial x} \right) + z \frac{\partial}{\partial z} \left( x \frac{\partial}{\partial z} \right) \right\} \\
J_x J_y &= -\hbar^2 \left\{ y \frac{\partial}{\partial x} + yz \frac{\partial^2}{\partial z \partial x} - yx \frac{\partial^2}{\partial z^2} - z^2 \frac{\partial^2}{\partial y \partial x} + zx \frac{\partial^2}{\partial y \partial z} \right\} \quad \dots(1)
\end{aligned}$$

$$\begin{aligned}
(ii) \quad J_y J_x &= (-i\hbar) \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right) (-i\hbar) \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) \\
&= -\hbar^2 \left\{ z \frac{\partial}{\partial x} \left( y \frac{\partial}{\partial z} \right) - z \frac{\partial}{\partial x} \left( z \frac{\partial}{\partial y} \right) - x \frac{\partial}{\partial z} \left( y \frac{\partial}{\partial z} \right) + x \frac{\partial}{\partial z} \left( z \frac{\partial}{\partial y} \right) \right\} \\
J_y J_x &= -\hbar^2 \left\{ zy \frac{\partial^2}{\partial x \partial z} - z^2 \frac{\partial^2}{\partial x \partial y} - xy \frac{\partial^2}{\partial z^2} + x \frac{\partial}{\partial y} + xz \frac{\partial^2}{\partial z \partial y} \right\} \quad \dots(2)
\end{aligned}$$

$x, y$  and  $z$  are perfect differentials.  $\frac{\partial^2}{\partial x \partial z} = \frac{\partial^2}{\partial x \partial z}, \dots$

Eq. (1) – Eq. (2) gives

$$[J_x, J_y] = J_x J_y - J_y J_x = -\hbar^2 \left\{ y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \right\}$$

$$[J_x, J_y] = i\hbar J_z.$$

Similarly, it can be proved that

$$[J_y, J_z] = i\hbar J_x, [J_z, J_x] = i\hbar J_y.$$

$$\therefore [J_x, J_y] = i\hbar J_z, [J_y, J_z] = i\hbar J_x, [J_z, J_x] = i\hbar J_y. \quad \dots(3)$$

### 10.11.3. Ladder Operators $J_+$ and $J_-$

We now introduce the raising and lowering operators.

We define the *ladder operators*

$$J_+ = J_x + iJ_y \text{ and } J_- = J_x - iJ_y$$

These operators are not hermitian, but they are adjoints of each other :

$$J_+^\dagger = J_- \text{ and } J_-^\dagger = J_+$$

**Commutation relation of  $J_z$  with ladder operators  $J_+ = (J_x + iJ_y)$  and  $J_- = (J_x - iJ_y)$**

$$\begin{aligned} [J_z, J_+] &= J_z J_+ - J_+ J_z = J_z (J_x + iJ_y) - (J_x + iJ_y) J_z = J_z J_x - J_x J_z - i(J_y J_z - J_z J_y) \\ &= [J_z, J_x] - i[J_y, J_z] = i\hbar J_y - i(i\hbar J_x) = i\hbar J_y + \hbar J_x = \hbar (J_x + iJ_y) = \hbar J_+ \\ [J_z, J_+] &= \hbar J_+ \end{aligned}$$

Similarly,  $[J_z, J_-] = -\hbar J_-$

**Commutation relation of  $J_+$  and  $J_-$  mutually**

$$\begin{aligned} [J_+, J_-] &= J_+ J_- - J_- J_+ = (J_x + iJ_y)(J_x - iJ_y) - (J_x - iJ_y)(J_x + iJ_y) \\ &= J_x (J_x - iJ_y) + iJ_y (J_x - iJ_y) - J_x (J_x + iJ_y) + iJ_y (J_x + iJ_y) \end{aligned}$$

On simplification, we get

$$\begin{aligned} &= 2i(J_y J_x - J_x J_y) = 2i[J_y, J_x] = 2i(-i\hbar J_z) = 2\hbar J_z \\ [J_+, J_-] &= 2\hbar J_z. \end{aligned}$$

**Commutation Relations of  $J^2$  with  $J_+$  and  $J_-$**

$$(i) \quad [J^2, J_+] = [J^2, J_x + iJ_y] = [J^2, J_x] + i[J^2, J_y]$$

Using Eqs. (3) and (4), of section 10.11.1, we get

$$[J^2, J_+] = 0 + 0$$

$$[J^2, J_+] = 0$$

$$(ii) \quad [J^2, J_-] = [J^2, J_x - iJ_y] = [J^2, J_x] - i[J^2, J_y] = 0$$

$$[J^2, J_-] = 0.$$

## 10.12 EIGEN VALUES OF $J^2$ AND $J_z$

Since  $J^2$  and  $J_z$  commute, they posses simultaneous eigenfunctions.

Let  $a$  and  $b$  be eigenvalues of operators  $J^2$  and  $J_z$  in the state  $|\psi\rangle$ .

- Eigenvalue equations of  $J^2$  and  $J_z$  in the state  $|\psi\rangle$  are

$$J^2 |\psi\rangle = a |\psi\rangle \quad \dots(1)$$

$$J_z |\psi\rangle = b |\psi\rangle \quad \dots(2)$$

We know that

$$[J_z, J_+] = J_z J_+ - J_+ J_z.$$

or

$$J_z J_+ = [J_z, J_+] + J_+ J_z$$

Operating by  $J_z J_+$  on  $|\psi\rangle$ , we get

$$\begin{aligned} J_z J_+ |\psi\rangle &= \{[J_z, J_+] + J_+ J_z\} |\psi\rangle \\ &= [J_z, J_+] |\psi\rangle + J_+ J_z |\psi\rangle \end{aligned}$$

But

$$[J_z, J_+] = \hbar J_+ \text{ and } J_z |\psi\rangle = b |\psi\rangle$$

$$J_z J_+ |\psi\rangle = \hbar J_+ |\psi\rangle + J_+ b |\psi\rangle$$

∴

$$J_z J_+ |\psi\rangle = (b + \hbar) J_+ |\psi\rangle \quad \dots(3)$$

Eq. (3) shows that  $J_+$  has eigenvalue  $(b + \hbar)$  in the state  $J_+ |\psi\rangle$ .

- The operation of  $J_+$  on  $\psi$  increases the eigenvalue of  $J_z$  by an amount  $\hbar$ .

Operating again by  $J_+$  on  $J_+ |\psi\rangle$ , we have

$$J_+ J_+ |\psi\rangle = J_+^2 |\psi\rangle$$

$$\begin{aligned}
J_z J_+^2 |\psi\rangle &= J_z J_+ (J_+ |\psi\rangle) \\
&= \{[J_z, J_+] + J_+ J_z\} J_+ |\psi\rangle \\
&= (\hbar J_+ + J_+ J_z) J_+ |\psi\rangle \\
&= \hbar J_+^2 |\psi\rangle + J_+ J_z J_+ |\psi\rangle \\
&= \hbar J_+^2 |\psi\rangle + J_+ (b + \hbar) J_+ |\psi\rangle \quad \text{using Eq. (3)} \\
J_z J_+^2 |\psi\rangle &= (b + 2\hbar) J_+^2 |\psi\rangle \quad \dots(4)
\end{aligned}$$

Eq. (4) shows that  $J_z$  has eigenvalue  $(b + 2\hbar)$  in the state  $J_+^2 |\psi\rangle$ .

The operation of  $J_+$  on  $|\psi\rangle$  each time means the increase in eigenvalue of  $J_z$  by  $\hbar$  each time.

In general, we write

$$J_z J_+^n |\psi\rangle = (b + n\hbar) J_+^n |\psi\rangle \quad \dots(5)$$

Similarly, for  $J_-$  operator, we have

$$J_z J_-^n |\psi\rangle = (b - n\hbar) J_-^n |\psi\rangle \quad \dots(6)$$

Eqs. (5) and (6) show that there is a discrete spectrum for the eigenvalues of  $J_z$  depending on the integer  $n$ .

The eigen values are

$$\begin{aligned}
n = -n, \dots &\quad n = -2, \quad n = -1, \quad n = 0, \quad n = 1, \quad n = 2, \dots \quad n = n \\
(b - n\hbar), \dots &\quad (b - 2\hbar), \quad (b - \hbar), \quad b, \quad (b + \hbar), \quad (b + 2\hbar), \dots \quad (b + n\hbar)
\end{aligned}$$

Thus we see that as  $n$  tends from  $-\infty$  to  $+\infty$ , the eigen values of  $J_z$  tend from  $-\infty$  to  $\infty$ .

Angular momentum of a system is finite. Therefore, eigenvalues of  $J^2$  must be finite i.e.,  $a$  is finite.

$$\text{Now, } J^2 = J_x^2 + J_y^2 + J_z^2, \text{ i.e., } J_z^2 < J^2.$$

Therefore, the eigenvalues of  $J_z$  must be finite and hence terminated after certain terms so that inequality  $J_z^2 < J^2$  is true.

$$\therefore J_z J_+^{l+1} |\psi\rangle = 0 \text{ and } J_z J_-^{k+1} |\psi\rangle = 0 \quad \dots(7)$$

i.e., maximum and minimum values of integers are  $l$  and  $k$  respectively defined by Eq. (7).

Therefore, the eigenvalues of  $J_z$  are restricted to the region

$$\begin{aligned}
n = -k, \dots &\quad n = -1, \quad n = 0, \quad n = 1, \dots \quad n = l \\
(b - k\hbar), \dots &\quad (b - \hbar), \quad b, \quad (b + \hbar), \dots \quad (b + l\hbar).
\end{aligned}$$

We know that

$$\begin{aligned}
J_+ &= J_x + iJ_y, J_- = J_x - iJ_y \\
\therefore J_z J_+ &= (J_x - iJ_y)(J_x + iJ_y) = J_x^2 + J_y^2 + i[J_x, J_y] \\
\text{or } J_x^2 + J_y^2 &= J_z J_+ - i(i\hbar J_z). \\
J_x^2 + J_y^2 &= J_z J_+ + \hbar J_z \quad \dots(8a)
\end{aligned}$$

If we start with  $J_+ J_-$  we get

$$\begin{aligned}
J_x^2 + J_y^2 &= J_+ J_- - \hbar J_z \\
\therefore J^2 &= J_x^2 + J_y^2 + J_z^2 = (J_x^2 + J_y^2) + J_z^2 \quad \dots(8b)
\end{aligned}$$

$$\text{Using Eq. (8a), } J^2 = J_z^2 + J_+ J_- + \hbar J_z, \quad \dots(9a)$$

$$\text{Using Eq. (8b), } J^2 = J_z^2 + J_+ J_- - \hbar J_z \quad \dots(9b)$$

Now let us find eigenvalues of  $J^2$  in terms of  $J_z$ .

$[J_z, J^2] = 0$ . Therefore,  $J_z$  and  $J^2$  have the same eigen function, i.e.,

If  $J_+^l |\psi\rangle$  is an eigenfunction of  $J_z$ , it will be also for  $J^2$ .

$$\begin{aligned}
\therefore J^2 J_+^l |\psi\rangle &= (J_z^2 + J_+ J_- + \hbar J_z) J_+^l |\psi\rangle \\
&= J_z^2 J_+^l |\psi\rangle + J_+ J_- J_+^l |\psi\rangle + \hbar J_z J_+^l |\psi\rangle.
\end{aligned}$$

Using Eqs. (5) and (7), we get

$$\begin{aligned}\hat{J}^2 J_+^l |\psi\rangle &= J_z J_z J_+^l |\psi\rangle + J_- J_+^{l+1} |\psi\rangle + \hbar J_z J_+^l |\psi\rangle \\&= J_z (b + l\hbar) J_+^l |\psi\rangle + 0 + \hbar(b + l\hbar) J_+^l |\psi\rangle \\&= \{(b + l\hbar)(b + l\hbar) + \hbar(b + l\hbar)\} J_+^l |\psi\rangle \\&= (b + l\hbar)(b + l\hbar + \hbar) J_+^l |\psi\rangle.\end{aligned}\quad \dots(10)$$

This equation gives the eigenvalues for  $J^2$ .

Again

$$\begin{aligned}\hat{J}^2 J_-^k |\psi\rangle &= (J_z^2 + J_+ J_- - \hbar J_z) J_-^k |\psi\rangle \\&= \{(b + k\hbar)^2 + 0 - \hbar(b - k\hbar)\} J_-^k |\psi\rangle \\&= (b - k\hbar)(b - k\hbar - \hbar) J_-^k |\psi\rangle.\end{aligned}\quad \dots(11)$$

Equating these two eigenvalues given by Eqs. (10) and (11) of  $J^2$ , which represent eigenvalues ' $a$ ', we have

$$a = (b + l\hbar)(b + l\hbar + \hbar) = (b - l\hbar)(b - k\hbar - \hbar). \quad \dots(12)$$

Solving for  $b$ , we get

$$b = \frac{\hbar}{2}(k - l) \quad \dots(13)$$

From Eq. (12), we have

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

We substitute  $\frac{k+l}{2} = j$  where  $\left( \frac{k+l}{2} \right)$  and  $\left( \frac{k-l}{2} \right)$  are necessarily integers.

$$a = j(j+1)\hbar^2 \quad \dots(14)$$

and

$$b = \frac{\hbar}{2}(k - l).$$

Now

$$j = \frac{k+l}{2}$$

∴

$$k_{\max} = 2j; k_{\min} = 0$$

$$l_{\max} = 2j; l_{\min} = 0$$

When

$$k = k_{\max} = 2j \quad l = l_{\min} = 0, \quad \text{we have } j_{\max} = j.$$

When

$$k = k_{\min} = 0 \quad l = l_{\max} = 2j, \quad \text{we have } j_{\min} = -j.$$

Thus eigenvalues  $b$  may be expressed as

$$-j\hbar, (-j+1)\hbar, \dots, -2\hbar, -\hbar, 0, \hbar, 2\hbar, \dots, (j-1)\hbar, j\hbar. \quad \dots(15)$$

Therefore,  $J_z$  has eigenvalues from  $-j\hbar$  to  $+j\hbar$  i.e., the total values are  $(2j+1)$ .

We define  $b = m_z$  when  $m_z$  has these all  $(2j+1)$  values.  $m_z$  is called *magnetic quantum number*.

The eigenvalues of  $J^2$  are  $j(j+1)\hbar^2$

The eigenvalues of  $J_z$  are  $m_z\hbar$  ...(16)

Equation (16) represents the eigenvalues of  $J^2$  and  $J_z$ . The matrix elements of  $J_z$  which is diagonal in above representation are determined by equation  $J_z \psi(jm) = m_z \psi(jm)$ .

If  $\psi(jm)$  are normalized,  $\{\psi(j'm'), J_z \psi(jm)\} = \langle j'm' | J_z | jm \rangle = m_z \hbar \delta_{mm'} \delta_{jj'}$ .

## EXERCISE

1. The operator for momentum is

- (a)  $-i\hbar\nabla$       (b)  $i\hbar\nabla$       (c)  $i\hbar \frac{\partial}{\partial t}$       (d)  $-i\hbar \frac{\partial}{\partial t}$  (B.U. 2014)

2. The quantum operator for energy is  
 (a)  $i\hbar\nabla$       (b)  $-i\hbar\nabla$       (c)  $i\hbar\frac{\partial}{\partial t}$       (d)  $-i\hbar\frac{\partial}{\partial t}$  (B.U. 2013)
3. Expectation value of momentum is  
 (a)  $\int \psi^*(i\hbar)\psi dx$       (b)  $\int \psi^*(-i\hbar)\psi dx$       (c)  $\int \psi^*(i\hbar)\nabla\psi dx$       (d)  $\int \psi^*(-i\hbar)\nabla\psi dx$  (B.U. 2008)
4. At normalisation condition, the expectation value of the operator  $A$  is .....  
 (a)  $\langle A \rangle = \int \psi^* A\psi d\tau$     (b)  $\langle A \rangle = \int |\psi|^2 d\tau$     (c)  $\langle A \rangle = \int A^2\psi d\tau$     (d)  $\langle A \rangle = \int A\psi d\tau$  (B.U. 2015)
5. The wave function of a particle in a classically forbidden region is  
 (a) a sine function      (b) a cosine function  
 (c) a positive exponential      (d) a negative exponential
6. If the function remains unchanged after operation, the operator is called .....  
 (a) null operator      (b) identity operator      (c) del operator      (d) momentum operator (B.U. 2011)
7. Probability current density is  
 (a)  $\frac{i\hbar}{2m} \left( \psi \frac{\partial \psi^*}{\partial x} - \psi^* \frac{\partial \psi}{\partial x} \right)$       (b)  $\frac{i\hbar}{2m} \left( \psi^* \frac{\partial \psi}{\partial x} + \psi \frac{\partial \psi^*}{\partial x} \right)$   
 (c)  $-\frac{i\hbar}{2m} \left( \psi \frac{\partial \psi^*}{\partial x} + \psi^* \frac{\partial \psi}{\partial x} \right)$       (d)  $-\frac{i\hbar}{2m} \left( \psi \frac{\partial \psi^*}{\partial x} - \psi^* \frac{\partial \psi}{\partial x} \right)$  (B.U. 2008)
8.  $\psi_1(x)$  and  $\psi_2(x)$  are said to be orthogonal, if  $\int \psi_1(x)\psi_2^*(x) dx = \dots$   
 (a) 0      (b) 1      (c) -1      (d)  $\infty$  (B.U. 2013)  
**Ans.** 1. (a), 2. (c), 3. (d), 4. (a), 5. (d), 6. (b), 7. (a), 8. (a)
9. State the principle of superposition of eigen states.
10. Which of the following wave functions are acceptable in quantum mechanics?  
 (a)  $\sin x$  (b)  $\tan x$  (c)  $\operatorname{cosec} x$  (d)  $\cos x + \sin x$   
 [Sol. (a) and (d) are acceptable wave functions. (b) and (c) are not acceptable because (b)  $\rightarrow \infty$  at  $x = \pi/2$  and (c)  $\rightarrow \infty$  at  $x = 0$ .]
11. Explain the postulates of quantum mechanics. (B.U. 2015)
12. Show that the operator  $L^2$  is  

$$L^2 = -\hbar^2 \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right] \quad (\text{Awadh 2006})$$
13. Find the eigen value of  $L^2$  and  $L_z$ . (Garhwal 1991, 92, Rohilkhand 94, 96)
14. Prove that two commuting operators have simultaneous eigen functions. (Purvanchal 2006)
15. Show that any two commuting set of operators have common eigen values. (Meerut 1980)
16. Define a commutator. Derive the commutation relation between Hamiltonian  $H$  and momentum  $p$ . (B.U. 2004)
17. Calculate the commutator for position and linear momentum operator and hence show that  

$$(xp_x - p_x x)\psi = i\hbar\psi. \quad (\text{P.U. 2000})$$
18. Prove that no two of the three components of angular momentum  $L$  commute with each other, but all of them commute with  $L^2$ . (Rohilkhand 96)
19. If Ladder operator  $L_+ = L_x + iL_y$  and  $L_- = L_x - iL_y$ , then find the value of commutators  $[L_z, L_+]$ ,  $[L_z, L_-]$  and  $[L_+, L_-]$ . (Rohilkhand 1994)

20. What is meant by Hermitian operator? Show that Hermitian operators have real eigen values.  
*(Purvanchal 2005)*
21. Show that the eigen functions of Hermitian operator belonging to different eigen values are orthogonal.  
*(Awadh 2005)*
22. What do you mean by parity of a wave-function? How do you test the symmetry of a wave function through this operator?  
*(Rohilkhand 1981)*
23. Explain probability current density.  
*(B.U. 2013)*
24. Prove the relation  $\frac{\partial P}{\partial t} + \vec{\nabla} \cdot \vec{J} = 0$  where  $\vec{J}$  is probability current density and  $P$  the probability density.  
*(Luck. U. 2002)*
25. State and prove Ehrenfest theorem.  
*(B.U. 2015)*
26. Explain formation of a Gaussian wave packet. Plot graphically and explain its properties. *(P.U. 2008)*
27. Solve linear harmonic oscillator problem using operator methods. Obtain the energy eigenvalues and eigenfunctions.
28. Define Hilbert space and illustrate its significance in the study of quantum mechanics.  
*(Rohilkhand 1991)*
29. Starting from commutation relation, derive the eigen values of operator  $J^2$  and  $J_z$  where  $J^2$  and  $J_z$  represent the square and the Z-component of the angular momentum operator.  
*(Rohilkhand 1994)*

## IDENTICAL PARTICLES AND SPIN

**After reading this chapter, you should be able to**

- ◆ Discuss the application of quantum mechanics to the structure of atoms
- ◆ Distinguish between symmetric and antisymmetric wave functions
- ◆ Explain and prove the Pauli exclusion principle
- ◆ Explain Pauli's spin matrices  $\sigma_x$ ,  $\sigma_y$  and  $\sigma_z$ .

## 11.1 IDENTICAL PARTICLES

*Identical particles in a system are regarded as those particles which when interchanged in the system will not make any change in it.* These particles can be distinguished from one another, only when their wave packets do not overlap. Similarly if the particles have spins which are aligned in different positions in the process of any interaction, they can be identified from one another as a result of such interactions.

*Thus the word “identical” in quantum mechanics is to describe particles that can be substituted for each other under the most general possible circumstances with no change in physical situation.*

For example, the component of spin along some particular axis is assumed not to change during elastic collisions. It means the particles can be distinguished if they have different spin components.

## 11.2 SYMMETRIC AND ANTISYMMETRIC WAVE FUNCTIONS

The Schrödinger equation for  $n$  identical particles is

$$H(1, 2, 3, \dots, n)\psi(1, 2, 3, \dots, n; t) = i\hbar \frac{\partial}{\partial t} \psi(1, 2, 3, \dots, n, t) \quad ..(1)$$

Here each of the numbers represents all the coordinates (positional and spin) of one of the particles. The Hamiltonian  $H$  is symmetrical in its arguments due to the identity of particles. The identity of particles means that they can be substituted for each other without changing  $H$  or indeed any other observable.

Now we define symmetric and antisymmetric wave functions.

(i) **Symmetric wave function  $\psi_s$ .** A wave function is symmetric if the interchange of any pair of particles among its arguments leaves the wave function unchanged.

(ii) **Antisymmetric wave function  $\psi_A$ .** A wave function is antisymmetric if the interchange of any pair of particles among its arguments changes the sign of the wave function.

If  $P$  is an exchange operator, then we must have

$$\begin{aligned} P\psi_s(1, 2) &= \psi_s(2, 1), \\ P\psi_A(1, 2) &= -\psi_A(2, 1). \end{aligned}$$

## 11.3 CONSTRUCTION OF SYMMETRIC AND ANTISYMMETRIC WAVE FUNCTIONS FROM UNSYMMETRISED FUNCTIONS

We now show how  $\psi_s$  or  $\psi_A$  can be constructed from a general unsymmetrized solution  $\psi$  of Eq. (1) [Sec. 11.2]. If the arguments of the wave function  $\psi$  are permuted in any way, the resulting wave function is also a solution of Eq. (1). This is because the same permutation applied throughout Eq. (1) does not impair its validity, since it corresponds simply to a relabeling of the particles. Since  $H$  is symmetric, the permuted  $H$  is the same as the original  $H$ , and the resulting equation is the same as Eq. (1) for the permuted  $\psi$ . In this way  $n!$  solutions can be obtained from any one solution, each of which corresponds to one of the  $n!$  permutations of the  $n$  arguments of  $\psi$ . Evidently any linear combination of these functions is also a solution of the wave equation (1).

The sum of all these functions is a symmetric (unnormalised) wave function  $\psi_s$ , since the interchange of any pair of particles changes any one of the component functions into another of them and the latter into the former, leaving the entire wave function unchanged.

An antisymmetric unnormalised wave function can be constructed by adding together all the permuted wave functions that arise from the original solution by means of an even number of interchanges of pairs of particles, and subtracting the sum of all the permuted wave functions that arise by means of an odd number of interchanges of pairs of particles in the original solution.

For a system of 3 particles, the Schroedinger's equation is

$$H(1, 2, 3) \psi(1, 2, 3) = E\psi(1, 2, 3) \quad \dots(1)$$

This equation has  $3! = 6$  solutions corresponding to the same eigen values  $E$ .

The six possible functions obtained by exchanging the indices of the particles are

$$\psi(1, 2, 3), \psi(2, 3, 1), \psi(3, 2, 1), \psi(1, 3, 2), \psi(2, 1, 3), \psi(3, 1, 2).$$

Out of these six functions, those arising by an even number of interchanges of the pairs of particles in original wave function  $\psi(1, 2, 3)$  are

$$\psi(1, 2, 3), \psi(2, 3, 1), \psi(3, 1, 2),$$

The functions arising by odd number of interchanges of pairs of particles in original function  $\psi(1, 2, 3)$  are

$$\psi(2, 1, 3), \psi(1, 3, 2), \psi(3, 2, 1)$$

So the *symmetric wave function* can be written as

$$\begin{aligned} \psi_S &= \psi(1, 2, 3) + \psi(2, 3, 1) + \psi(3, 2, 1) + \psi(1, 3, 2) \\ &\quad + \psi(2, 1, 3) + \psi(3, 1, 2), \end{aligned} \quad \dots(2)$$

The *antisymmetric wave function* is

$$\begin{aligned} \psi_A &= \psi(1, 2, 3) + \psi(2, 3, 1) + \psi(3, 1, 2) - \{\psi(1, 3, 2) \\ &\quad + \psi(2, 1, 3) + \psi(3, 2, 1)\}. \end{aligned} \quad \dots(3)$$

Here  $\psi_S$  and  $\psi_A$  are unnormalised wave functions.

## 11.4 DISTINGUISHABILITY OF IDENTICAL PARTICLES

Two identical particles will be distinguishable from each other if the sum of probabilities of the individual wave functions in two states is equal to the probability derived by the symmetrised wave function as given by the following equations:

$$\begin{aligned} \psi_S &= \psi(1, 2) + \psi(2, 1) \\ \psi_A &= \psi(1, 2) - \psi(2, 1) \\ |\psi(1, 2)|^2 + |\psi(2, 1)|^2 &= |\{\psi(1, 2) \pm \psi(2, 1)\}|^2 \\ &= |\psi(1, 2)|^2 + |\psi(2, 1)|^2 \\ &\quad \pm 2 \operatorname{Re} [\psi(1, 2) \psi^*(2, 1)] \end{aligned} \quad \dots(1)$$

Here,  $\operatorname{Re}$  denotes the real part of  $\{\psi(1, 2) \psi^*(2, 1)\}$ .

This is possible only when overlap of wave-functions  $\psi(1, 2)$  and  $\psi(2, 1)$  is zero  
or  $2 \operatorname{Re} \{\psi(1, 2) \psi^*(2, 1)\} = 0$ .

Thus when the co-ordinates (space and spin) of two particles are not the same between exchange degenerate functions, the interference term, i.e.,  $2 \operatorname{Re} \psi(1, 2) \psi^*(2, 1)$  becomes zero and particle co-ordinates do not overlap.

## 11.5 THE PAULI EXCLUSION PRINCIPLE

Schroedinger's time dependent equation for a system of  $N$ -particles is

$$i\hbar \frac{\partial \psi}{\partial t}(1, 2, 3, r \dots s \dots N; t) = \hat{H}(1, 2, \dots, N) \psi(1, 2, 3, \dots, r, \dots, s, \dots, N; t) \quad \dots(1)$$

As Hamiltonian does not involve time, the stationary solution can be found.

Eq. (1) can be written as

$$[\hat{H}(1, 2, 3, \dots, N) - E] \psi(1, 2, 3, \dots, r, \dots, s, \dots, N) = 0 \quad \dots(2)$$

A useful zero-order approximation can be obtained by neglecting the interactions between the particles that make up the system under consideration. The approximate (unperturbed) Hamiltonian is the sum of equal Hamiltonian functions for the separate particles.

$$H_0(1, 2, 3 \dots N) = H'_0(1) + H'_0(2) + \dots + H'_0(N), \quad \dots(3)$$

Here,  $H_0(1, 2, 3 \dots N)$  is unperturbed Hamiltonian for the complete system.  $H'_0(1)$  etc., are the unperturbed Hamiltonian for the individual particles.

The approximate energy eigenfunction is a product of one-particle eigenfunctions of  $H'_0$ .

$$\psi(1, 2, 3 \dots N) = \psi_{n1}(1) \psi_{n2}(2) \dots \psi_{nN}(N) \quad \dots(4)$$

Here,  $n_1, n_2, n_N$  indicate the totality of the quantum numbers required to represent the state. If  $E$  represents the total energy of the system,

$$E = E_{n1} + E_{n2} + \dots + E_{nN}, \quad \dots(5)$$

Moreover, the wave functions are such that they obey

$$H'_0(1) \psi_{n1}(1) = E_{n1} \psi_{n1}(1) \text{ etc.} \quad \dots(6)$$

Thus for Fermi particles, the antisymmetric state which only describes the actual state of affairs can be written as follows :

$$\psi_A(1, 2, 3 \dots N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_{n1}(1) & \psi_{n1}(2) & \dots & \psi_{n1}(N) \\ \psi_{n2}(1) & \psi_{n2}(2) & \dots & \psi_{n2}(N) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \psi_{nN}(1) & \psi_{nN}(2) & \dots & \psi_{nN}(N) \end{vmatrix} \quad \dots(7)$$

Here  $\psi_A$  is solution of the equation

$$H_0(1, 2, 3 \dots N) \psi_A(1, 2, \dots N) = E \psi_A(1, 2 \dots N) \quad \dots(8)$$

Eq. (7) suggests that if two  $\psi_s$ 's are the same, the state does not exist. Thus the unperturbed Hamiltonian  $H_0$  has no solutions for which there is more than one electron in any one of the states 1, 2, 3 ..., N. This result is known as the *Pauli's exclusion principle*. It states that *no two particles described by antisymmetric wave functions* (or obeying Fermi-Dirac Statistics) *can exist in same quantum state*.

## 11.6 SPIN

Spin is an intrinsic property of particles. This property was deduced from the Stern-Gerlach experiment. From other observations we conclude that spin is a property of elementary particles, such as protons, neutrons, and electrons. Spin is a vector quantity. The characteristic values of the component of the spin in any prescribed direction are given by  $\hbar/2$  and  $-\hbar/2$ . Moreover, the spin is measured in unit of  $\hbar$  and the characteristic values are  $\frac{1}{2}$  and  $-\frac{1}{2}$ . This type of behavior can be described by a spin quantum number  $s = \frac{1}{2}$ .

## 11.7 PAULI'S SPIN MATRICES FOR ELECTRON

Like orbital angular momentum operators  $L_x, L_y, L_z$ , the spin operators  $S_x, S_y$  and  $S_z$  to be associated with the components of spin angular momentum satisfy the commutation relations:

$$\begin{aligned} [S_x, S_y] &= S_x S_y - S_y S_x = i\hbar S_z \\ [S_y, S_z] &= S_y S_z - S_z S_y = i\hbar S_x \\ [S_z, S_x] &= S_z S_x - S_x S_z = i\hbar S_y \end{aligned} \quad \dots(1)$$

We consider the electron, with spin  $= \frac{1}{2}$ . Then, according to Uhlenbeck and Goudsmit hypothesis, each of operators,  $S_x$ ,  $S_y$  and  $S_z$  must have just two eigen values  $\frac{1}{2}\hbar$  and  $-\frac{1}{2}\hbar$ . Now introduce new auxiliary operators  $\sigma_x$ ,  $\sigma_y$  and  $\sigma_z$  such that

$$\left. \begin{aligned} S_x &= \frac{1}{2}\hbar\sigma_x \\ S_y &= \frac{1}{2}\hbar\sigma_y \\ S_z &= \frac{1}{2}\hbar\sigma_z \end{aligned} \right\} \quad \dots(2)$$

The following properties of  $\sigma$ 's may be noted:

Since the eigen values of each  $S$  are to be just  $\frac{1}{2}\hbar$  and  $-\frac{1}{2}\hbar$ , the eigen values of each  $\sigma$  must be 1 and  $-1$ . Each of the operators  $\sigma_x^2$ ,  $\sigma_y^2$  and  $\sigma_z^2$  must therefore have only the eigen value 1 and such operator is only unit operator. Therefore,

$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1 \quad \dots(3)$$

According to Eqs. (2) and (3), the commutation rules satisfied by  $\sigma$ 's must be

$$\left. \begin{aligned} [\sigma_x, \sigma_y] &= \sigma_x \sigma_y - \sigma_y \sigma_x = 2i\sigma_z \\ [\sigma_y, \sigma_z] &= \sigma_y \sigma_z - \sigma_z \sigma_y = 2i\sigma_x \\ [\sigma_z, \sigma_x] &= \sigma_z \sigma_x - \sigma_x \sigma_z = 2i\sigma_y \end{aligned} \right\} \quad \dots(4)$$

Now

$$\begin{aligned} 2i(\sigma_x \sigma_y + \sigma_y \sigma_x) &= (2i\sigma_x)\sigma_y + \sigma_y(2i\sigma_x) \\ &= (\sigma_y \sigma_z - \sigma_z \sigma_y)\sigma_y + \sigma_y(\sigma_y \sigma_z - \sigma_z \sigma_y) \\ &= -\sigma_z \sigma_y^2 + \sigma_y^2 \sigma_z \\ &= -\sigma_z + \sigma_z \\ &= 0. \end{aligned}$$

Hence

$$\sigma_x \sigma_y = -\sigma_y \sigma_x.$$

$\therefore \sigma_x$  and  $\sigma_y$  anticommute.

Similarly, any two of the  $\sigma$ 's anticommute as

$$\left. \begin{aligned} \sigma_x \sigma_y + \sigma_y \sigma_x &= 0 \\ \sigma_y \sigma_z + \sigma_z \sigma_y &= 0 \\ \sigma_z \sigma_x + \sigma_x \sigma_z &= 0 \end{aligned} \right\} \quad \dots(5)$$

Finally from Eqs. (4) and (5) we have

$$\sigma_x \sigma_y = i\sigma_z, \quad \dots(6a)$$

$$\sigma_y \sigma_z = i\sigma_x, \quad \dots(6b)$$

$$\sigma_z \sigma_x = i\sigma_y, \quad \dots(6c)$$

Each  $\sigma$  has two eigen values. So  $(2 \times 2)$  matrix may be expected to fulfil the purpose. We begin by associating with  $\sigma_z$ , the simplest  $(2 \times 2)$  matrix having the eigen values 1 and  $-1$ .

$$\sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \dots(7)$$

$$\sigma_x \sigma_z = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} a & -b \\ c & -d \end{bmatrix} \quad \dots(8)$$

and  $\sigma_z \sigma_x = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ -c & -d \end{bmatrix} \quad \dots(9)$

But  $\sigma_x$  and  $\sigma_z$  anticommute.

$$\therefore \sigma_x \sigma_z + \sigma_z \sigma_x = \begin{bmatrix} a & -b \\ c & -d \end{bmatrix} + \begin{bmatrix} a & b \\ -c & -d \end{bmatrix} = 0$$

i.e.,  $\begin{bmatrix} 2a & 0 \\ 0 & -2d \end{bmatrix} = 0.$

This yields  $a = d = 0$ .

So every matrix that anticommutes with (7) as  $\sigma_x$  and  $\sigma_y$  must have the form

$$\sigma_x = \begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix}. \quad \dots(10)$$

The eigen values of (10) are  $\pm\sqrt{(bc)}$ , so that if they are to be 1 and -1, we must set  $bc = 1$ . And simplest possibility is to take  $b = c = 1$ .

$$\therefore \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

It then follows from (6c) that the matrix to be associated with  $\sigma_y$  is

$$\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$$

Hence complete list of  $\sigma$ 's becomes

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \dots(11)$$

These matrices  $\sigma_x$ ,  $\sigma_y$ , and  $\sigma_z$  are called *Pauli spin matrices* associated with the components of spin angular momentum

$$S_x = \frac{1}{2}\hbar \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad S_y = \frac{1}{2}\hbar \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad S_z = \frac{1}{2}\hbar \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \dots(12)$$

Thus we have  $\sigma^2 = \sigma_x^2 + \sigma_y^2 + \sigma_z^2 = 3$

$$S^2 = S_x^2 + S_y^2 + S_z^2 = \frac{\hbar^2}{4}(\sigma_x^2 + \sigma_y^2 + \sigma_z^2) = \frac{3}{4}\hbar^2$$

## 11.8 COMMUTATION RELATIONS SATISFIED BY THREE COMPONENTS $\sigma_x$ , $\sigma_y$ AND $\sigma_z$

(i) We have

$$[\sigma_x, \sigma_y] = \sigma_x \sigma_y - \sigma_y \sigma_x.$$

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$[\sigma_x, \sigma_y] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} - \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\begin{aligned}
 &= \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} - \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} + \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \\
 &= 2 \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} = 2i \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\
 &= 2i\sigma_z.
 \end{aligned}$$

$$\begin{aligned}
 (ii) \quad [\sigma_y, \sigma_z] &= \sigma_y \sigma_z - \sigma_z \sigma_y \\
 &= \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \\
 &= \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} - \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix} \\
 &= \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} + \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} \\
 &= 2 \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} = 2i \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\
 &= 2i\sigma_x.
 \end{aligned}$$

$$\begin{aligned}
 (iii) \quad [\sigma_z, \sigma_x] &= \sigma_z \sigma_x - \sigma_x \sigma_z \\
 &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\
 &= \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} - \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \\
 &= \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \\
 &= 2 \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = -2 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \\
 &= 2i^2 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = 2i \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \\
 &= 2i\sigma_y.
 \end{aligned}$$

(iv) Commutation relation satisfied by  $\sigma^2$  and its components  $\sigma_x$ ,  $\sigma_y$  and  $\sigma_z$

We know that

$$\sigma^2 = \sigma_x^2 + \sigma_y^2 + \sigma_z^2.$$

Let us take

$$\begin{aligned}
 [\sigma^2, \sigma_x] &= [\sigma_x^2 + \sigma_y^2 + \sigma_z^2, \sigma_x] \\
 &= [\sigma_x^2, \sigma_x] + [\sigma_y^2, \sigma_x] + [\sigma_z^2, \sigma_x]
 \end{aligned}$$

But

$$[\sigma_x^2, \sigma_x] = \sigma_x^3 - \sigma_x^3 = 0.$$

So

$$\begin{aligned}
 [\sigma^2, \sigma_x] &= [\sigma_y^2, \sigma_x] + [\sigma_z^2, \sigma_x] \\
 &= [\sigma_y \sigma_y, \sigma_x] + [\sigma_z \sigma_z, \sigma_x]
 \end{aligned}$$

We know that

$$[ab, c] = a [b, c] + [a, c] b$$

So

$$\begin{aligned}
 [\sigma^2, \sigma_x] &= \sigma_y [\sigma_y, \sigma_x] + [\sigma_y, \sigma_x] \sigma_y + \sigma_z [\sigma_z, \sigma_x] + [\sigma_z, \sigma_x] \sigma_z \\
 &= \sigma_y (-2i\sigma_z) + (-2i\sigma_z) \sigma_y + \sigma_z (2i\sigma_y) + (2i\sigma_y) \sigma_z
 \end{aligned}$$

$$\begin{aligned}
 &= -2i\sigma_y\sigma_z - 2i\sigma_z\sigma_y + 2i\sigma_z\sigma_y + 2i\sigma_y\sigma_z \\
 &= 0 \\
 \therefore \quad [\sigma^2, \sigma_x] &= 0 \\
 \text{Similarly,} \quad [\sigma^2, \sigma_y] &= 0 \\
 [\sigma^2, \sigma_z] &= 0.
 \end{aligned}$$

Thus  $\sigma^2$  commutes with each component  $\sigma_x$ ,  $\sigma_y$  and  $\sigma_z$ .

### 11.9 TWO COMPONENT WAVE FUNCTIONS

Pauli operators  $\sigma_x$ ,  $\sigma_y$ ,  $\sigma_z$  are  $2 \times 2$  matrices. Therefore, the Pauli operands must be two-component column matrices. Pauli theory retains in essence the Schroedinger operators for dynamical variables having classical analogues. So the components of these column symbols must be functions of  $(x, y, z)$  or  $(r, \theta, \phi)$  and so on. Thus simplest Pauli operand  $\psi$  has the form

$$\psi = \begin{pmatrix} \psi_1(x, y, z) \\ \psi_2(x, y, z) \end{pmatrix} \quad \dots(1a)$$

$$\text{or} \quad \psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}. \quad \dots(1b)$$

A Pauli wave-function  $\psi$  is said to be *well behaved* if at least one of its components is non-zero and if each of its non-zero components is well behaved in the sense of Schrodinger's theory. The complex conjugate of Pauli wave function corresponding to (1) is the row matrix

$$\psi^* = [\psi_1^*, \psi_2^*] \quad \dots(2)$$

$$\therefore \quad \psi^*\psi = [\psi_1^*, \psi_2^*] \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \psi_1^*\psi_1 + \psi_2^*\psi_2 \quad \dots(3)$$

The Pauli wave function is said to be normalised if

$$\int \psi^*\psi d\tau = 1 \quad \dots(4a)$$

$$\text{or} \quad \int (\psi_1^*\psi_1 + \psi_2^*\psi_2) d\tau = 1 \quad \dots(4b)$$

Here, the integration extends over the entire three-dimensional space.

### EXERCISE

1. What are symmetric and antisymmetric wave functions? Show how they lead to the Pauli exclusion principle. *(Rohilkhand, 1984)*
2. Construct symmetric and antisymmetric wave functions for three spinless identical particles from unsymmetrised wave function. Show that the antisymmetric wave function obeys Pauli exclusion principle. *(Agra, 1976)*
3. What are Pauli spin operators? Express Pauli spin functions in the form of  $2 \times 2$  matrices.
4. Explain why Pauli introduced a set of  $2 \times 2$  spin matrices and obtain the commutation relations satisfied by the three components of the spin vector. *(Agra, 1970)*

## SCATTERING THEORY

**After reading this chapter, you should be able to**

- ◆ Develop a knowledge and understanding of the scattering theory and partial wave analysis
- ◆ Distinguish between incoming and outgoing waves.
- ◆ Relate scattering cross-section and scattering amplitude
- ◆ Prove the optical theorem.

## 12.1 INTRODUCTION

Scattering theory plays a very important role in quantum mechanics particularly in the field of high energy physics.

Scattering experiments have provided most of the information about interactions among atoms, nuclei and particles. The general technique is to have a collimated beam of projectiles (particles) with well defined energies and sometimes other properties, such as polarization, as well. The particles are then scattered by a target of atoms. The scattered particles are detected by means of a suitable detector. The measurement could be simply the intensity variation as a function of angle or could involve more detailed information like the change in the energy or the polarization state, etc. When a target nucleus is bombarded by means of energetic projectile particles, two types of scattering are possible.

**(i) Elastic scattering.** In this case, the incident particle strikes the target nucleus and leaves without loss of energy, but its direction may change.

**Example:** Scattering of  $\alpha$ -particles from a thin gold foil (Rutherford Scattering)



The target nucleus remains unaffected.

If the incident particle is denoted by ' $a$ ' and the target nucleus by  $X$ , then the elastic scattering process is written as

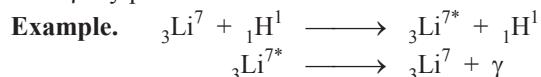


It is a process in which the kinetic energy is conserved.

Thus if the incident particles and the target nuclei are simply scattered by each other without any change in their relative energy, the process is elastic scattering.

**(ii) Inelastic scattering.** In this case, the kinetic energy is not conserved.

The incident particle loses a part of its energy in exciting the target nucleus to a higher allowed energy level. The excited nucleus later decays to the ground state, radiating the excess energy in the form of a  $\gamma$ -ray photon.



The {\*} is used to indicate that, after scattering, the target nucleus is left in an excited state.

In this process, the target nucleus, after the interaction, is not left in its initial state but is raised to an *excited* state. So the emergent particle, although the same as the incident particle, emerges with an energy lower than the energy of incident particle by an amount equal to the excitation energy given to the target nucleus. Symbolically, inelastic scattering is represented as



The quantum mechanical interpretation of the scattering process is rather different from that in a classical case.

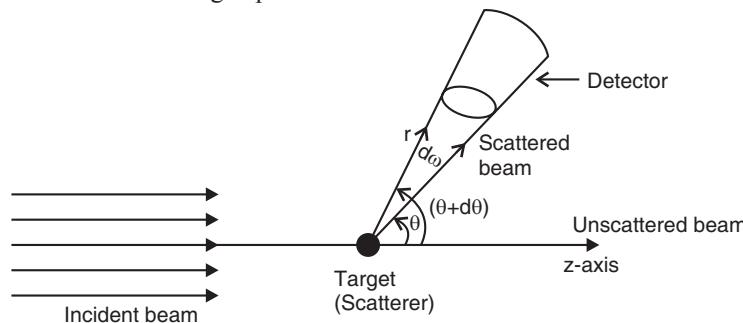
The motion of a particle cannot be described with complete accuracy by classical orbits. We must use wave packet whose average coordinates give the classical orbits.

Since the particle cannot be localised, a well defined impact parameter cannot be assigned to it.

The scattering must therefore, be described by the wave functions that are solutions of the Schrodinger equations. The probability of scattering is computed by determining the flux at a point. The scattering process is best described by a quantity called the *scattering cross section*. The most commonly used parameter which displays the results of scattering experiments is the *scattering cross-section*.

## 12.2 SCATTERING EXPERIMENT—DEFINITIONS OF CROSS-SECTIONS

Figure 12.1 shows a scattering experiment.



**Fig. 12.1**

An incident beam of monoenergetic particles is scattered by the target.

The number of particles scattered per unit time into an element of solid angle centred about a direction specified by polar angles  $(\theta, \phi)$  is counted with the aid of a detector (See Section 4.1).

$J$  is the magnitude of the incident flux *i.e.*, the number of incident particles crossing per unit time a unit surface area placed perpendicular to the direction of incident beam.

$\rho$  is the number of particles per unit volume in the incident beam.

$v$  is the velocity of the incident particles (Fig. 12.2).  
Then,

$$J = \rho v \quad \dots(1)$$

$J$  is called the current density.

$n$  is the number of particles scattered per unit time into a solid angle  $d\omega$  in the direction  $(\theta, \phi)$  [Fig. 12.3]. Then,

$n$  is directly proportional to the incident current *i.e.*,

$$n \propto J d\omega$$

or

$$n = \Sigma(\omega) J d\omega.$$

Here,  $\Sigma(\omega)$  is a constant of proportionality.

It is called *Scattering cross-section of the particle and the target in the direction  $\omega(\theta, \phi)$* .

### Differential Scattering Cross-section

Consider the target to be made up of a large number  $N$  of atomic or nuclear scattering centres.

Then  $n$  is directly proportional to  $N$  also *i.e.*,  $n \propto N$ .

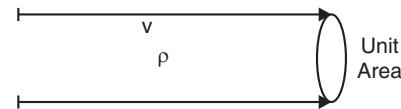
$$n \propto N J d\omega$$

$$\therefore n = \sigma(\omega) N J d\omega$$

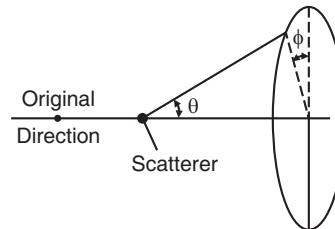
Here,  $\sigma(\omega)$ , the constant of proportionality, is called the *scattering cross-section* of the particle by the scattering centre in the direction  $\omega(\theta, \phi)$  or briefly the *differential scattering cross-section*.

$$\sigma(\omega) = \frac{n}{N J d\omega} \quad \dots(2)$$

**Differential Scattering Cross-section** is defined as the ratio of the number of particles scattered into the direction  $(\theta, \phi)$  per unit time, per unit solid angle, divided by the incident flux.



**Fig. 12.2**



**Fig. 12.3**

### Total Scattering Cross-section

The total number of particles scattered in unit time is obtained by integrating  $n$  over all angles.

$$N_{\text{total}} = \int \sigma(\omega) N J d\omega = NJ \sigma_{\text{total}}$$

Here,

$$\sigma_{\text{total}} = \int \sigma(\omega) d\omega \quad \dots(3)$$

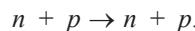
$\sigma_{\text{total}}$  is the total scattering cross section.

The total scattering cross-section  $\sigma_{\text{total}}$  is defined as the integral of the differential scattering cross-section over all solid angles.

In nuclear physics, the scattering centres have the dimensions of the order of  $10^{-12}$  cm. Hence, cross-sections having the dimensions of area are measured in barns.

$$1 \text{ barn} = 10^{-24} \text{ cm}^2 = 10^{-28} \text{ m}^2.$$

We shall be mainly concerned with elastic scattering in which the internal energies of the incident and target particles do not change and in which no further particles are created. An example is the scattering of neutrons by protons at low energies:



#### 12.2.1. Centre of Mass and Laboratory Reference Frames

The scattering of particles can be visualised in two kinds of coordinates :

1. *Laboratory frame or system (L-system)* : It is that coordinate system in which the bombarded particle (or target) is initially at rest [Fig. 12.4 (a)].

2. *Centre of mass coordinate system (C-system)* : It is that coordinate system in which the centre of mass of two colliding particles is at rest (initially and always) [Fig. 12.4 (b)].

In the centre of mass system, the two particles approach each other with the same momentum.

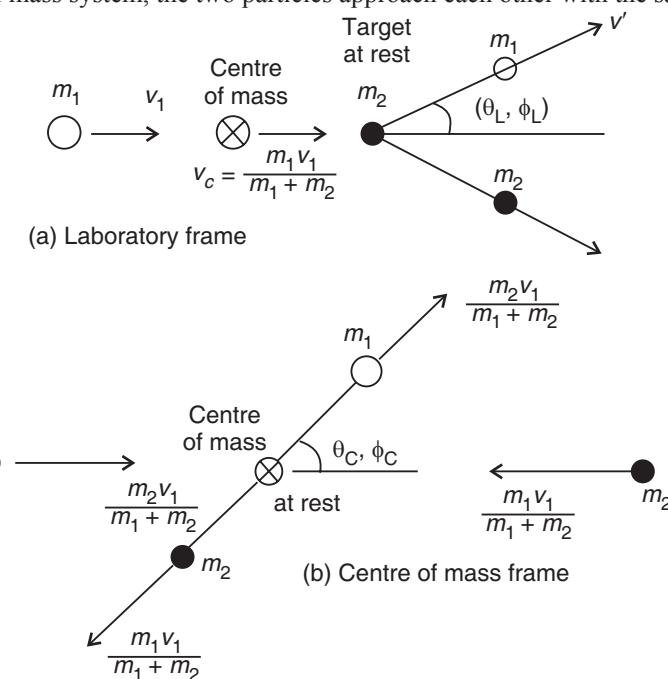


Fig. 12.4

### Transformation from Centre of Mass to laboratory Frame of Reference

#### Relation between Angles in L-system and C-system

(a) **Laboratory Frame.** Consider a particle of mass  $m_1$  moving with velocity  $v_1$  towards the target particle of mass  $m_2$  which is initially at rest [Fig. 12.4 (a)].

The centre of mass moves with velocity

$$v_c = \frac{m_1 v_1}{m_1 + m_2} \quad \dots(1)$$

' is the speed of the particle of mass  $m_1$  after scattering.

After scattering, the incident particle moves in a direction given by the angle  $(\theta_L, \phi_L)$ .

The target particle also suffers a recoil and moves in a different direction.

(b) **Centre of Mass Frame.** In the centre of mass system, the centre of mass is at rest and the two particles will approach each other with velocities

$$\frac{m_2 v_1}{(m_1 + m_2)} \text{ and } \frac{m_1 v_1}{(m_1 + m_2)}.$$

The total momenta of the two particles before scattering is zero.

Hence they must fly off in opposite directions after scattering [Fig. 12.4 (b)].

In the centre of mass system, the scattering angle is denoted by  $(\theta_c, \phi_c)$ .

As collision is symmetric in  $\phi$ -coordinate,

$$\phi_c = \phi_L \quad \dots(2)$$

For elastic scattering, the particles should move away with their original speeds.

In order to get back to the laboratory system we must superpose the velocity of the centre of mass.

$$v' \sin \theta_L = \frac{m_2 v_1}{m_2 + m_1} \sin \theta_c \quad \dots(3)$$

$$v' \cos \theta_L = \frac{m_2 v_1}{m_2 + m_1} \cos \theta_c + v_c \quad \dots(4)$$

∴

$$\tan \theta_L = \frac{\sin \theta_c}{\frac{m_1}{m_2} + \cos \theta_c} \quad \dots(5)$$

When  $m_2 \gg m_1$ , the centre of mass is almost at rest,  $\theta_L \approx \theta_C$  and the two system of coordinates coincide.

#### Relation between cross-sections in L-system and C-system

Since the laboratory and centre of mass scattering angles are not equal, the angular distribution of the cross-section will be different in the two systems. The relation between the two scattering cross-sections is obtained by noting that the number of particles scattered in a given element of solid angle must be the same in two systems. Thus

$$\sigma(\theta_L, \phi_L) \sin \theta_L d\theta_L d\phi_L = \sigma(\theta_c, \phi_c) \sin \theta_c d\theta_c d\phi_c \quad \dots(6)$$

The azimuthal angles in the two frames are equal:  $\phi_L = \phi_c$

or

$$d\phi_L = d\phi_c \quad \dots(7)$$

∴

$$\sigma(\theta_L, \phi_L) = \sigma(\theta_c, \phi_c) \frac{\sin \theta_c}{\sin \theta_L} \left| \frac{d\theta_c}{d\theta_L} \right| \quad \dots(8)$$

$$\text{From Eq. (5), } \frac{d\theta_c}{d\theta_L} = \frac{(m_1/m_2)^2 + 2(m_1/m_2) \cos \theta_c + 1}{(m_1/m_2) \cos \theta_c + 1} \quad \dots(9)$$

$$\therefore \sigma(\theta_L, \phi_L) = \frac{\left[ 1 + \frac{m_1^2}{m_2^2} + 2 \cdot \frac{m_1}{m_2} \cdot \cos \theta_c \right]^{3/2}}{\left( 1 + \frac{m_1}{m_2} \cos \theta_c \right)} \sigma(\theta_c, \phi_c) \quad \dots(10)$$

It is  $\sigma(\theta_c, \phi_c)$  which is obtained from the theory.

Eq. (10) tells us how to transform it to the laboratory system which should be compared with the experimental data.

### 12.3 REDUCTION OF THE TWO-BODY PROBLEM INTO ONE-BODY PROBLEM

(i) We first give the simplification of the problem according to *classical mechanics*. If a system of particles is not subject to any external forces (i.e., forces other than the mutual interactions of the particles themselves) then the centre of mass of the system moves with constant velocity  $v_c$ . In a reference frame which moves along with the same velocity  $v_c$ , the centre of mass of the particle-system appears to remain at rest. Any such reference frame is called a *centre of mass frame*. In the particular case of a two-body system, the motion as seen from the centre of mass frame is especially simple since, with the centre of mass remaining fixed, the instantaneous position of either particle completely determines that of the other. The two-body problem then becomes effectively a one-body problem.

#### (ii) Simplification of the Problem in Quantum Mechanics

For a two body problem, whenever the potential energy depends only on the magnitude of the distance between the two particles, the problem can always be reduced to a one body problem along with a uniform translational motion of the centre of mass.

Consider a two-particle system with the Hamiltonian

$$H = -\frac{\hbar^2}{2m_1} \nabla_1^2 - \frac{\hbar^2}{2m_2} \nabla_2^2 + V(\mathbf{x}_1 - \mathbf{x}_2) \quad \dots(1)$$

The subscripts 1 and 2 refer to the two particles.

We define the *centre of mass* and *relative* co-ordinate variables as

$$\mathbf{X} = -\frac{m_1 \mathbf{x}_1 + m_2 \mathbf{x}_2}{m_1 + m_2} \text{ and } \mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2 \quad \dots(2)$$

We denote the corresponding gradient operators by

$$\nabla_X \equiv \left( \frac{\partial}{\partial X}, \frac{\partial}{\partial Y}, \frac{\partial}{\partial Z} \right) \text{ and } \nabla = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \quad \dots(3)$$

Both  $\nabla_1 \equiv (\partial/\partial x_1, \partial/\partial y_1, \partial/\partial z_1)$  and  $\nabla_2$  can be expressed in terms of these.

$$\begin{aligned} \frac{\partial}{\partial x_1} &= \frac{\partial X}{\partial x_1} \frac{\partial}{\partial X} + \frac{\partial x}{\partial x_1} \frac{\partial}{\partial x} = \frac{m_1}{m_1 + m_2} \frac{\partial}{\partial X} + \frac{\partial}{\partial x} \\ \frac{\partial^2}{\partial x_1^2} &= \left( \frac{m_1}{m_1 + m_2} \right)^2 \frac{\partial^2}{\partial X^2} + \frac{2m_1}{m_1 + m_2} \frac{\partial^2}{\partial X \partial x} + \frac{\partial^2}{\partial x^2}, \text{ etc.} \end{aligned} \quad \dots(4)$$

$$\nabla = \left( \frac{m_1}{m_1 + m_2} \right)^2 \nabla_X^2 + \frac{2m_1}{m_1 + m_2} \nabla_X \cdot \nabla + \nabla^2 \quad \dots(5)$$

$$\nabla = \left( \frac{m_2}{m_1 + m_2} \right)^2 \nabla_X^2 - \frac{2m_2}{m_1 + m_2} \nabla_X \cdot \nabla + \nabla^2 \quad \dots(6)$$

On substituting Eqs. (5) and (6) into Eq. (1), it becomes

$$H = -\frac{\hbar^2}{2M} \nabla_X^2 - \frac{\hbar^2}{2\mu} \nabla^2 + V(\mathbf{x}) \quad \dots(7)$$

Here,

$$M = m_1 + m_2, \mu = \frac{m_1 m_2}{m_1 + m_2} \quad \dots(8)$$

$\mu = m_1 m_2 / (m_1 + m_2)$  is called the *reduced mass* of the system.

In this form,  $H$  is sum of two commuting parts depending on independent sets of variables.

(i) The first part,  $H_{cm} = -(\hbar^2/2M) \nabla_X^2$  describes the free motion of the centre of mass.

It represents the kinetic energy of the system (with total mass  $M$ ) as a whole.

(ii) The second part,  $H_{rel} = -\frac{\hbar^2}{2\mu} \nabla^2 + V(\mathbf{x})$  depends on the relative variable  $\mathbf{x}$  only. It is identical in form with the Hamiltonian of a single particle of mass  $\mu$  moving in a potential  $V(\mathbf{x})$ .

A complete set of eigen functions of  $H$  is now given by

$$(2\pi)^{-3/2} \exp(i\mathbf{K} \cdot \mathbf{X}) \Psi_n(\mathbf{x}) \quad \dots(9)$$

(for all  $\mathbf{K}, n$ )

(i) The factor  $(2\pi)^{-3/2} \exp[i\mathbf{K} \cdot \mathbf{X}]$  describes the motion of the centre of mass with momentum  $\hbar\mathbf{K}$ .

(ii)  $\Psi_n(\mathbf{x})$  are eigen functions of  $H_{rel}$  given by

$$-\frac{\hbar^2}{2M} \nabla_X^2 [(2\pi)^{-3/2} \exp(i\mathbf{K} \cdot \mathbf{X})] = \frac{\hbar^2 K^2}{2M} [(2\pi)^{-3/2} \exp(i\mathbf{K} \cdot \mathbf{X})] \quad \dots(10)$$

$$\left[ -\frac{\hbar^2}{2\mu} \nabla^2 + V(\mathbf{x}) \right] \Psi_n(\mathbf{x}) = E_n \Psi_n(\mathbf{x}) \quad \dots(11)$$

Thus the problem reduces essentially to that of solving the single-particle Schrodinger equation.

## 12.4 QUANTUM SCATTERING THEORY

Consider an incident *plane wave*,

$$\psi(z) = A e^{ikz} \quad \dots(1)$$

travelling in the  $z$  direction (Fig. 12.5).

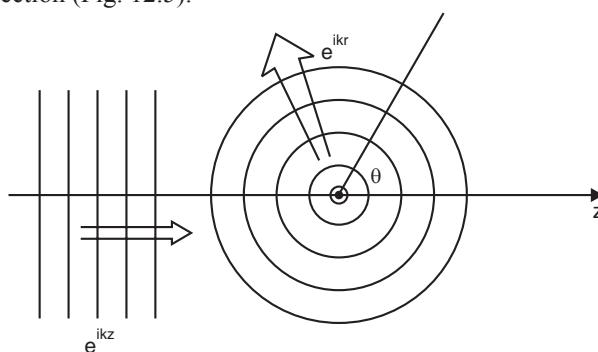


Fig. 12.5

The incident *plane* wave encounters a scattering potential, producing an outgoing *spherical wave*.

We look for solutions to the Schrodinger equation of the general form

$$\psi(r, \theta) \approx A \left\{ e^{ikz} + f(\theta) \frac{e^{ikr}}{r} \right\}, \text{ for large } r. \quad \dots(2)$$

(The spherical wave carries a factor of  $1/r$ , because this portion of  $|\psi|^2$  must go like  $1/r^2$  to conserve probability.) The **wave number**  $k$  is related to the energy  $E$  of the incident particles in the usual way:

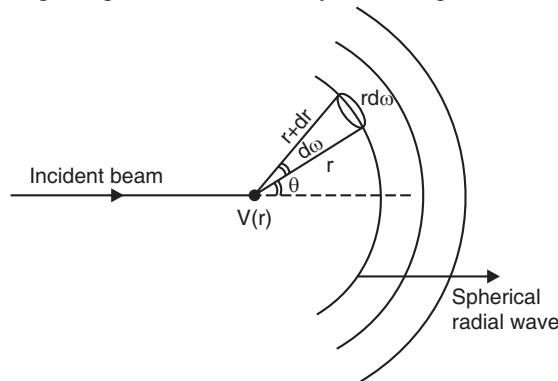
$$k = \frac{\sqrt{2mE}}{\hbar} \quad \dots(3)$$

We make the following assumptions in the scattering theory:

- (a) The incident beam is parallel.
- (b) Scattering potential is spherically symmetric.
- (c) When the incident particle is at a large distance from the target, the effect of the potential is negligible.
- (d) The scattered particle is detected at a distance far away from the scatterer.

#### 12.4.1. Relation between Scattering Amplitude and Differential Scattering Cross-section

Consider the scattering of a particle of mass  $m$  by a central potential  $V(\mathbf{r})$  (Fig. 12.6).



**Fig. 12.6**

$V(\mathbf{r})$  tends to zero more rapidly than  $1/r$  as  $r \rightarrow \infty$ .

Let  $E$  be the energy and  $p = \hbar k$  the initial momentum of the particle where  $\mathbf{k}$  is the wave-vector. Schrödinger equation for the central potential  $V(\mathbf{r})$  is

$$\left( -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \right) \psi_k(\mathbf{r}) = E \psi_k(\mathbf{r}) \quad \dots(1)$$

The wave-function  $\psi_k$  is written as a function of  $\theta, \phi$  and radial distance  $r$  *between the two particles*.

$$\psi_k = \psi_k(\mathbf{r}, \theta, \phi)$$

The scattering is determined by the asymptotic form of  $\psi_k(r, \theta, \phi)$  in the region where  $V = 0$ , when the colliding particles are far apart (or  $r \rightarrow \infty$ ).

$$\psi_k(r, \theta, \phi) = \lim_{r \rightarrow \infty} e^{i\mathbf{k}\cdot\mathbf{r}} + f(\omega) \frac{e^{ikr}}{r} \quad \dots(2)$$

**First Term:** The plane wave term  $e^{ikr}$  represents a wave of unit density and of current density  $\frac{\hbar k}{m}$ . It represents the incident beam of monoenergetic particles of momentum  $\hbar k$  and of unit density.

**Second Term:** The term  $\frac{f(\omega)}{r} e^{ikr}$  represent a wave of density  $\frac{|f(\omega)|^2}{r^2}$  and of current density  $\frac{|f(\omega)|^2 \hbar k}{r^2 m}$  directed along the direction  $\omega$  towards increasing  $r$  (outgoing-wave). It represents a beam of particles emitted radially from the scattering centre. It represent a beam of the scattered particles.

We can calculate the number of particles emitted per unit time into the solid angle  $d\omega$  located in the direction  $\omega$ .

The scattering wave-function is  $f(\omega) \frac{e^{ikr}}{r}$ .

Density of scattered particles is

$$\rho_s = \left| f(\omega) \frac{e^{ikr}}{r} \right|^2 = \frac{1}{r^2} |f(\omega)|^2 \quad \dots(3)$$

From Fig. 12.6, small elementary area  $= r \cdot r d\omega = r^2 d\omega$

The volume element between  $r$  and  $r + dr = r^2 d\omega \cdot dr$ .

$\rho_s$  is the number of scattered particles per unit volume.

$\therefore$  No. of particles in this elementary volume,

$$N_s = \rho_s r^2 d\omega dr.$$

Substituting value of  $\rho_s$  from Eq. (3), we get

$$N_s = \frac{|f(\omega)|^2}{r^2} \cdot r^2 d\omega dr = |f(\omega)|^2 dr d\omega \quad \dots(4)$$

$\therefore$  The number of scattered particles per unit time

$$\begin{aligned} \frac{dN_s}{dt} &= |f(\omega)|^2 d\omega \frac{dr}{dt} = |f(\omega)|^2 d\omega v = |f(\omega)|^2 d\omega \cdot \frac{\hbar k}{m} \\ \frac{dN_s}{dt} &= |f(\omega)|^2 \frac{\hbar k}{m} d\omega \end{aligned} \quad \dots(5)$$

If  $J$  represents the current density, then we have

$$J = \rho v. \quad \dots(6)$$

But a beam of particles is travelling in the same direction with velocity  $v$ . The flux of the beam is the number of particles crossing unit area (perpendicular to the beam) per unit time. These are number of particles in a volume of unit cross-section and length  $v$ .

For incident particles,  $\rho = 1$ .

$$\therefore J = v = \frac{\hbar k}{m} \quad \dots(7)$$

Let  $\sigma(\omega)$  represent differential scattering cross-section.

Then, number of particles scattered in solid angle  $d\omega$  per unit time is

$$J \sigma(\omega) d\omega = \frac{\hbar k}{m} \cdot \sigma(\omega) d\omega \quad [\text{using Eq. (7)}] \quad \dots(8)$$

Comparing Eqs. (5) and (8), we get

$$\frac{\hbar k}{m} \sigma(\omega) d\omega = |f(\omega)|^2 \frac{\hbar k}{m} d\omega$$

$$\therefore \sigma(\omega) = |f(\omega)|^2. \quad \dots(9)$$

Here  $f(\omega)$  is called the *scattering amplitude*.

i.e., The differential scattering cross section = (The scattering amplitude)<sup>2</sup>

The two techniques commonly employed to evaluate the scattering amplitude are:

(i) partial wave analysis and (ii) Born approximation.

Therefore, the total scattering cross section is

$$\sigma_{\text{total}} = \int |f(\omega)|^2 d\omega \quad \dots(10)$$

## 12.5 PARTIAL WAVE ANALYSIS

The method of partial waves is based upon an expansion of the wave function in terms of angular momentum eigenfunctions. It is applicable to spherically symmetric potentials. It is useful only for low-energy incident particles.

Consider a plane wave incident along  $z$ -axis in a region having interaction potential function  $V(r)$ .

Fig. 12.7 shows scattering from a localized potential.

scattering region (darker shading),

intermediate region (lighter shading).

Then total wave function may be expressed as

$$\Psi(r) = e^{ikz} + f(\theta) \frac{e^{ikr}}{r} \quad \dots(1)$$

incident                    scattered  
wave                    wave

It is a solution of three dimensional Schrödinger equation

$$\nabla^2 \Psi + \frac{2\mu}{\hbar^2} [E - V(r)] \Psi = 0 \quad \dots(2)$$

The solution of Eq. (2) may be expressed as

$$\Psi(r) = \sum_{l=0}^{\infty} R_l(r) Y_l^{(0)} \quad \dots(3)$$

This is a superposition of a number of waves. Each term in above equation is called a **partial wave**, corresponding to a particular value of  $l$ .

The functions  $R_l(r)$  are called radial wave functions.

$$\text{Put } \frac{2\mu E}{\hbar^2} = k^2 \quad \text{and} \quad \frac{2\mu V(r)}{\hbar^2} = U(r).$$

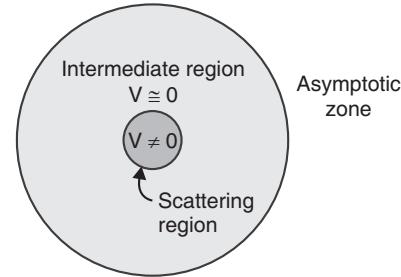
Eq. (2) becomes,

$$\nabla^2 \Psi + [k^2 - U(r)] \Psi = 0 \quad \dots(4)$$

As there is symmetry about polar axis i.e.,  $z$ -axis ( $m = 0$ ) and potential energy function does not involve  $\phi$ , the solution of Eq. (4) may be expressed as

$$\Psi(r, \theta, \phi) = \Psi(r, \theta) = \sum_l R_l(r) P_l(\cos \theta) \quad \dots(5)$$

$$\text{Put } \chi_l = r R_l(r) \quad \dots(6)$$



**Fig. 12.7**

Eq. (5) becomes,

$$\Psi(r, \theta) = \sum_l r^l \chi_l(r) P_l(\cos \theta) \quad \dots(7)$$

Here,  $\chi_l(r)$  satisfies the equation

$$\frac{d^2 \chi_l}{dr^2} + \left[ k^2 - U(r) - \frac{l(l+1)}{r^2} \right] \chi_l = 0 \quad \dots(8)$$

### (i) Asymptotic region (very large r)

In the asymptotic region,  $U$  and  $l$  terms in Eq. (8) may be neglected in comparison with  $k^2$ .

So Eq. (8) becomes,

$$\frac{d^2 \chi_l}{dr^2} + k^2 \chi_l = 0. \quad \dots(9)$$

or

$$\chi_l = e^{\pm ikr} \quad \dots(10)$$

Eq. (10) gives radial wave only.

### (ii) Intermediate region

For better approximation, we define a distance “ $a$ ” such that when  $r > a$ ,  $V(r) = 0$ .

But the  $l$  term in Eq. (8) cannot be ignored.

Then Eq. (8) becomes

$$\frac{d^2 \chi_l}{dr^2} + \left[ k^2 - \frac{l(l+1)}{r^2} \right] \chi_l = 0 \quad \dots(11)$$

This is spherical Bessel equation.

On solving Eq. (11), we get

$$R_l(r) = \frac{A_l}{kr} \sin \left( kr - \frac{l\pi}{2} + \delta_l \right) \quad \dots(12)$$

Here,  $A_l$  is a constant.

$\delta_l$  is called the *phase shift* of the  $l$ th partial wave caused by scattering potential  $U(r)$ .

The potential is spherically symmetric. So the wave function cannot depend on  $\phi$ .

Eq. (5) becomes,

$$\begin{aligned} \Psi(r, \theta) &= \sum_l \frac{A_l}{kr} \sin \left( kr - \frac{l\pi}{2} + \delta_l \right) P_l(\cos \theta) \\ \text{or} \quad \Psi(r, \theta) &= \sum_l \frac{A_l}{kr} \left\{ \frac{e^{i(kr - \frac{l\pi}{2} + \delta_l)} - e^{-i(kr - \frac{l\pi}{2} + \delta_l)}}{2i} \right\} \times P_l(\cos \theta) \end{aligned} \quad \dots(13)$$

Here,  $P_l$  is the  $l$ th Legendre polynomial.

This equation is identical with asymptotic form of Eq. (1), *i.e.*,

$$\Psi(r) = e^{ikz} + f(\theta) \frac{e^{ikr}}{r}$$

The asymptotic expansion of the plane wave

$$e^{ikz} = e^{ikr} \cos \theta$$

$$e^{ikz} = \sum_{l=0}^{\infty} (2l+1)i^l \left( \frac{1}{kr} \right) \sin \left( kr - \frac{1}{2} l\pi \right) P_l(\cos \theta) \quad \dots(14)$$

$$\begin{aligned}\therefore \Psi(r) &= \sum_l (2l+1)i^l \cdot \frac{1}{kr} \sin\left(kr - \frac{l\pi}{2}\right) P_l(\cos \theta) + f(\theta) \frac{e^{ikr}}{r} \\ \Psi(r) &= \sum_l (2l+1)i^l \cdot \frac{1}{kr} \left( \frac{e^{i\left(kr - \frac{l\pi}{2}\right)} - e^{-i\left(kr - \frac{l\pi}{2}\right)}}{2i} \right) P_l(\cos \theta) + r^{-1} f(\theta) e^{ikr} \quad \dots(15)\end{aligned}$$

Equations (13) and (15) represent the same function. Hence the coefficients of  $e^{ikr}$  and  $e^{-ikr}$  should be identical.

Equating the coefficients gives us

$$\sum_l A_l \frac{e^{i\left(-\frac{l\pi}{2} + \delta_l\right)}}{2i kr} P_l(\cos \theta) = \sum_l \frac{(2l+1) i^l e^{-il\pi/2}}{2i kr} \cdot P_l(\cos \theta) + r^{-1} \cdot f(\theta) \quad \dots(16)$$

$$\text{and } \sum_l A_l \frac{e^{-i\left(\frac{l\pi}{2} + \delta_l\right)}}{(2ikr)} \times P_l(\cos \theta) = \sum_l \frac{(2l+1) i^l e^{il\pi/2}}{2ikr} \cdot P_l(\cos \theta) \quad \dots(17)$$

From Eq. (17), we get

$$A_l = (2l+1) i^l e^{i\delta_l} \quad \dots(18)$$

Substituting this value of  $A_l$  in Eq. (16), we get

$$\begin{aligned}\sum_l \frac{(2l+1)i^l e^{i\delta_l} e^{i\left(-\frac{l\pi}{2} + \delta_l\right)}}{(2ikr)} P_l(\cos \theta) &= \sum_l \frac{(2l+1) i^l e^{-il\pi/2}}{2i kr} \cdot P_l(\cos \theta) + \frac{1}{r} f(\theta) \\ i^l &= e^{il\pi/2}. \quad (l \text{ is an integer}).\end{aligned}$$

$$f(\theta) = (2ik)^{-} \sum_{l=0}^{\infty} (2l+1) (e^{i\delta_l} - 1) P_l(\cos \theta)$$

$$\text{or scattering amplitude, } f(\theta) = \frac{1}{k} \sum_{l=0}^{\infty} (2l+1) e^{i\delta_l} \sin \delta_l P_l(\cos \theta) \quad \dots(19)$$

This gives  $f(\theta)$  as a sum of contributions from partial waves with  $l = 0$  to  $\infty$ .

The differential scattering cross-section,

$$\sigma(\theta) = |f(\theta)| = \frac{1}{k^2} \left| \sum_{l=0}^{\infty} (2l+1) P_l(\cos \theta) e^{i\delta_l} \sin \delta_l \right|^2 \quad \dots(20)$$

The total scattering cross-section,

$$\begin{aligned}\sigma_{\text{total}} &= 2\pi \int_0^\pi \sigma(\theta) \sin \theta d\theta \\ \sigma_{\text{total}} &= \frac{4\pi}{k^2} \sum_{l=0}^{\infty} (2l+1) \sin^2 \delta_l \quad \dots(21)\end{aligned}$$

(I used the orthogonality of the Legendre polynomials to do the angular integration.)

Equation (21) gives total scattering cross-section in terms of phase shifts.

A knowledge of the phase shifts  $\delta_l$  enables us to calculate the total scattering cross-section.

### Optical Theorem

The scattering amplitude  $f(\theta)$  is complex.

The total scattering cross-section can also be written in terms of the imaginary part of the *forward* scattering amplitude ( $\theta = 0$ ).

For  $\theta = 0$ ,  $P_l(\cos \theta) = P_l(\cos 0^\circ) = P_l(1) = 1$  for all values of  $l$ .

Then Eq. (19) for  $\theta = 0$  gives

$$f(0) = \frac{1}{k} \sum_l (2l+1) e^{i\delta_l} \sin \delta_l$$

$$\therefore \quad \text{Im } f(0) = \frac{1}{k} \sum_l (2l+1) \sin^2 \delta_l$$

Here,  $\text{Im } f(0)$  denotes the imaginary part of  $f(0)$ .

Equation (21) for total scattering cross-section can be written as

$$\sigma_{\text{total}} = \frac{4\pi}{k} \text{Im } f(0) \quad \dots(22)$$

Eq. (22) is called the *optical theorem*.

It relates the imaginary part of the forward ( $\theta = 0$ ) scattering amplitude to the total scattering cross-section.

### Phase Shifts

The total scattering cross-section is given by

$$\sigma_{\text{total}} = \frac{4\pi}{k^2} \sum_{l=0}^{\infty} (2l+1) \sin^2 \delta_l$$

Thus phase shift is a complete measure of scattering.

- The scattering cross-section vanishes if  $\delta_l = 0$  or  $\pi$ .
- The scattering cross-section is maximum if the value of  $\delta_l = \pm \frac{\pi}{2}, \pm \frac{3\pi}{2}$  etc.
- Each term in the sum is a contribution of the  $l$ th partial wave.
- The phase shift  $\delta_l$  gives the phase difference of the radial function (of a given angular momentum  $l$ ) from the no scattering case ( $\delta_l = 0$ ).
- The phase shift is a function of energy, *i.e.*,  $k$ .

### EXERCISE

1. What is meant by differential scattering cross-section, the total scattering cross-section and the scattering amplitude? *(Rohilkhand 1997)*
2. Relate the laboratory and centre of mass co-ordinates systems for a scattering process. *(Rohilkhand 1977)*
3. Give the theory of reducing two-body problem into a one body problem. *(Madras, 2006)*
4. Find the relationship between differential scattering cross-section and scattering amplitude. *(Madras, 2006)*
5. Explain the method of partial waves to calculate the phase shifts and scattering amplitude. *(Agra 1982)*
6. What is meant by scattering cross-section? Show that the scattering amplitude by the method of partial waves is given by

$$f(\theta) = \frac{1}{k} \sum_{l=0}^{\infty} (2l+1) e^{i\delta_l} \sin \delta_l P_l(\cos \theta) \quad \quad \quad \text{(Rohilkhand 1994)}$$

7. State optical theorem.

## PERTURBATION THEORY

---

**After reading this chapter, you should be able to**

- ◆ Develop stationary perturbation theory and apply it to illustrative problems.

### 13.1 TIME-INDEPENDENT PERTURBATION THEORY FOR A NON-DEGENERATE ENERGY LEVEL

In the present chapter, we are concerned with the approximate determination of the discrete eigenenergies and corresponding eigenfunctions for the *stationary states* of a time-independent Hamiltonian. Perturbation theory studies the changes induced in a system by a small disturbance. Perturbation theory analyses the modifications of discrete energy levels and of the corresponding eigenfunctions of a system when a perturbation is applied.

If corresponding to each energy level or eigenvalue there is only one wavefunction, the system is called *non-degenerate*.

Schördinger wave equation in perturbed state is

$$H\psi = E\psi, \quad \dots(1)$$

$H$  represents the operator

$$H = \left( -\frac{\hbar^2}{2m} \nabla^2 + V \right) \quad \dots(2)$$

The Hamiltonian  $H$  can be expressed as

$$H = H^{(0)} + \lambda H^{(1)} + \lambda^2 H^{(2)} + \dots \quad \dots(3)$$

The quantity  $\lambda$  is a real parameter.

$H^{(0)}$  is unperturbed Hamiltonian and other terms are perturbation terms.

$H^{(0)}$  is large compared with  $H^{(1)}$ , i.e., the energy associated with  $H^{(0)}$  is large compared with that associated with  $H^{(1)}$ .

Let the eigenvalues and eigenfunctions of the unperturbed problem be  $E_1^{(0)}, E_2^{(0)}, E_3^{(0)}, \dots, E_n^{(0)}$ , ..., and  $\psi_1^{(0)}, \psi_2^{(0)}, \psi_3^{(0)}, \dots, \psi_n^{(0)}, \dots$ , respectively.

$$H^{(0)}\psi_n^{(0)} = E_n^{(0)}\psi_n^{(0)}, \quad n = 1, 2, 3, \dots \quad \dots(4)$$

Eq. (4) is the wave equation for the unperturbed system.

$E_n^{(0)}$  is the energy of the  $n$ th level of the system.

We want to solve the eigenvalue problem

$$H\psi_n = E_n\psi_n, \quad \dots(5)$$

Here,  $E_n$  are the energy eigenvalues of the modified Hamiltonian.

We assume that both the eigenvalues and the eigenfunctions of  $H$  can be expanded in powers of the perturbation parameter  $\lambda$ :

$$E_n = E_n^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \dots \quad \dots(6)$$

$$\psi_n = \psi_n^{(0)} + \lambda \psi_n^{(1)} + \lambda^2 \psi_n^{(2)} + \dots \quad \dots(7)$$

The power of  $\lambda$  is the order of the perturbation.

- $E_n^{(1)}, E_n^{(2)}, \dots$ , and  $\psi_n^{(1)}, \psi_n^{(2)}, \dots$ , are quantities to be determined.

Substituting Eqs. (3), (6) and (7) into the Schrödinger equation (5), we have

$$(H^{(0)} + \lambda H^{(1)}) (\psi_n^{(0)} + \lambda \psi_n^{(1)} + \lambda^2 \psi_n^{(2)} + \dots) = (E_n^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \dots) (\psi_n^{(0)} + \lambda \psi_n^{(1)} + \lambda^2 \psi_n^{(2)} + \dots) \quad \dots(8)$$

$$\text{or } H^{(0)}\psi_n^{(0)} + \lambda(H^{(0)}\psi_n^{(1)} + H^{(1)}\psi_n^{(0)}) + \lambda^2(H^{(0)}\psi_n^{(2)} + H^{(1)}\psi_n^{(1)} + \dots) + \dots = E_n^{(0)}\psi_n^{(0)} + \lambda(E_n^{(0)}\psi_n^{(1)} + E_n^{(1)}\psi_n^{(0)}) + \lambda^2(E_n^{(0)}\psi_n^{(2)} + E_n^{(2)}\psi_n^{(0)}) + \lambda^3(E_n^{(0)}\psi_n^{(3)} + \dots) + \dots \quad \dots(9)$$

In the above substitution, we have limited  $H$  to two terms, namely,  $H^0$  the unperturbed and major part and  $H^{(1)}$  the perturbed and small part.

- We now equate the coefficients of equal powers of  $\lambda$  on both sides of Eq. (9).

$$H^{(0)}\psi_n^{(0)} = E_n^{(0)}\psi_n^{(0)} \quad \dots(10)$$

$$H^{(0)}\psi_n^{(1)} + H^{(1)}\psi_n^{(0)} = E_n^{(0)}\psi_n^{(1)} + E_n^{(1)}\psi_n^{(0)} \quad \dots(11)$$

Eqs. (10) and (11) represent unperturbed, and first order perturbation equations respectively.

**Evaluation of first order energy  $E_n^{(1)}$ .** The first order perturbation equation is

$$H^{(0)}\psi_n^{(1)} + H^{(1)}\psi_n^{(0)} = E_n^{(0)}\psi_n^{(1)} + E_n^{(1)}\psi_n^{(0)}$$

The unknown functions  $\psi_n^{(1)}$  can be expanded in terms of the known functions

$$\psi_1^{(0)}, \psi_2^{(0)}, \psi_3^{(0)}, \dots, \psi_m^{(0)}, \dots$$

as

$$\psi_n^{(1)} = \sum_{m=0}^{\infty} C_m \psi_m^{(0)} \quad \dots(12)$$

Substituting Eq. (12) in Eq. (11), we have

$$\sum_m C_m H^{(0)}\psi_m^{(0)} + H^{(1)}\psi_n^{(0)} = \sum_m C_m E_n^{(0)}\psi_m^{(0)} + E_n^{(1)}\psi_n^{(0)}. \quad \dots(13)$$

But

$$\begin{aligned} H^{(0)}\psi_m^{(0)} &= E_m^{(0)}\psi_m^{(0)} \\ \therefore \sum_m C_m E_m^{(0)}\psi_m^{(0)} - \sum_m C_m E_n^{(0)}\psi_m^{(0)} + H^{(1)}\psi_n^{(0)} &= E_n^{(1)}\psi_n^{(0)}, \end{aligned}$$

or

$$\sum_m C_m [E_m^{(0)} - E_n^{(0)}]\psi_m^{(0)} + H^{(1)}\psi_n^{(0)} = E_n^{(1)}\psi_n^{(0)}. \quad \dots(14)$$

In order to obtain the first-order energy correction  $E_n^{(1)}$ , we premultiply Eq. (14) by  $\psi_n^{(0)*}$  and integrate over all coordinates. This gives

$$\int \sum_m \psi_n^{(0)*} C_m [E_m^{(0)} - E_n^{(0)}] \psi_m^{(0)} d\tau + \int \psi_n^{(0)*} H^{(1)} \psi_n^{(0)} d\tau = \int E_n^{(1)} \psi_n^{(0)*} \psi_n^{(0)} d\tau \quad \dots(15)$$

$$\text{We know that } \left. \begin{aligned} \int \psi_i^{(0)*} \psi_j^{(0)} d\tau &= 0, \text{ if } i \neq j \\ &= 1, \text{ if } i = j \end{aligned} \right\} \quad \dots(16)$$

Applying Eq. (16) to Eq. (15), we have

$$0 + \int \psi_n^{(0)*} H^{(1)} \psi_n^{(0)} d\tau = E_n^{(1)}$$

or

$$E_n^{(1)} = \int \psi_n^{(0)*} H^{(1)} \psi_n^{(0)} d\tau. \quad \dots(17)$$

$\therefore$

$$E_n^{(1)} = \langle \psi_n^{(0)} | H^{(1)} | \psi_n^{(0)} \rangle. \quad \dots(18)$$

This is a very *important result*. It tells us that the first-order correction to the energy for a non-degenerate level is just the perturbation  $H^{(1)}$  averaged over the corresponding unperturbed state of the system.

- First - order energy correction for a non-degenerate system is just the expectation value of first order Hamiltonian  $H^{(1)}$  over the unperturbed state, here  $n$ th state of the system.

Substituting the value of  $E_n^{(1)}$  in Eq. (6), we have

$$E_n = E_n^{(0)} + \lambda \int \psi_n^{(0)*} H^{(1)} \psi_n^{(0)} d\tau$$

**Evaluation of first order wavefunction  $\psi_n^{(1)}$**

We premultiply Eq. (14) by  $\psi_m^{(0)*}$  and integrate over all space.

$$\int \sum_m C_m [E_m^{(0)} - E_n^{(0)}] \psi_m^{(0)*} \psi_m^{(0)} d\tau + \int \psi_m^{(0)*} H^{(1)} \psi_m^{(0)} d\tau = \int E_n^{(1)} \psi_m^{(0)*} \psi_m^{(0)} d\tau \quad \dots(19)$$

Applying condition (16), we have

$$C_m [E_m^{(0)} - E_n^{(0)}] + \int \psi_m^{(0)*} H^{(1)} \psi_m^{(0)} d\tau = 0$$

or

$$C_m = -\frac{\int \psi_m^{(0)*} H^{(1)} \psi_n^{(0)} d\tau}{E_m^{(0)} - E_n^{(0)}}, m \neq n. \quad \dots(20)$$

Substituting the value of  $C_m$  in Eq. (12), we have

$$\psi_n^{(1)} = -\sum_{m=0}^{\infty}' \frac{\int \psi_m^{(0)*} H^{(1)} \psi_n^{(0)} d\tau}{E_m^{(0)} - E_n^{(0)}} \psi_m^{(0)} \quad \dots(21)$$

Here,  $\Sigma'$  indicates the summation over the values of  $m$  from 0 to  $\infty$ , excluding  $m = n$ .

$$\psi_n^{(1)} = -\sum_{m=0}^{\infty}' \frac{\langle \psi_m^{(0)} | H^{(1)} | \psi_n^{(0)} \rangle}{E_m^{(0)} - E_n^{(0)}} \psi_m^{(0)}$$

Substituting the value of  $\psi_n^{(1)}$  in Eq. (7), we have

$$\psi_n = \psi_n^{(0)} - \lambda \sum_{m=0}^{\infty}' \frac{\int \psi_m^{(0)*} H^{(1)} \psi_n^{(0)} d\tau}{E_m^{(0)} - E_n^{(0)}} \psi_m^{(0)}.$$

## 13.2 APPLICATION OF NON-DEGENERATE PERTURBATION THEORY NORMAL HELIUM ATOM (GROUND STATE OF HELIUM)

Consider the normal state of helium atom.

The helium atom contains a nucleus of charge  $Ze$  and two electrons each having a charge  $-e$  (Fig. 13.1).

The potential energy of the system is given by

$$V = \left( -\frac{Ze^2}{r_1} - \frac{Ze^2}{r_2} + \frac{e^2}{r_{12}} \right) \frac{1}{4\pi\epsilon_0} \quad \dots(1)$$

- $r_1$  and  $r_2$  are the distances of electrons 1 and 2, respectively from nucleus.
- $r_{12}$  is the distance between the two electrons.
- $m$  is the electron mass.
- We shall neglect the effects of spin-orbit interaction and the motion of the nucleus.

The Schrodinger wave equation for the two electrons is

$$\left( \frac{\partial^2 \psi}{\partial x_1^2} + \frac{\partial^2 \psi}{\partial y_1^2} + \frac{\partial^2 \psi}{\partial z_1^2} + \frac{\partial^2 \psi}{\partial x_2^2} + \frac{\partial^2 \psi}{\partial y_2^2} + \frac{\partial^2 \psi}{\partial z_2^2} \right) + \frac{2m}{\hbar^2} \left[ E + \frac{1}{4\pi\epsilon_0} \left( \frac{Ze^2}{r_1} + \frac{Ze^2}{r_2} - \frac{e^2}{r_{12}} \right) \right] \psi = 0. \quad \dots(2)$$

The variables  $x_1, y_1, z_1$  are cartesian coordinates of one electron and  $x_2, y_2, z_2$  those of the other.

- If the term  $\frac{e^2}{4\pi\epsilon_0 r_{12}}$  is omitted, the wave equation can be solved exactly.

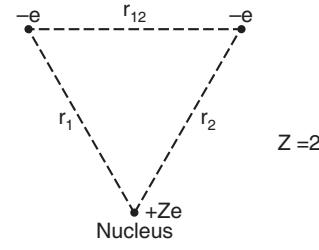
Hence this term is regarded as perturbation term.

$$H^{(1)} = \frac{e^2}{r_{12}} \frac{1}{4\pi\epsilon_0} \quad \dots(3)$$

The perturbation is the electrostatic repulsion between the two electrons.

The unperturbed wave equation can be separated into two equations by substituting

$$\psi^{(0)}(x_1 y_1 z_1, x_2 y_2 z_2) = u_1^{(0)}(x_1 y_1 z_1) u_2^{(0)}(x_2 y_2 z_2). \quad \dots(4)$$



**Fig. 13.1**

Considering the wavefunctions in polar coordinates  $r_1, \theta_1, \phi_1$  and  $r_2, \theta_2, \phi_2$  and in normal state,

$$\psi_{100,100}^{(0)} = \psi_{100}^{(0)}(r_1, \theta_1, \phi_1) \psi_{100}^{(0)}(r_2, \theta_2, \phi_2) = u_{1s}^{(0)}(r_1, \theta_1, \phi_1), u_{1s}^{(0)}(r_2, \theta_2, \phi_2)$$

The corresponding eigenvalue is

$$E_{100,100}^{(0)} = E_1^{(0)} + E_2^{(0)} = -2Z^2 E_H. \quad \dots(5)$$

The first order perturbed energy  $E^{(1)}$  is the average value of the perturbation function  $H^{(1)} = \frac{e^2}{4\pi\epsilon_0 r_{12}}$  over the unperturbed state of the system.

$$\therefore E^{(1)} = \int \psi_n^{(0)*} H^{(1)} \psi_n^{(0)} d\tau = \frac{1}{4\pi\epsilon_0} \int \frac{e^2}{r_{12}} \psi_{100,100}^{(0)} d\tau. \quad \dots(6)$$

$$\text{We know that } u_{1s}^{(0)} = \psi_{100}^{(0)} = \sqrt{\left(\frac{Z^3}{\pi a_0^3}\right)} e^{-\rho/2} \quad \dots(7)$$

$$\text{Here, } \rho = \frac{2Zr}{a_0} \text{ and } a_0 = \frac{\hbar^2}{me^2} (4\pi\epsilon_0).$$

$$\psi_{100,100}^{(0)} = \frac{Z^3}{\pi a_0^3} \exp(-\rho_1/2) \exp(-\rho_2/2). \quad \dots(8)$$

Volume element for two electrons in spherical polar coordinates  $(r, \theta, \phi)$  is

$$d\tau = r_1^2 dr_1 \sin \theta_1 d\theta_1 d\phi_1 r_2^2 dr_2 \sin \theta_2 d\theta_2 d\phi_2, \quad \dots(9)$$

The integral for  $E^{(1)}$  becomes

$$E^{(1)} = \frac{Ze^2}{(4\pi\epsilon_0)^2 \pi^2 a_0} \int_0^\infty \int_0^\pi \int_0^{2\pi} \int_0^\infty \int_0^\pi \int_0^{2\pi} \frac{\exp(-\rho_1 - \rho_2)}{\rho_{12}} \rho_1^2 d\rho_1 \sin \theta_1 d\theta_1 d\phi_1 \rho_2^2 d\rho_2 \sin \theta_2 d\theta_2 d\phi_2 \quad \dots(10)$$

$$\text{Here, } \rho_{12} = 2Zr_{12}/a_0.$$

The value of the integral is calculated by considering the electrostatic interaction energy of two spherically symmetrical distributions of electricity, with density functions  $\exp(-\rho_1)$  and  $\exp(-\rho_2)$ , respectively.

$$E^{(1)} = (5/4) Z E_H. \quad \dots(11)$$

Thus, the total energy of the system is

$$E = -\left(2Z^2 - \frac{5}{4}Z\right) E_H. \quad \dots(12)$$

Thus, we can calculate the energy of He, Li<sup>+</sup>, Be<sup>++</sup>, etc., with  $Z=2, 3, 4$  etc., respectively.

## EXERCISE

1. (a) Give the first order perturbation theory for a non-degenerate case. Show that the first order perturbation energy for a non-degenerate state of a system is just the perturbation function averaged over the corresponding unperturbed state of the system. Derive an expression for the eigen function of the perturbed system.  
*(Agra 1992, Nagpur 95)*
1. (b) Apply the perturbation theory to derive the energy of the helium atom in its normal state.  
*(Nagpur 1995, 87)*
2. What is the principle of perturbation theory?  
*(Rohilkhand 2001)*

**P A R T - IV**

# **MOLECULAR PHYSICS**

**[Molecular Spectra, Atomic Spectra and X-ray Spectra]**

**14. Molecular Spectra**

**15. Spectra of Alkali and Alkaline Earth Elements**

**16. X-Ray Spectra**

## MOLECULAR SPECTRA

**After reading this chapter, you should be able to**

- ◆ Discuss the different types of molecular spectra, their origin, characteristics, methods of production and their applications.
- ◆ Illustrate using a diagram, the electronic, rotational, and vibrational energy levels of a diatomic molecule.
- ◆ Explain why energy levels in rotational spectra are not equally spaced.
- ◆ Explain why atomic spectra consists of lines whereas molecular spectra at room temperature are broad and continuous.
- ◆ Justify the difference in intensity between Stokes and anti-Stokes lines.
- ◆ Compare and contrast the advantages of IR and Raman spectroscopy.
- ◆ List the salient features observed in the electronic spectrum of a diatomic molecule.

## 14.1 INTRODUCTION

Apart from the ground state, a molecule can be in a higher energy state, and transitions between the various energy levels give rise to the observed molecular spectra.

The energy of a diatomic molecule arises from three modes:

- (i) the electronic configuration of the electrons in the molecule,
- (ii) the vibration of the atoms about the equilibrium position, and
- (iii) the rotation of the molecule as a whole about its centre of mass.

The total energy of a molecule can be expressed as the sum of three independent terms:

$$E = E_e + E_v + E_r \quad \dots(1)$$

$E_e, E_v, E_r$  are quantised and  $E_v > E_r$ .

If the molecule remains in its ground state level of electronic and vibrational energies, and if it suffers transitions between different rotational energies, we get the *pure rotational spectrum* of the molecule. These spectra are in the microwave and far infrared regions.

If the molecule remains in its ground state level of electronic energy, and if transitions occur between different vibrational and rotational energies, we get the *vibration-rotation spectrum* of the molecule. These spectra are in the near infrared region.

If transitions occur between different electronic-vibrational-rotational energy levels, we get the *electronic spectrum* of the molecule. These spectra are in the visible and ultraviolet regions.

### 14.1.1. Origin of Molecular Spectra

*The molecular spectra arise due to transitions between allowed energy states of molecules.*

Molecular spectra (band spectra) are produced due to excitation of a substance in the *molecular state*.

The transition between two energy states of a molecule, as a result of absorption or emission of energy will give rise to a spectral line. When the molecule suffers a transition from higher energy state with energy  $E$  to a lower energy state with energy  $E'$ , the difference of energy  $\Delta E = (E - E')$  is emitted as radiation of frequency  $\nu$  given as

$$\nu = \frac{E - E'}{h}.$$

For each allowed transition, we get one line in the spectrum of the molecule. Similarly we have the frequency in the absorption spectrum when the molecule suffers a transition from lower energy state to higher energy state.

### 14.1.2. Nature of Molecular Spectra

Molecular spectrum is a band spectrum. Usually a band has a sharp intense edge on one side called '*band head*' and gradually decreases in intensity towards the other side.

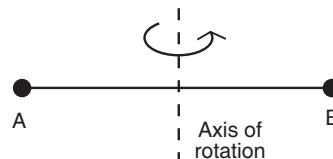
With high resolving power instruments, band spectra are found to have threefold structure:

- (i) Each band is composed of a large number of lines, which are crowded together at the long wavelength side called '*band head*'. Going away from the band head, the lines become more widely spaced and also weaker.
- (ii) There is a regular sequence in which several bands follow one another, and thus constitute a group of bands.
- (iii) These different groups of bands are quite close and thus because of their regular arrangement, form a band system.

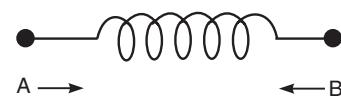
#### 14.1.3. Different Modes of Molecular Excitation

When a molecule is supplied with energy, it may be excited in any of the following modes :

1. The molecule as a whole may rotate about an axis passing through the centre of mass (Fig. 23.1).



**Fig. 14.1**



**Fig. 14.2**

2. The molecule may make vibrational motion. *i.e.*, the constituent atoms may vibrate relative to one another (Fig. 14.2).

- In a diatomic molecule, the two atoms usually vibrate along the line joining them, somewhat like a linear harmonic oscillator.
- We may think of a molecule as an arrangement of masses held together by massless springs. The masses represent the atomic nuclei and springs represent interatomic forces.
- 3. There may be changes produced in its electronic configuration.

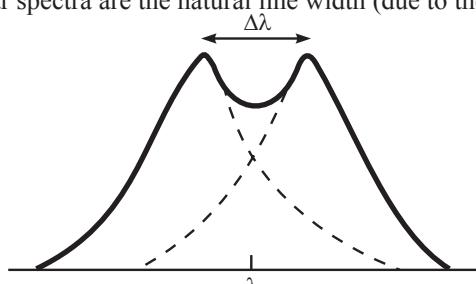
#### 14.1.4. Factors Affecting Line Width of Molecular Spectra

When a molecule makes transition from a higher energy state to a lower energy state, molecular emission spectrum is produced. The spectrum is found to have some line-width, even if all attempts are made to narrow down the slit.

The factors that affect the line width of molecular spectra are the natural line width (due to the uncertainty principle), Doppler effect at moderate frequencies and intermolecular collisions (collision broadening) at lower frequencies.

Ideally, the analysis of any spectrum requires that the lines be well separated or resolved. In addition to the various mechanisms that can cause line broadening, lines can also be "broadened" by the measuring instrument itself. We define *resolving power* of an instrument as a measure of its ability to distinguish lines that overlap. It is usually expressed as the ratio of the observed wavelength (frequency) to the smallest difference between two wavelengths (frequencies) that can be measured, that is,  $\lambda / \Delta\lambda$  ( $v / \Delta v$ ) (Fig. 14.3).

Finally, rate processes such as dissociation, rotation, and electron and proton transfer reactions can also cause the broadening.



**Fig. 14.3**

#### 14.1.5. Factors Affecting Intensity of Molecular Spectra

The following three factors affect the intensity of molecular spectra.

1. *The transition probability.* *i.e.*, the likelihood of a system in one state changing to another state.

2. *The population of molecules.* i.e., the number of molecules initially in the state from which the transition occurs. In general, the intensity of the absorption lines is proportional to the population in the lower state.

3. *The concentration or path length of the sample.* i.e., the amount of material present, giving rise to the spectrum.

#### 14.1.6. Born-Oppenheimer Approximation

So far we have considered rotation and vibration quite separately. The energies of the two motions are very different. We may, as a first approximation, consider that a diatomic molecule can execute rotations and vibrations quite independently. This is called the Born-Oppenheimer Approximation.

#### 14.1.7. Rotation of Linear System

The simplest molecule is a diatomic molecule.  
*Example :* HCl, CO. A diatomic molecule is an example of a linear system. In a diatomic molecule, two atoms are bound by attractive forces. It can be regarded as a rigid dumb-bell. To a first approximation, it is assumed that the distance between two nuclei remains constant, i.e., the molecule behaves as a *rigid rotator* (Fig. 14.4).

The three directions of rotation may be taken as

- (i) about the bond axis
- (ii) end-over-end rotation in the plane of the paper and
- (iii) end-over-end rotation at right angles to the plane.

For linear molecules, the moments of inertia of (ii) and (iii) are equal and that of (i) is zero.

$$I_a = I_b \text{ and } I_c = 0.$$

The principal axes of inertia are conventionally given the subscripts of *a*, *b*, and *c*, which may be thought of as the *x*, *y*, and *z* directions.

#### Microwave Spectroscopy (Pure Rotational Spectra)

- Spectroscopy in the microwave region is concerned with the study of rotating molecules.

### 14.2 THEORY OF THE ORIGIN OF PURE ROTATIONAL SPECTRUM OF A MOLECULE

#### [Diatomic Molecule as Rigid Rotator]

We start with this, the simplest of all linear molecules, shown in Fig. 14.5.

We may picture a diatomic molecule as consisting of atoms of masses  $m_1$  and  $m_2$  a distance  $R$  apart. The diatomic molecule can rotate about its centre of mass as shown in Fig. 14.6.

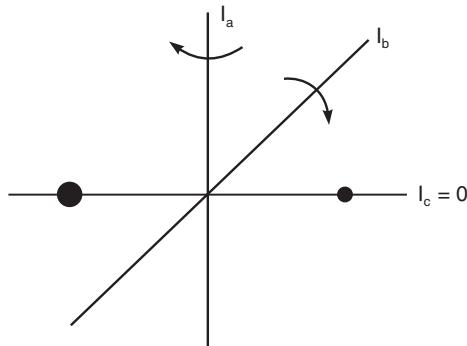


Fig. 14.4

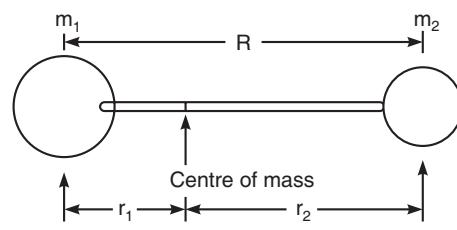


Fig. 14.5

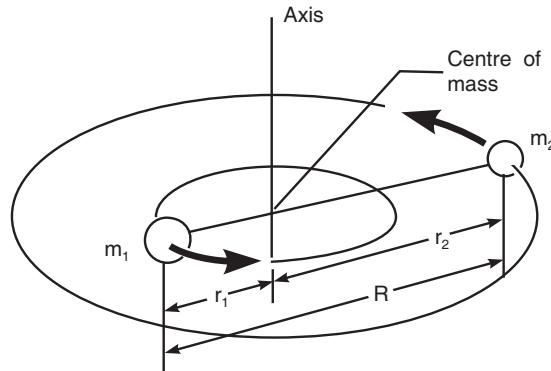


Fig. 14.6

The moment of inertia of this molecule about an axis passing through its centre of mass and perpendicular to a line joining the atoms is

$$I = m_1 r_1^2 + m_2 r_2^2 \quad \dots(1)$$

Here,  $r_1$  and  $r_2$  are the distances of  $m_1$  and  $m_2$ , respectively from the centre of mass.

From the definition of centre of mass

$$m_1 r_1 = m_2 r_2 \quad \dots(2)$$

$$\text{Further} \quad R = r_1 + r_2 \quad \dots(3)$$

From equations (2) and (3)

$$r_1 = \frac{m_2}{m_1 + m_2} R \text{ and } r_2 = \frac{m_1}{m_1 + m_2} R.$$

Putting these values of  $r_1$  and  $r_2$  in Eq. (1),

$$I = \frac{m_1 m_2^2}{(m_1 + m_2)^2} R^2 + \frac{m_2 m_1^2}{(m_1 + m_2)^2} R^2 = \frac{m_1 m_2}{(m_1 + m_2)} R^2$$

or

$$I = m' R^2 \quad \dots(4)$$

Here, the reduced mass  $m'$  is defined as

$$m' = \frac{m_1 m_2}{(m_1 + m_2)} \quad \dots(5)$$

The angular momentum  $L$  of the molecule has the magnitude  $L = I\omega$

The angular momentum of the rotating diatomic molecule is quantised according to

$$L = \sqrt{J(J+1)} \hbar \quad J = 0, 1, 2, 3, \dots \quad \dots(6)$$

$J$  is called the rotational quantum number.

The rotational kinetic energy of a diatomic molecule is  $\frac{1}{2} I\omega^2$ .

$$\therefore E_J = \frac{1}{2} I\omega^2 = \frac{L^2}{2I} = \frac{J(J+1)\hbar^2}{2I} \quad \dots(7)$$

Rotational transitions occur only in those molecules which possess a permanent electric dipole moment. For this reason nonpolar diatomic molecules such as  $H_2$  and symmetric polyatomic molecules such as  $CO_2$  and  $CH_4$  do not exhibit rotational spectra.

The selection rule for rotational transitions is  $\Delta J = \pm 1$   $\dots(8)$

Rotational spectra are always obtained in absorption. Hence for a transition from some initial state of quantum number  $J$  to the next higher state of quantum number  $J+1$ , the frequency of the absorbed photon is

$$\nu_{J \rightarrow J+1} = \frac{\Delta E}{h} = \frac{E_{J+1} - E_J}{h} = \frac{\hbar}{2\pi I} (J+1) \quad \dots(9)$$

The spectrum of a rigid molecule therefore consists of equally spaced lines (Fig. 14.7).

**Importance of the pure rotational spectrum for determining the properties of a molecule:** The frequency difference between two successive lines in the pure rotational spectrum of a diatomic molecule is given by

$$\Delta\nu = \nu_{J+1 \rightarrow J+2} - \nu_{J \rightarrow J+1} = \frac{\hbar}{2\pi I}.$$

By measuring  $\Delta\nu$ , the moment of inertia  $I$  of the molecule can be calculated. Now,

$$I = m' R^2 = \left\{ \frac{m_1 m_2}{(m_1 + m_2)} \right\} R^2$$

Thus the distance  $R$  between the nuclei can be calculated. This gives the length of the chemical bond between the atoms.

**EXAMPLE 1:** In the CO molecule the wavenumber difference between the successive absorption lines in the pure rotational spectrum is  $384 \text{ m}^{-1}$ . Calculate the moment of inertia of the molecule and the equilibrium bond length of the molecule. Masses of the  $C^{12}$  and  $O^{16}$  atoms are respectively  $1.99 \times 10^{-26} \text{ kg}$  and  $2.66 \times 10^{-26} \text{ kg}$ .

**SOL.** Here,

$$\Delta\bar{\nu} = \frac{\hbar}{2\pi Ic} = 384 \text{ m}^{-1}$$

$$\therefore I = \frac{\hbar}{2\pi c \Delta\bar{\nu}} = \frac{1.054 \times 10^{-34}}{2\pi \times (3 \times 10^8) \times 384} = 1.456 \times 10^{-46} \text{ kg m}^2$$

The reduced mass  $m'$  of the CO molecule is

$$m' = \frac{m_1 m_2}{m_1 + m_2} = \left[ \frac{1.99 \times 2.66}{1.99 + 2.66} \right] \times 10^{-26} \text{ kg} = 1.138 \times 10^{-26} \text{ kg}.$$

$$\therefore \text{Bond length} = R = \sqrt{\frac{I}{m'}} = \sqrt{\frac{1.456 \times 10^{-46}}{1.138 \times 10^{-26}}} = 0.1131 \text{ nm}.$$

**EXAMPLE 2:** The lines in the pure rotational spectrum of HCl are spaced as  $20.8 \times 10^2 \text{ per metre}$ . Calculate the moment of inertia and the internuclear distance. Mass of proton =  $1.67 \times 10^{-27} \text{ kg}$ ; mass of chlorine =  $58.5 \times 10^{-27} \text{ kg}$ .

**SOL.** The reduced  $m'$  is defined as

$$m' = \frac{m_1 m_2}{m_1 + m_2} = \frac{(1.67 \times 10^{-27})(58.5 \times 10^{-27})}{(1.67 + 58.5) \times 10^{-27}} = 1.62 \times 10^{-27} \text{ kg}$$

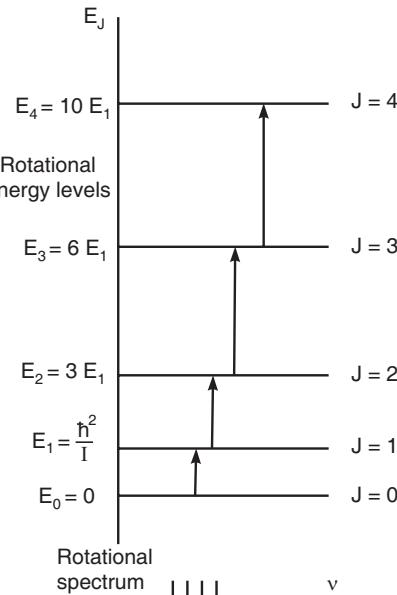


Fig. 14.7

Here,

$$\Delta \bar{v} = \frac{\hbar}{2\pi I c} = 20.8 \times 10^2 \text{ m}^{-1}$$

∴

$$I = \frac{\hbar}{2\pi c \Delta \bar{v}} = \frac{1.054 \times 10^{-34}}{2\pi(3 \times 10^8)(20.8 \times 10^2)} = 2.689 \times 10^{-47} \text{ kg m}^2$$

∴

$$R = \sqrt{\frac{I}{m'}} = \sqrt{\frac{2.689 \times 10^{-47}}{1.62 \times 10^{-27}}} \text{ m} = 1.286 \times 10^{-10} \text{ m.}$$

### 14.3 NON-RIGID ROTATOR [DIATOMIC MOLECULE AS NON-RIGID ROTATOR]

On the wave number scale, the frequency difference between two successive lines in the pure rotational spectrum of a diatomic molecule is given by

$$\Delta \bar{v} = \frac{\hbar}{2\pi I c}.$$

The rotational spectrum can be recorded. The absorption lines are equi-spaced. The separation between the adjacent lines is identified as  $2B$ .

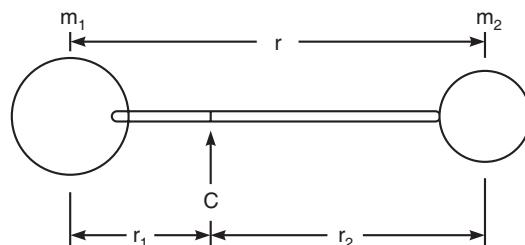
$$\Delta \bar{v} = 2B = \frac{\hbar}{2\pi I c}$$

∴

$$B = \frac{\hbar}{4\pi I c}.$$

By measuring  $\Delta \bar{v}$ , the rotational constant  $B$  can be calculated. From this, the moment of inertia of the molecule  $I$  can be calculated.

$I = m' r^2$  where  $m'$  is the reduced mass of the molecule and  $r$  the bond length. (Fig. 14.8).



**Fig. 14.8** A diatomic molecule treated as two masses,  $m_1$  and  $m_2$ , joined by a non-rigid bar of length  $r = r_1 + r_2$

Knowing the value of  $m'$ , we can calculate  $r$ .

Experimental investigations have shown that the successive lines in the far infra-red spectrum are not evenly spaced, but that the frequency separation decreases slightly with increasing value of  $J$  (larger speed of rotation). It shows that bond length ( $r$ ) increases with  $J$ . Therefore, our assumption that the molecule is a rigid rotator is false. In fact, all bonds are elastic to some extent. The more quickly a diatomic molecule rotates, the greater is the centrifugal force tending to move the atoms apart.

Now we discuss the consequences of the change in bond length with  $J$ .

(i) When a bond is elastic, it will stretch and compress periodically with a certain functional frequency dependent upon the masses of the atoms and the elasticity (or force constant  $k$ ) of the bond. This means that the molecule may have vibrational energy.

If the vibrational motion is simple harmonic, the force constant  $k$  is given by

$$k = 4\pi^2 \omega^2 c^2 m, \quad \dots(1)$$

Here,  $\omega$  is the vibration frequency (in  $m^{-1}$ ).  $m$  is the reduced mass of the molecule.

The variation of  $B$  with  $J$  is determined by the force constant, *i.e.*, the weaker the bond, more readily will it distort under centrifugal forces.

(ii) The second consequence of elasticity is that the quantities  $r$  and  $B$  vary during a vibration.

### Energy Levels

Consider a single particle of mass  $m$  rotating about a fixed point with an angular velocity,  $\omega$ . Let the particle be at a distance  $r_0$  from the fixed point when there is no rotation. Let this length increase to  $r$ , when the particle rotates.

Centrifugal force during rotation =  $m r \omega^2$ .

Restoring force due to bond stretching =  $k(r - r_0)$ .

The above two forces balance each other at any instant of rotation.

$$\therefore k(r - r_0) = m r \omega^2 \quad \dots(2)$$

$$\text{or } r = \frac{k r_0}{(k - m \omega^2)} \quad \dots(3)$$

This gives the distorted bond length.

Total energy of the rotating system =  $K.E. + P.E.$

$$\begin{aligned} E &= \frac{1}{2} I \omega^2 + \frac{1}{2} k (r - r_0)^2 = \frac{1}{2} I \omega^2 + \frac{1}{2} k \cdot \frac{m^2 r^2 \omega^4}{k^2} \\ &\quad \left[ \because (r - r_0) = \frac{m r \omega^2}{k} \text{ from Eq. (2)} \right] \\ &= \frac{1}{2} I \omega^2 + \frac{1}{2} \frac{(I \omega^2)^2}{k r^2} \quad [\because I = m r^2] \\ &= \frac{L^2}{2I} + \frac{1}{2k r^2} (L^2/I)^2 \quad [\because L = I \omega] \\ E &= \frac{L^2}{2mr^2} + \frac{L^4}{2kr^2 m^2 r^4} = \frac{L^2}{2mr^2} + \frac{L^4}{2m^2 r^6 k}. \end{aligned}$$

The quantum restriction that the angular momentum  $I\omega$  be quantized according to  $\sqrt{J(J+1)}\hbar$  will convert this classical result to a quantum-mechanical result.

$$L = \sqrt{J(J+1)}\hbar.$$

The correct allowed energies are,

$$E_J = \frac{\hbar^2}{2mr^2} J(J+1) + \frac{\hbar^4}{2m^2 r^6 k} J^2 (J+1)^2 \quad \dots(4)$$

In Eq. (4), the first term is of major importance. The second is a minor term. It is finally necessary to relate the distorted distance  $r$  in the first term to  $r_0$  by means of Eq. (3). In this way, and approximating  $r$  by  $r_0$  in the second term, we get

$$E_J = \frac{\hbar^2}{2mr_0^2} J(J+1) - \frac{\hbar^4}{m^2 r_0^6 k} J^2 (J+1)^2 + \frac{\hbar^4}{2m^2 r_0^6 k} J^2 (J+1)^2$$

$$= \frac{\hbar^2}{2mr_0^2} J(J+1) - \frac{\hbar^4}{2m^2 r_0^6 k} J^2 (J+1)^2 \quad \dots(5)$$

We are interested in the wave number

$$\bar{v} = \frac{\Delta E}{hc}$$

of the radiation emitted or absorbed as a consequence of changes between energy levels. So we consider energies in these units.

In terms of wave number units,

$$\begin{aligned} \varepsilon_J &= \frac{\hbar}{2\pi I c} J(J+1) - \frac{\hbar^3 m}{4\pi I^3 k c} J^2 (J+1)^2 \\ &= BJ(J+1) - \frac{4B^3}{\omega^2} J^2 (J+1)^2 \end{aligned}$$

Here,  $\omega$  is the vibrational energy term  $\frac{1}{2\pi c} \sqrt{k/m}$ .

Thus, the expression for the rotational energy levels is

$$\varepsilon_J = BJ(J+1) - D J^2 (J+1)^2 \quad \dots(6)$$

Here,  $D$  is called the centrifugal distortion constant.

$$D = 4B^3 / \omega^2 \quad \dots(7)$$

$D$  is very small compared with  $B$ .

Fig. 14.9 shows the lowering of rotational levels when passing from the rigid to the non-rigid diatomic molecule. The difference between the energy levels increases with increasing value of  $J$ .

The rotational absorption spectrum is produced due to molecular transition from the state  $J$  to the state  $(J+1)$ . The selection rule is  $\Delta J = +1$ .

$$\bar{v}_{J+1} - \bar{v}_J = \Delta \bar{v}$$

$$\begin{aligned} \Delta \bar{v} &= B[(J+1)(J+2) - J(J+1)] \\ &\quad - D[(J+1)^2 (J+2)^2 - J^2 (J+1)^2] \\ \Delta \bar{v} &= 2B(J+1) - 4D(J+1)^3 \text{ m}^{-1} \quad \dots(8) \end{aligned}$$

Thus, we see that the spectrum of a non-rigid diatomic molecule is similar to that of the rigid molecule except that each line is displaced slightly to low frequency (Fig. 14.10).

We note from Eq. (8) that the lines are no longer exactly equidistant but their separation decreases slightly with increasing  $J$ . The effect, however, is small owing to the smallness of  $D$  as compared to  $B$ .

#### Determination of Force Constant from Rotational Spectrum

By making observations on a number of lines in the rotation spectrum of a molecule, and by curve fitting method, the values of  $B$  and  $D$  are found out.

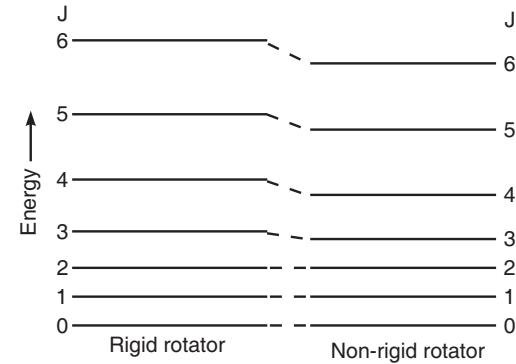


Fig. 14.9

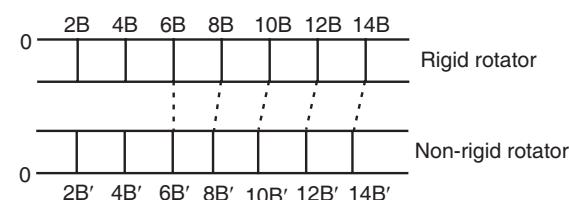


Fig. 14.10

Once  $D$  is known, the  $J$  values of lines in the observed spectrum can be determined. Further, knowing  $B$  and  $D$ , the angular frequency of vibration of the rotating molecule can be calculated using the relation

$$\omega^2 = \frac{4B^3}{D}$$

For hydrogen fluoride,  $\omega = 405000 \text{ m}^{-1}$ .

The reduced mass of the molecule  $m$  is known.

The force constant  $k$  is calculated using the equation

$$k = 4\pi^2 c^2 \omega^2 m.$$

For hydrogen fluoride, the force constant is  $960 \text{ Nm}^{-1}$  which indicates that H-F is a relatively strong bond.

**EXAMPLE 1.** *HCl molecule has a rotational constant  $B$  value of  $1059.3 \text{ m}^{-1}$  and a centrifugal distortion constant  $D$  of  $5.3 \times 10^{-2} \text{ m}^{-1}$ . Estimate the vibrational frequency and force constant of the molecule.*

**SOL.**  $\omega^2 = \frac{4B^3}{D}$ . Here,  $\omega$  is the vibrational frequency in wave number unit.

$$\omega = \sqrt{\frac{4B^3}{D}} = \sqrt{\frac{4 \times (1059.3)^3}{5.3 \times 10^{-2}}} = 299516.3 \text{ m}^{-1}.$$

Force constant  $k = 4\pi^2 c^2 \omega^2 m$ .

$$m = \frac{m_1 m_2}{m_1 + m_2} = \frac{(1.008)(35.46)}{(1.008 + 35.46)} \cdot \frac{1}{6.023 \times 10^{26}} = 1.627 \times 10^{-27} \text{ kg.}$$

$$k = 4\pi^2 \times (3 \times 10^8)^2 \times (299516.3)^2 \times 1.627 \times 10^{-27} = 518.6 \text{ Nm}^{-1}.$$

#### Infra-Red Spectroscopy (Vibration-Rotation Spectra)

- We consider first the case of a diatomic molecule and the spectrum which arises if its only motion is vibration.
- Then, we shall deal with the more practical case of a diatomic molecule undergoing vibration and rotation simultaneously.

### HARMONIC VIBRATIONS OF DIATOMIC MOLECULES

#### 14.4 THE ENERGY OF A DIATOMIC MOLECULE

Consider a diatomic molecule consisting of two atoms  $A$  and  $B$  separated by equilibrium distance  $r_e$ . Fig. 14.11 shows how the potential energy of a diatomic molecule varies with the internuclear distance  $r$ . The minimum of this curve corresponds to the normal configuration of the molecule. The zero of curve is found at  $r = r_e$ . Any energy in excess of this, for example,  $\epsilon_1$ , arises because of extension or compression of the bond. If one atom ( $A$ ) is considered to be stationary on the  $r = 0$  axis, the other will oscillate between  $B'$  and  $B''$ . If the energy is increased to  $\epsilon_2$ , the oscillation will become more vigorous.

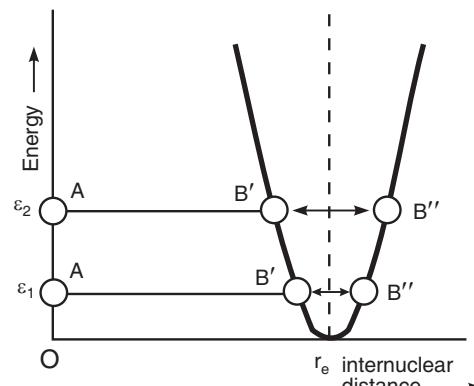


Fig. 14.11

Suppose the bond is distorted from its equilibrium length  $r_e$  to new length  $r$ . The bond, like a spring, obeys Hooke's law. Then,

$$f = -k(r - r_e) \quad \dots(1)$$

Here,  $f$  is the restoring force,  $k$  the force constant, and  $r$  the internuclear distance.

In this case, the energy curve is parabolic and has the form

$$E = \frac{1}{2}k(r - r_e)^2 \quad \dots(2)$$

This model of a vibrating diatomic molecule is called "simple harmonic oscillator" model.

#### 14.5 | VIBRATING DIATOMIC MOLECULE AS A HARMONIC OSCILLATOR

**Frequency of Vibrations.** In the diatomic molecule, we have two bodies of masses  $m_1$  and  $m_2$  joined by a spring. Suppose the bond is distorted from its equilibrium length  $r_e$  to a new length  $r$ . Then, restoring forces on each atom of a diatomic molecule are

$$m_1 \frac{d^2 r_1}{dt^2} = -k(r - r_e) \quad \dots(1)$$

$$\text{and} \quad m_2 \frac{d^2 r_2}{dt^2} = -k(r - r_e) \quad \dots(2)$$

Here,  $k$  represents the *force constant*. It is a measure of the stiffness of the bond.  $r_1$  and  $r_2$  are the positions of atoms 1 and 2 relative to the centre of mass of the molecule.

From Art. 14.2, we have,

$$r_1 = \frac{m_2 r}{m_1 + m_2}$$

$$r_2 = \frac{m_1 r}{m_1 + m_2}.$$

Putting the value of  $r_1$  in Eq. (1), we get

$$\frac{m_1 m_2}{m_1 + m_2} \frac{d^2 r}{dt^2} = -k(r - r_e) \quad \dots(3)$$

$$\text{Since } r_e \text{ is constant, } \frac{d^2 r}{dt^2} = \frac{d^2(r - r_e)}{dt^2}$$

$$\therefore \frac{m_1 m_2}{m_1 + m_2} \frac{d^2(r - r_e)}{dt^2} = -k(r - r_e) \quad \dots(4)$$

$$\text{Put} \quad r - r_e = x, \quad \text{and} \quad \frac{m_1 m_2}{m_1 + m_2} = \mu.$$

Here,  $x$  represents the displacement of the bond length from its equilibrium position.

Eq. (4) becomes

$$\mu \frac{d^2 x}{dt^2} = -kx$$

$$\text{or} \quad \frac{d^2 x}{dt^2} + \frac{k}{\mu} x = 0$$

$$\text{or } \frac{d^2x}{dt^2} + \omega^2 x = 0$$

where  $\omega^2 = k/\mu$ .

This is the equation of simple harmonic motion with frequency of vibration,

$$\nu = \frac{1}{2\pi} \sqrt{\left(\frac{k}{\mu}\right)}.$$

In  $\text{m}^{-1}$  unit (wave number),

$$\bar{\nu} = \frac{1}{2\pi c} \sqrt{\left(\frac{k}{\mu}\right)}.$$

**Energy levels.** Vibrational energies, like other molecular energies, are quantised. The permitted vibrational energies for any particular system can be calculated from Schrodinger equation. The eigen values for the energy (in joules) of a linear harmonic oscillator are given by

$$E_v = \left(v + \frac{1}{2}\right) h\nu.$$

Here  $v$  is the *vibrational quantum number*, equal to zero or an integer.

Converting to the spectroscopic units,  $\text{m}^{-1}$ , we have,

$$\varepsilon_v = \frac{E_v}{hc} = \left(v + \frac{1}{2}\right) \bar{\nu}.$$

Here,  $\varepsilon_v$  is called *term value*. Its unit is  $\text{m}^{-1}$ .

$\bar{\nu}$  is the vibrational frequency of the oscillator in wave numbers.

Zero-point energy in  $\text{m}^{-1} = \frac{1}{2} \bar{\nu}$ .

The *selection rule* for the harmonic oscillator undergoing vibrational changes is

$$\Delta v = \pm 1.$$

Vibrational energy changes will only give rise to an observable spectrum if the vibration can interact with radiation, *i.e.*, if the vibration involves a change in the dipole moment of the molecule. Thus, vibrational spectra will be observable only in heteronuclear diatomic molecules (like HF, HCl, HBr) since homonuclear molecules (like  $\text{H}_2$ ,  $\text{N}_2$  and  $\text{O}_2$ ) have no dipole moment. Applying the selection rule, we have

$$\varepsilon_{v+1 \rightarrow v} = (v+1+1/2)\bar{\nu} - (v+1/2)\bar{\nu} = \bar{\nu}$$

for emission and

$$\varepsilon_{v \rightarrow v+1} = \bar{\nu}$$

for absorption, whatever the initial value of  $v$ .

It means that all the vibrational lines obtained from harmonic oscillator are of the same frequency. The allowed vibrational energy levels and transitions between them for a diatomic molecule undergoing simple harmonic motion are shown in Fig. 14.12.

Pure vibrational spectra are observed only in liquids. This is because interactions between neighbouring molecules prevent their rotational motion.

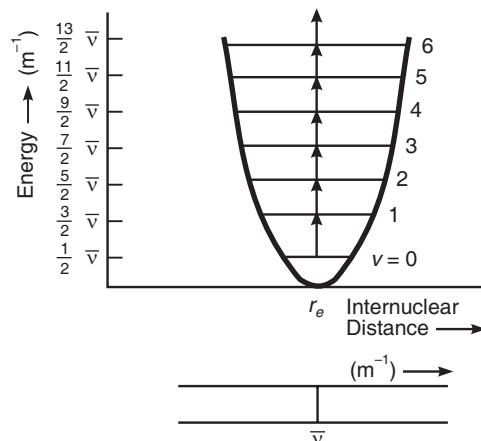


Fig. 14.12

**EXAMPLE 1.** In the near infra-red spectrum of HCl molecule there is single intense band at  $2885.9 \text{ cm}^{-1}$ . Assuming that it is due to the transition between vibrational levels, show that the force constant  $k$  is  $480 \text{ Nm}^{-1}$ . (Given :  $M_H = 1.68 \times 10^{-27} \text{ kg}$ ). (Meerut 77, 86)

**SOL.**  $\bar{v} = 2885.9 \text{ cm}^{-1} = 288590 \text{ m}^{-1}$ .

Now,  $\bar{v} = 1/\lambda = v/c$ .

Frequency of vibration is

$$v = \bar{v}c = 288590 \times (3 \times 10^8) = 8.6577 \times 10^{13} \text{ Hz.}$$

Now,  $v = \frac{1}{2\pi} \sqrt{\left(\frac{k}{\mu}\right)}$ , where  $k$  is the force constant and  $\mu$  is the reduced mass of the molecule.

Therefore,

$$k = 4\pi^2 \mu v^2.$$

The reduced mass of  $H^1 Cl^{35}$  is given by

$$\mu = \frac{M_H M_{Cl}}{M_H + M_{Cl}} = \frac{1 \times 35}{1+35} M_H = \frac{35}{36} \times (1.68 \times 10^{-27} \text{ kg}) = 1.63 \times 10^{-27} \text{ kg}$$

$$\therefore k = 4 \times (3.14)^2 \times (1.63 \times 10^{-27}) (8.6577 \times 10^{13})^2 = 482.3 \text{ Nm}^{-1}.$$

**EXAMPLE 2.** The force constant of the bond in CO molecule is  $187 \text{ Nm}^{-1}$ . Calculate the frequency of vibration of the molecule and the spacing between its vibrational energy levels in eV. Given that reduced mass of CO =  $1.14 \times 10^{-26} \text{ kg}$ ,  $h = 6.6 \times 10^{-34} \text{ Js}$  and  $1 \text{ eV} = 1.60 \times 10^{-19} \text{ J}$ .

**SOL.** The frequency of vibration of the molecule is

$$v_{osc} = \frac{1}{2\pi} \sqrt{\left(\frac{k}{\mu}\right)} = \frac{1}{2 \times 3.14} \sqrt{\left(\frac{187 \text{ Nm}^{-1}}{1.14 \times 10^{-26} \text{ kg}}\right)} = 2.04 \times 10^{13} \text{ Hz}$$

The vibrational energy of a diatomic molecule is

$$E_v = hv(v + 1/2), v = 0, 1, 2 \dots$$

The separation between two successive vibrational energy levels is

$$\begin{aligned} \Delta E &= E_{v+1} - E_v = hv\left(v + \frac{3}{2}\right) - hv\left(v + \frac{1}{2}\right) = hv \\ &= (6.63 \times 10^{-34}) \times (2.04 \times 10^{13}) \text{ J} = 8.44 \times 10^{-2} \text{ eV.} \end{aligned}$$

**EXAMPLE 3.** Given that the spacing between the vibrational levels of CO molecule is  $8.45 \times 10^{-2} \text{ eV}$  of energy. Find the force constant of the molecule.

**SOL.**  $\Delta E = 8.45 \times 10^{-2} \text{ eV} = 8.45 \times 10^{-2} \times 1.602 \times 10^{-19} \text{ J}$

$$\therefore v = \frac{\Delta E}{h} = \frac{8.45 \times 10^{-2} \times 1.602 \times 10^{-19}}{6.63 \times 10^{-34}} = 2.04 \times 10^{13} \text{ hertz.}$$

The reduced mass of CO molecule =  $\mu = 1.14 \times 10^{-26} \text{ kg}$ .

$$\begin{aligned} k &= 4\pi^2 v^2 \mu = 4 \times (3.14)^2 \times (2.04 \times 10^{13})^2 \times 1.14 \times 10^{-26} \\ &= 187 \text{ Nm}^{-1}. \end{aligned}$$

**EXAMPLE 4.** Calculate the vibrational energy levels of an HCl molecule, assuming the force constant to be  $516 \text{ Nm}^{-1}$ .

**SOL.**

$$\mu = (35 / 36) \times 1.0078 \times 1.66 \times 10^{-27} \text{ kg.}$$

$$v_0 = \frac{1}{2\pi} \left( \frac{516}{\mu} \right)^{1/2}$$

The energy levels are then given by

$$\begin{aligned} hv_0(v + 1/2) &= 6.6 \times 10^{-34} \times \frac{1}{2\pi} \left( \frac{516 \times 36}{35 \times 1.0078 \times 1.66 \times 10^{-27}} \right)^{1/2} (v + 1/2) \text{ J} \\ &\sim 5.9 \times 10^{-20} (v + 1/2) \text{ J} \\ &\sim 0.37 (v + 1/2) \text{ eV } (v = 0, 1, 2, \dots) \end{aligned}$$

## 14.6 THE ANHARMONIC OSCILLATOR

Real molecules do not obey exactly the laws of simple harmonic motion. Real molecules dissociate at large values of  $r$  and acquire very large amounts of energy as  $r$  decreases. Fig. 14.13 shows the energy curve for a typical diatomic molecule undergoing anharmonic extensions and compressions together with (dotted) the ideal simple harmonic parabola.

A purely empirical expression which fits this curve to a good approximation was derived by P. M. Morse, and is called the Morse function :

$$E = D_e [1 - \exp \{a(r_e - r)\}]^2 \quad \dots(1)$$

where  $D_e$  is the dissociation energy,  $a$  is constant for a particular molecule,  $r_e$  is the value of internuclear distance  $r$  that corresponds to the minimum of Morse curve.

**Energy Levels.** When Eq. (1) is used in the Schrodinger equation, the allowed vibrational energy levels are found to be

$$\epsilon_v = (v + 1/2) \bar{v}_e - (v + 1/2)^2 \bar{v}_e x_e \text{ m}^{-1} \quad (v = 0, 1, 2, \dots) \quad \dots(2)$$

Here,  $\bar{v}_e$  is called the *equilibrium oscillation frequency* of the molecule expressed in wave numbers and is the value for small displacements.  $x_e$  is the corresponding anharmonicity constant.  $x_e \approx 0.01$ .

The selection rules for the anharmonic oscillator are found to be  $\Delta v = \pm 1, \pm 2, \pm 3, \dots$

Normally only the lines of  $\Delta v = \pm 1, \pm 2$  and  $\pm 3$ , have observable intensity.

Vibration bands are almost invariably observed in absorption. At room temperature nearly all the molecules in a particular sample have only the zero point vibrational energy and exist in  $v = 0$  state. The absorption of radiation must, therefore, result in transitions starting from  $v = 0$ .

1. The *fundamental* absorption band corresponds to

$v = 0 \rightarrow v = 1, \Delta v = +1$ , with considerable intensity.

$$\Delta \epsilon = \epsilon_{v=1} - \epsilon_{v=0}$$

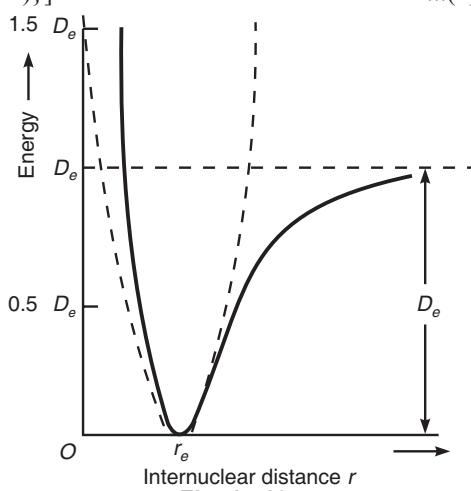


Fig. 14.13

$$\begin{aligned}
 &= (1+1/2)\bar{v}_e - x_e (1+1/2)^2 \bar{v}_e - \{1/2\bar{v}_e - (1/2)^2 x_e \bar{v}_e\} \\
 &= \bar{v}_e (1-2x_e) m^{-1}
 \end{aligned} \quad \dots(4)$$

2. Band corresponding to  $v=0 \rightarrow v=2$ ,  $\Delta v=+2$ , with small intensity, is called *first overtone*.

$$\begin{aligned}
 \Delta \varepsilon &= (2+1/2)\bar{v}_e - x_e (2+1/2)^2 \bar{v}_e - \{1/2\bar{v}_e - (1/2)^2 x_e \bar{v}_e\} \\
 &= 2\bar{v}_e (1-3x_e) m^{-1}
 \end{aligned} \quad \dots(5)$$

3. Band corresponding to  $v=0 \rightarrow v=3$ ,  $\Delta v=+3$ , with normally negligible intensity is called *second overtone*.

$$\begin{aligned}
 \Delta \varepsilon &= (3+1/2)\bar{v}_e - (3+1/2)^2 \bar{v}_e x_e - \{1/2\bar{v}_e - (1/2)^2 x_e \bar{v}_e\} \\
 &= 3\bar{v}_e (1-4x_e) m^{-1}
 \end{aligned} \quad \dots(6)$$

The allowed vibrational energy levels and these three absorption transitions are shown in Fig. 14.14.

4. If the temperature is raised, the population of the  $v=1$  state may become appreciable. The transitions from energy levels with the selection rule  $\Delta v=1$  are called *hot bands*. The first hot band corresponds to

$v=1 \rightarrow v=2$ ,  $\Delta v=+1$ , normally very weak.

$$\begin{aligned}
 \Delta \varepsilon &= 2 \frac{1}{2}\bar{v}_e - 6 \frac{1}{4}x_e \bar{v}_e - \left\{ 1 \frac{1}{2}\bar{v}_e - 2 \frac{1}{4}x_e \bar{v}_e \right\} \\
 &= \bar{v}_e (1-4x_e) m^{-1}
 \end{aligned}$$

Their nature may be confirmed by raising the temperature of the sample when a true hot band will increase in intensity.

**EXAMPLE 1.** The fundamental band for CO is centred at  $2143.3 \text{ cm}^{-1}$  and first overtone at  $4259.7 \text{ cm}^{-1}$ . Calculate  $\bar{v}_e$  and  $\bar{v}_e x_e$ .

**SOL.** We know that frequencies of fundamental and first overtone bands are given by

$$\bar{v}_1 = (1-2x_e)\bar{v}_e$$

$$\bar{v}_2 = (1-3x_e)2\bar{v}_e$$

and

$$(1-2x_e)\bar{v}_e = 214330 \text{ m}^{-1} \quad \dots(1)$$

$$(1-3x_e)2\bar{v}_e = 425970 \text{ m}^{-1} \quad \dots(2)$$

Solving (1) and (2),

$$x_e = 0.006, \bar{v}_e = 224992 \text{ m}^{-1}$$

$$\bar{v}_e x_e = 1350 \text{ m}^{-1}$$

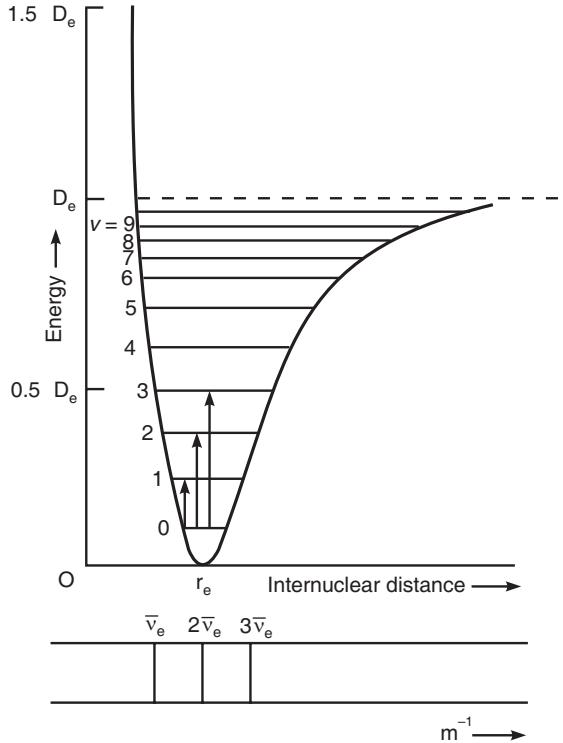


Fig. 14.14

**EXAMPLE 2.** The spectrum of HCl shows a very intense absorption at  $2886\text{ cm}^{-1}$ , a weaker one at  $5668\text{ cm}^{-1}$ , and very weak one at  $8347\text{ cm}^{-1}$ . Find the equilibrium frequency of the molecule and the anharmonicity constant.

**SOL.** We solve any two of the following three equations :

$$\begin{aligned}\bar{v}_e &= (1 - 2x_e) = 2886 \\ 2\bar{v}_e (1 - 3x_e) &= 5668 \\ 3\bar{v}_e (1 - 4x_e) &= 8347 \text{ cm}^{-1} \\ \bar{v}_e &= 2990 \text{ cm}^{-1}, \quad x_e = 0.0174.\end{aligned}$$

The force constant of the bond in HCl may be calculated by using the formula,

$$k = 4\pi^2 \bar{v}_e^2 c^2 \mu \text{ Nm}^{-1}.$$

#### 14.6.1. Fundamental and Overtone Frequencies

The frequency corresponding to  $v = 0$  to  $v = 1$  transition is called the *fundamental frequency*. We sometimes observe frequencies due to  $v = 0$  to  $v = 2$  or even  $v = 0$  to  $v = 3$  transitions. Such frequencies are called *overtone frequencies*. Since the frequencies of first and second overtone bands are approximately 2 to 3 times the frequency of the fundamental band, they appear in regions of the shorter wavelengths than do the fundamental.

#### 14.6.2. Calculation of Force Constant

The vibrational frequency of fundamental =  $\bar{v}_e (1 - 2x_e) \text{ m}^{-1}$

The vibrational frequency of the first overtone =  $2\bar{v}_e (1 - 3x_e) \text{ m}^{-1}$

The vibrational frequency of the second overtone =  $3\bar{v}_e (1 - 4x_e) \text{ m}^{-1}$ .

By studying the vibration spectra of a diatomic molecule, the above frequencies could be identified. Knowing these frequencies, we can solve any two of the above three equations and calculate the equilibrium vibrational frequency  $\bar{v}_e$ . The reduced mass of the molecule ( $\mu$ ) is known.

The force constant ( $k$ ) of the bond in the diatomic molecule may be calculated from the equation,

$$k = 4\pi^2 c^2 \mu \bar{v}_e^2.$$

**EXAMPLE 1.** The fundamental and the first overtone frequencies of the diatomic molecule HCl behaving as anharmonic oscillator are found to be  $2886\text{ cm}^{-1}$  and  $5668\text{ cm}^{-1}$  respectively. If the reduced mass of HCl be  $1.63 \times 10^{-27}\text{ kg}$ , calculate the force constant.

**SOL.**  $\bar{v}_e (1 - 2x_e) = 2886 \quad \dots(1)$

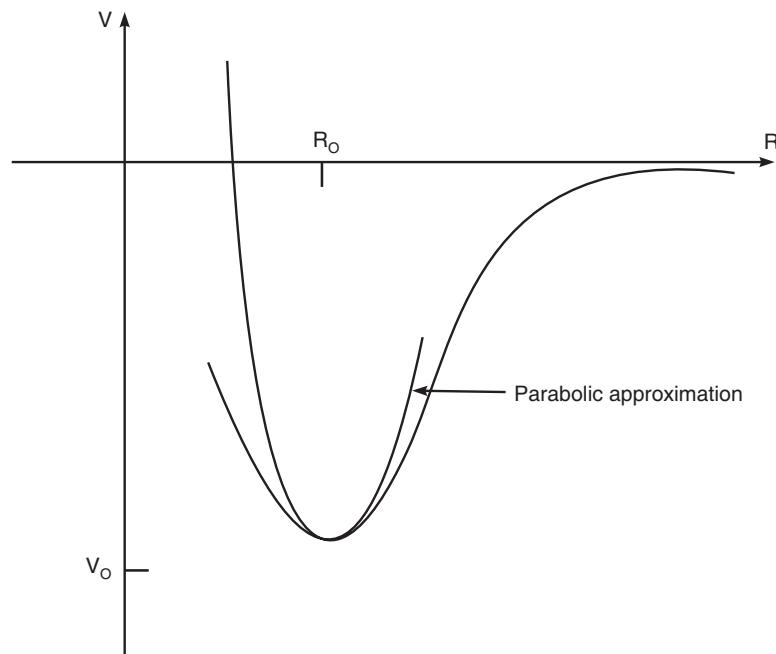
$$2\bar{v}_e (1 - 3x_e) = 5668 \quad \dots(2)$$

Solving Eqs. (1) and (2),

$$\begin{aligned}\bar{v}_e &= 2990 \text{ cm}^{-1} = 2990 \times 10^2 \text{ m}^{-1} \\ k &= 4\pi^2 c^2 \mu \bar{v}_e^2 \\ &= 4 \times (3.14)^2 \times (3 \times 10^8)^2 \times 1.63 \times 10^{-27} \times (2990 \times 10^2)^2 \\ &= 516 \text{ Nm}^{-1}.\end{aligned}$$

### 14.7 DIATOMIC VIBRATING ROTATOR—THEORY OF THE ORIGIN OF THE VIBRATION-ROTATION SPECTRUM OF A MOLECULE

When sufficiently excited, a molecule can vibrate as well as rotate. We shall consider only diatomic molecules. Fig. 14.15 shows how the potential energy of a diatomic molecule varies with



**Fig. 14.15**

the internuclear distance  $R$ . In the neighbourhood of the minimum of this curve, which corresponds to the normal configuration of the molecule, the shape of the curve is very nearly a parabola. In this region, then

$$V = V_0 + \frac{1}{2}k(R - R_0)^2 \quad \dots(1)$$

where  $R_0$  is the equilibrium separation of the atoms.

The interatomic force that gives rise to this potential energy

$$F = -dV/dR = -k(R - R_0) \quad \dots(2)$$

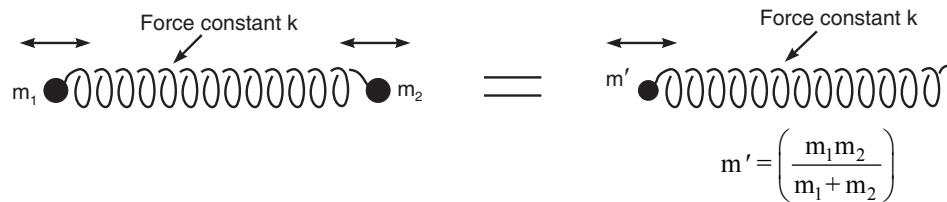
The force is just the restoring force that a stretched or compressed spring exerts and, as with a spring, a molecule suitably excited can undergo simple harmonic oscillations. Classically, the frequency of a vibrating body of mass  $m$  connected to a spring of force constant  $k$  is

$$\nu_0 = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad \dots(3)$$

In the diatomic molecule we have two bodies of masses  $m_1$  and  $m_2$  joined by a spring. The use of the “reduced mass” reduces the problem of two particles  $m_1$  and  $m_2$  vibrating about their common centre of mass to the problem of a single mass  $m'$  vibrating about a fixed point (Fig. 14.16).

The frequency of oscillation of such a two-body oscillator is

$$\nu_0 = \frac{1}{2\pi} \sqrt{\frac{k}{m'}} \quad \dots(4)$$

**Fig. 14.16**

When the harmonic oscillator problem is solved quantum mechanically, the energy of the oscillator is found to be restricted to the values

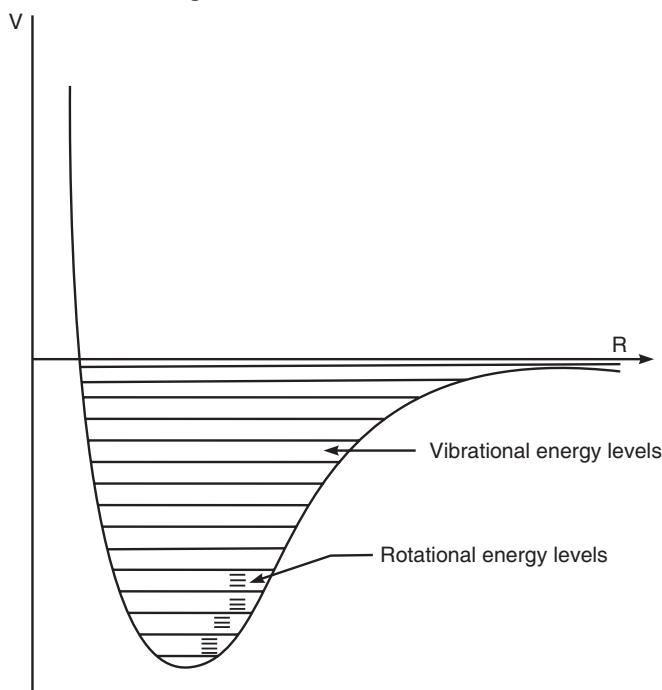
$$E_v = (v + \frac{1}{2}) \hbar v_0 \quad \dots(5)$$

Here  $v$  is the vibrational quantum number. It may have the values  $v = 0, 1, 2, 3, \dots$

Substituting the value of  $v_0$  from Eq. (4),

$$E_v = (v + \frac{1}{2}) \hbar \sqrt{\frac{k}{m'}} \quad \dots(6)$$

These levels are sketched on Fig. 14.17.

**Fig. 14.17**

For *large-amplitude* vibrations, the parabola does not fit the actual energy curve. Eq. (6) no longer holds. The energy well is wider than the parabola for the higher  $E_v$ 's, and therefore these energy levels are closer together than the lower  $E_v$ 's. This diagram also shows the fine structure in the vibrational levels caused by the simultaneous excitation of rotational levels.

The selection rule for transitions between vibrational states is

$$\Delta v = \pm 1 \quad \dots(7)$$

Pure vibrational spectra are observed only in liquids. This is because interactions between neighbouring molecules prevent their rotational motion.

The vibrations and rotations of a molecule take place independently of each other. Under these circumstances the energy levels of a diatomic molecule are specified by

$$E_{v,J} = (v + \frac{1}{2}) \hbar \sqrt{\frac{k}{m'}} + J(J+1) \frac{\hbar^2}{2I} \quad \dots(8)$$

$E_v$  is very much larger than  $E_r$ . Even at high temperatures only the vibrational states corresponding to  $v = 0$  and  $v = 1$  are excited. Fig. 14.18 shows the  $J = 0, 1, 2, 3$  and  $4$  levels of a diatomic molecule for the  $v = 0$  and  $v = 1$  vibrational states. It also shows the spectral lines in absorption. The selection rules are  $\Delta v = 1$  and  $\Delta J = \pm 1$ . The transition with  $\Delta J = 0$  is not allowed.

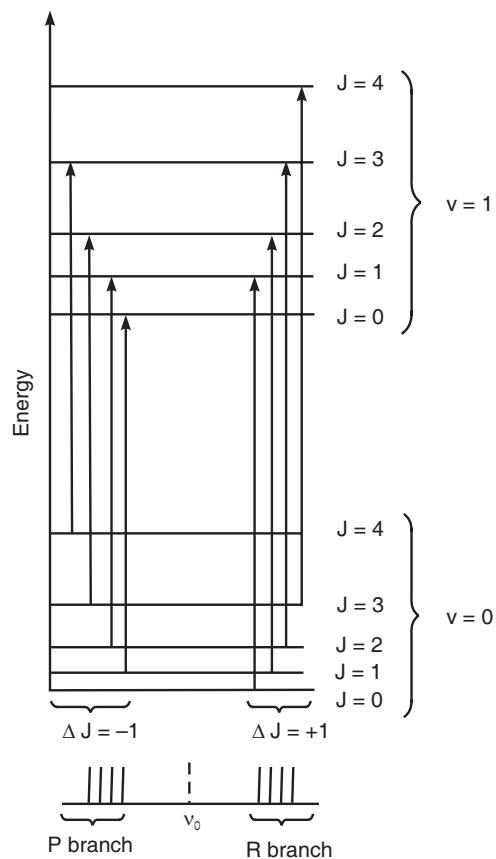


Fig. 14.18

The  $v = 0 \rightarrow v = 1$  transitions fall into two categories :

- (i) *P Branch:* Here,  $\Delta J = -1$  (i.e.,  $J \rightarrow J - 1$ ).

From Eq. (8) the frequencies of the spectral lines in the *P* branch are given by

$$\nu_p = \frac{E_{1,J-1} - E_{0,J}}{\hbar} = \frac{1}{2\pi} \sqrt{\frac{k}{m'}} + [(J-1)J - J(J+1)] \frac{\hbar}{4\pi I}$$

$$\text{or} \quad \nu_p = \nu_0 - J \frac{\hbar}{2\pi I} \quad J = 1, 2, 3, \dots \quad \dots(9)$$

(ii) *R Branch*: Here,  $\Delta J = +1$  (*i.e.*,  $J \rightarrow J+1$ )

$$\nu_R = \frac{E_{1,J+1} - E_{0,J}}{\hbar} = \frac{1}{2\pi} \sqrt{\frac{k}{m'}} + [(J+1)(J+2) - J(J+1)] \frac{\hbar}{4\pi I}$$

or  $\nu_R = \nu_0 + (J+1) \frac{\hbar}{2\pi I} \quad J = 0, 1, 2, 3, \dots \quad \dots(10)$

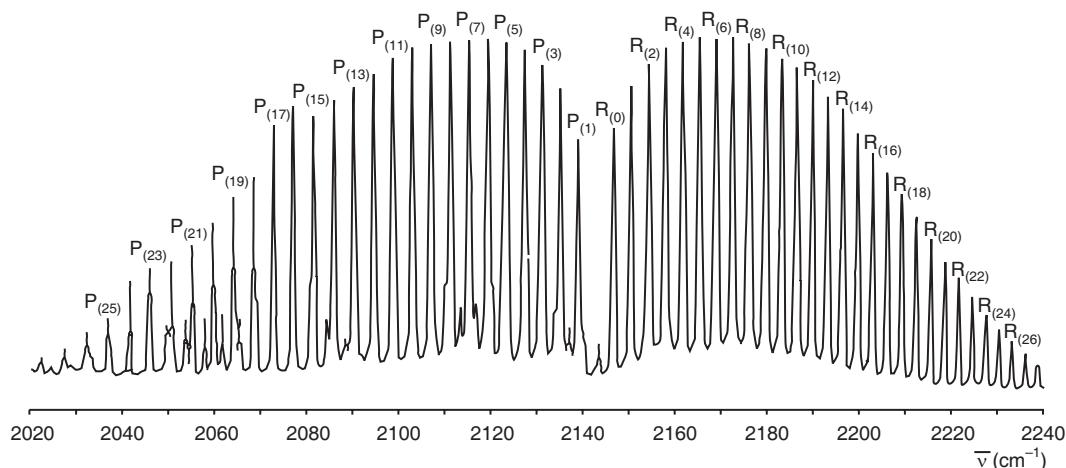
There is no line at  $\nu = \nu_0$  (the *Q branch*) because transitions for which  $\Delta J = 0$  are forbidden in diatomic molecules.

The spacing between the lines in both the *P* and the *R* branch is  $\Delta\nu = \frac{\hbar}{2\pi I}$ .

Hence, by measuring the frequencies of these lines, the moment of inertia of the molecule and the length of the chemical bond can be calculated.

### 14.8 VIBRATION-ROTATION SPECTRUM OF CARBON MONOXIDE

Fig. 14.9 shows the fundamental vibration-rotation band of carbon monoxide under high resolution.



**Fig. 14.19**

It is made by sending radiation from a source of continuous infrared wavelengths through a cell containing CO and recording the absorption spectra.

Table 14.1 gives the observed wavenumbers of the first five lines in each branch.

- There is a slight decrease in separation between the rotational lines as the wavenumber increases.
- The band centre is at about  $2143 \text{ cm}^{-1}$ . The average line separation near the centre is  $3.83 \text{ cm}^{-1}$ . This gives:  $2B = 3.83 \text{ cm}^{-1}$   $B = 1.915 \text{ cm}^{-1}$ .

This is in satisfactory agreement with the value  $B = 1.92118 \text{ cm}^{-1}$  derived by microwave studies. Hence we can calculate the moment of inertia and bond length from infra-red data alone. [See Example 1 in section 14.2]

**Table 14.1. Part of the infra-red spectrum of CO molecule**

Line	$\bar{v}$	Separation $\Delta \bar{v}$	Line	$\bar{v}$	Separation $\Delta \bar{v}$
P(1)	2139.43		R(0)	2147.08	
		3.88			3.78
P(2)	2135.55		R(1)	2150.86	
		3.92			3.73
P(3)	2131.63		R(2)	2154.59	
		3.95			3.72
P(4)	2127.68		R(3)	2158.31	
		3.98			3.66
P(5)	2123.70		R(4)	2161.97	

- From Table 14.1, we see that the band origin, at the midpoint of P(1) and R(0) is at  $2143.26\text{ cm}^{-1}$ . This, then, is the fundamental vibration frequency of CO molecule, if anharmonicity is ignored.
- The fundamental band for CO is centred at  $2143.26\text{ cm}^{-1}$ . The first overtone is centred at  $4260.04\text{ cm}^{-1}$ . We know that frequencies of fundamental and first overtone bands are given by

$$\bar{v}_1 = (1 - 2x_e)\bar{v}_e$$

and

$$\bar{v}_2 = (1 - 3x_e)2\bar{v}_e$$

$$(1 - 2x_e)\bar{v}_e = 2143.26\text{ cm}^{-1} \quad \dots (1)$$

$$(1 - 3x_e)2\bar{v}_e = 4260.04\text{ cm}^{-1} \quad \dots (2)$$

Solving (1) and (2),  $x_e = 0.0061$ ,  $\bar{v}_e = 2169.74\text{ cm}^{-1}$

**Fig. 14.20**

- Fig. 14.20 shows the spectrum of carbon monoxide under very low resolution. All rotational fine structure has been lost and a typical PR contour is seen.

## 14.9 INFRA-RED SPECTRA

Suppose a vibrational transition takes place within a given electronic level. Then the wavelength of the emitted radiation would be of the order of 12,000 nm which would be in the near infra-red region of the spectrum. The infra-red (I. R.) radiation absorbed by a pure liquid or by a material in solution can be studied and *well-defined* absorption lines can be recorded. This is because, in the liquid phase, no well-defined rotational energy levels exist (interaction between adjacent molecules inhibit rotation). So rotational structure is not found on a vibrational absorption band spectrum.

However, the I.R. spectrum of a simple (diatomic) molecule in the gas phase, when examined with high resolution instrument, shows rotational component lines in addition to vibrational lines. When vibrational energy changes, it is accompanied by rotational energy change. When this

happens, each absorption band (which is due to a particular vibrational energy change) is found to consist of a number of relatively closely spaced absorption lines. They appear as a broad streak, called the vibration-rotation band. The rotational components of the band can be related to the rotational energy changes that occur simultaneously with the vibrational energy change. A study of such vibration-rotation bands (I.R. spectrum) enables us to find force constant, moment of inertia and equilibrium bond length data of the molecule concerned.

A pure rotational transition without any change in either the vibrational or the electronic energy is possible. The resulting spectral lines will be in the far infra-red. These lines constitute *rotational* spectra.

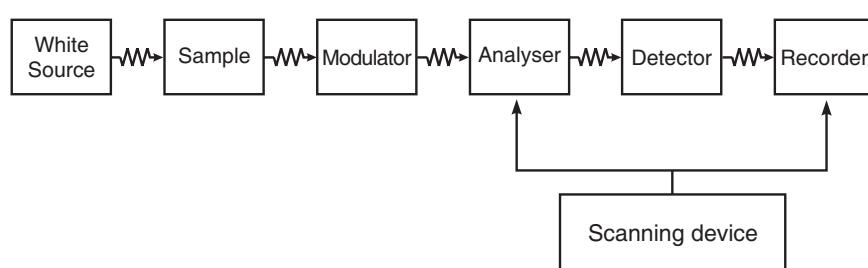
### 14.10 TECHNIQUES AND INSTRUMENTATION

The infra-red region of the spectrum extends beyond the red end of the visible spectrum (720 nm) up to 400 microns (1 micron =  $1\mu = 10^{-4}$  cm).

#### Absorption Spectroscopy

Light from a source, which emits radiations of all wavelengths, is passed through the sample under study. The transmitted light is spectrally studied to find the absorption at different wavelengths.

Fig. 14.21 shows the block diagram of an absorption spectrometer for use in the infra-red, visible, and ultra-violet regions.



**Fig. 14.21**

The radiation from a *white* source is directed by some guiding device (lens) on to the sample. From the sample, the radiation passes through an analyser (prism). The analyser selects the frequency reaching the detector at any given time. The signal from the detector passes to a recorder which is synchronized with the analyser so as to produce a trace of the absorbance as the frequency varies.

#### 14.10.1. Principle of a Double-Beam Spectrophotometer

In most absorption studies, the *double-beam principle* is used.

Fig. 14.22 shows the schematic diagram of a double-beam spectrophotometer.

The incident beam is split at the plate  $P_1$  into two beams  $A$  and  $B$ . The two beams are energetically and optically identical. The beam  $A$  passes through the absorption sample. The beam  $B$  passes through a variable intensity control  $R$ . The beams  $A'$  and  $B'$  are then to be balanced at each wavelength  $\lambda$ .

The percentage cut-off of intensity needed at  $R$  to secure balance gives the percentage absorption by the sample in path  $A$ .

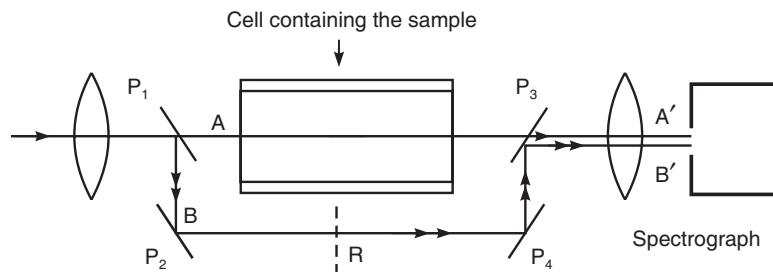
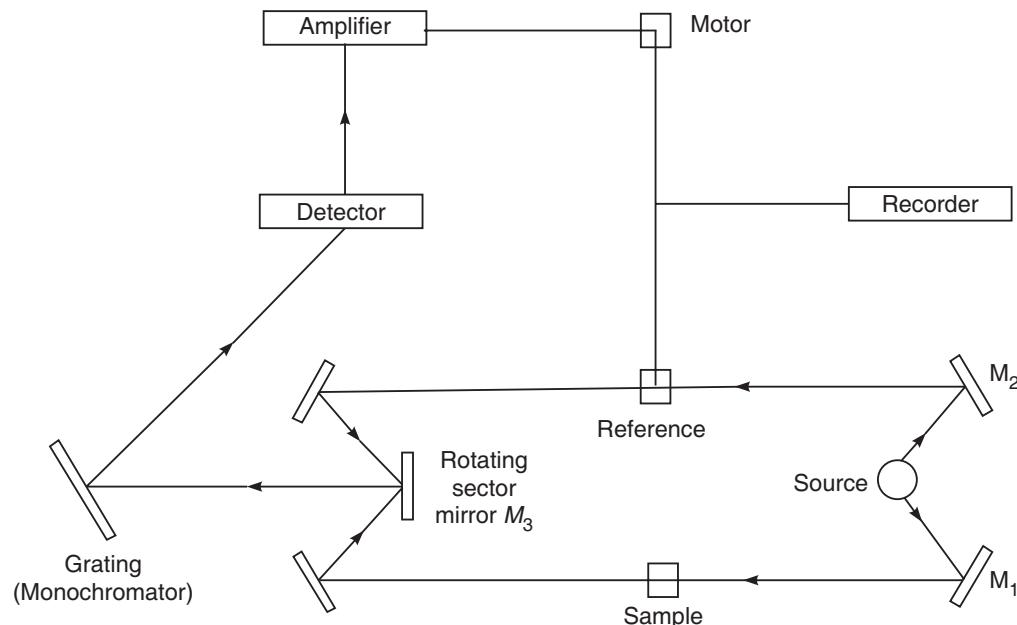


Fig. 14.22

**Advantage.** The double beam system makes the working independent of time variations in source intensity.

### 14.11 DOUBLE-BEAM INFRARED SPECTROMETER

Fig. 14.23 shows the block diagram of a double-beam infrared spectrometer.



(a) Schematic diagram of Double-beam infrared spectrometer

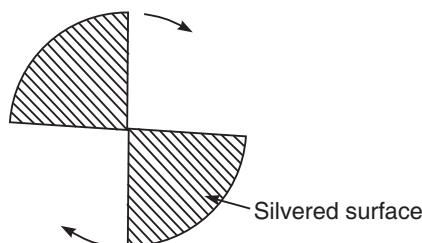
(b) a plan view of the rotating sector mirror, M<sub>3</sub>

Fig. 14.23

The essential components of the spectrometer are:

- (i) source of radiation      (ii) fore optics      (iii) monochromator
- (iv) detector with an amplifier and      (v) recorder.

**Principle.** The source radiation is divided into two by means of the mirrors  $M_1$  and  $M_2$ . One beam is brought to a focus at the sample space, while the other follows an exactly equivalent path and is referred to as the reference beam. The two beams meet at the sector mirror  $M_3$ . As this mirror rotates, it alternately reflects the reference beam, or allows the sample beam through the spaces, into the monochromator. Thus the detector 'sees' the sample beam and reference beam alternately.

If a sample, capable of absorbing energy from the beam at the particular frequency passed by the monochromator, is now placed in the sample beam, the detector will receive a signal alternating in intensity, since the sample beam carries less energy than the reference beam. This alternating signal is amplified electronically. A calibrated attenuator is driven into the reference beam until the signal is reduced to zero, i.e., until sample and reference beams are again balanced. The distance moved by the attenuator is a direct measure of amount of the energy absorbed by the sample.

**Working (i) Source of radiation.** The source is always some form of filament which is maintained at red-or white-heat by an electric current. Two common sources are the Nernst filament and the 'globar' filament. The Nernst glower consists of a spindle of rare earth oxides about 2.5 cm long and 0.25cm in diameter. It has very good emissivity in the shorter wavelengths. It requires pre-heating to make it conductive.

The globar is a low resistance rod of silicon carbide. It is useful in the longer wavelengths.

**(ii) The fore optics.** The beam is guided and focused by mirrors aluminized or silvered on their surfaces. Normally a focus is produced at the point where the sample is to be placed. Ordinary lenses and mirrors are not suitable as glass absorbs strongly over most of the frequencies used. Any windows which are essential (e.g., to contain a sample, or to protect the detector) must be made of mineral salts transparent to infra-red radiation which have been highly polished in order to reduce scattering to a minimum. NaCl, which is transparent above  $650\text{ cm}^{-1}$ , and KBr, transparent above  $400\text{ cm}^{-1}$ , are much used. For aqueous samples, AgCl and CaF<sub>2</sub> sample cells can be used, and are infra-red transparent above 400 and  $1200\text{ cm}^{-1}$ , respectively.

$M_1$  and  $M_2$  divide the source radiation into two equivalent beams. One beam passes through the sample. The other beam passes through an equivalent path and is called the reference beam. The two beams meet at the rotating sector mirror  $M_3$ . The rotating mirror alternately allows the sample beam through the spaces or reflects the reference beam with a predetermined period to the monochromator slit. Consequently, the detector receives the sample beam and reference beam alternately. Since both the beams have travelled the same path length, compensation for extraneous effects due to the atmosphere etc., have been taken care of.

**(iii) Monochromator.** It splits the polychromatic radiation into the component wavelengths by using prisms or gratings or both. Gratings give better resolving power. The prism or grating is rotatable mechanically. The resolution also depends on slit width and quality of mirrors.

**(iv) Detector and Amplifier.** Two main types of detectors are in common use, one sensing the heating effect of the radiation, the other depending on photoconductivity. In both, the greater the effect (temperature or conductivity rise) at a given frequency, the greater the transmittance (and the less the absorbance) of the sample at that frequency.

**Golay cell:** An example of the temperature method is the Golay cell which is pneumatic in operation. The radiation falls on to a very small cell containing air, and temperature changes

are measured in terms of pressure changes within the cell which can be recorded directly as 'transmittance'.

**Thermocouples:** Alternative examples of thermal detectors are small, sensitive thermocouples or bolometers.

**Photoconducting detectors.** Photoconductivity is the increase in electrical conductivity of a semiconductor caused by radiation incident on it. The radiation is allowed to fall on photoconducting material and the conductivity of the material is measured continuously by a bridge network. The detector does not show any signal when the energy transmitted by reference and sample cells are equal. If the sample absorbs radiation, there will be inequality in the beams and a signal is produced. The signal is then fed to a high gain amplifier which amplifies the low intensity signal upto a useable level with minimum thermal and electrical noise.

Photoconductive detectors in common use are indium antimonide (InSb) which can be used above  $1400\text{cm}^{-1}$  and mercury cadmium telluride (MCT) used above  $\sim 700\text{cm}^{-1}$ . These detectors operate at liquid nitrogen temperatures, which reduces noise, and have a faster response time and sensitivity. Where extremely high sensitivity is required, doped photoconductors such as copper-or gold-doped germanium detectors can be used-but these generally need to operate at liquid helium temperatures.

(v) **Recorder.** The amplified signal is used to move an attenuator (comb) which cuts down the radiation coming out of the reference beam until energy balance is restored. This is achieved by a motor which drives the comb into reference beam when an absorbing band is encountered and out of the beam when the band is passed over. The recorder pen is also coupled to this motor so that the comb movement is followed exactly by the pen.

**Advantages.** (i) By balancing sample and reference beams, the absorption of atmospheric  $\text{CO}_2$  and  $\text{H}_2\text{O}$  does not appear in the infra-red spectra since both beams are reduced in energy to the same extent.

(ii) It is much simpler to amplify the alternating signal produced than the d.c. signal resulting from a single-beam detector.

(iii) The sector mirror acts as a modulator since it interrupts the beam periodically and by amplifying only that component of the signal having the sector mirror frequency (usually 10-100 rotations per second), a great improvement in the signal-to-noise ratio results.

(iv) When examining the spectra of solutions, one can put a cell containing the appropriate quantity of pure solvent into the reference beam, thus eliminating the solvent spectrum from the final trace. On a single-beam instrument, the solvent spectrum must be taken separately and 'subtracted' from the solution spectrum in order to arrive at the spectrum of the substance of interest.

**Disadvantages.** (i) A double-beam instrument is never completely effective in removing traces of water vapour or  $\text{CO}_2$  from the spectra. No matter how carefully the instrument is assembled, small differences occur in the beam paths and a small residual spectrum results. This can usually be removed however by sweeping with dry, inert gas as well as using the double-beam principle.

(ii) The double-beam instrument only removes the spectral trace of  $\text{CO}_2$  and  $\text{H}_2\text{O}$ . The very strong absorption of energy by these gases still remains in both beams. This means that at some parts of the spectrum, the actual amount of energy reaching the detector may be extremely small. Under these conditions, unless the spectrometer is very carefully operated, the spectral trace of a substance may be quite false. Fortunately, regions of very high atmospheric absorption are few and narrow but they should be borne in mind when examining infra-red spectra. This disadvantage can only be removed by sweeping out or evacuating the spectrometer. Similar, but more pronounced, effects occur in regions of strong solvent absorbance when a compensating cell is put in the reference beam.

## RAMAN SPECTROSCOPY

### 14.12 SCATTERING OF LIGHT

(i) **Rayleigh scattering.** When radiation is incident on a molecule, it is in general, scattered by the molecule. The dominant mode of scattering is elastic scattering. Here, the frequency of the scattered light remains the same as that of the incident radiation. This phenomenon is called Rayleigh scattering.

Rayleigh showed theoretically that the intensity of scattered light is inversely proportional to the fourth power of the wavelength *i.e.*,  $I \propto \frac{1}{\lambda^4}$ .

This is called Rayleigh's law of scattering.  $I$  is also proportional to the square of the volume of the particle. The blue colour of the sky is due to the greater scattering of short wavelengths by the dust suspensions in the atmosphere or by the air molecules. This type of scattering simply produces separation of wavelengths originally present in the incident light. No wavelength can be present in the scattered light which is not present in the incident light *i.e.*, no new wavelength is created. For this reason, Rayleigh scattering is called coherent scattering.



C.V. Raman (1888-1970)

(ii) **Raman Effect.** While studying the scattering of light, Raman found that when a beam of monochromatic light was passed through organic liquids such as benzene, toluene, etc., the scattered light contained other frequencies in addition to that of the incident light. This is called *Raman effect*.

The scattering is inelastic.

### 14.13 EXPERIMENTAL STUDY OF RAMAN EFFECT

Fig. 14.24 shows the apparatus used for studying Raman effect in liquids.

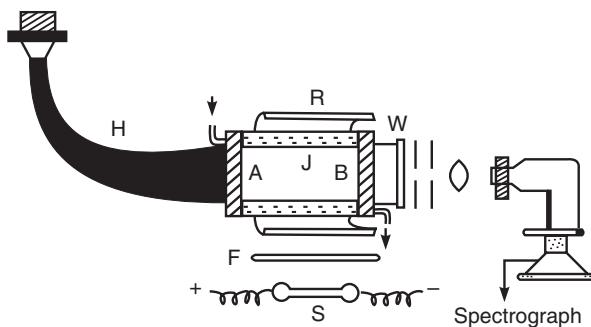


Fig. 14.24

It consists of a glass tube *AB* containing the pure experimental liquid free from dust and air bubbles. The tube is closed at one end by an optically plane glass plate *W* and at the other end it is drawn into a horn (*H*) and blackened on the outside. Light from a mercury arc *S* is passed through a filter *F* which allows only monochromatic radiation of  $\lambda = 4358 \text{ \AA}$  to pass through it. The tube is surrounded by a water-jacket (*J*) through which water is circulated to prevent overheating of the liquid. A semi-cylindrical aluminium reflector *R* is used to increase the intensity of illumination. The scattered light coming out of *W* is condensed on the slit of a spectrograph. The spectrograph must have a large light gathering power and the prism must have a large resolving power. A short focus camera is used to photograph the spectrum.

It is observed that the spectrum contains, in addition to the original wavelength ( $4358 \text{ \AA}$ ), some lines which have wavelengths longer as well as shorter than the original wavelength.

These lines of modified wavelengths are known as *Raman lines*. The lines having wavelengths greater than that of the incident wavelength are called *Stokes lines* and those having shorter wavelengths are called *anti-Stokes lines*. The Stokes lines are found to be more intense than the anti-Stokes lines. Most of the Raman lines are strongly polarised.

The displacement of the modified spectral lines from the exciting line, when measured in wavenumbers, depends only on the scattering substance and is independent of the wavenumber of the exciting radiation. The Raman spectrum is thus characteristic of the scattering substance.

The spectrum exhibits a number of Stokes' lines, a few anti-Stokes lines and a strong unmodified line (Fig. 14.25).

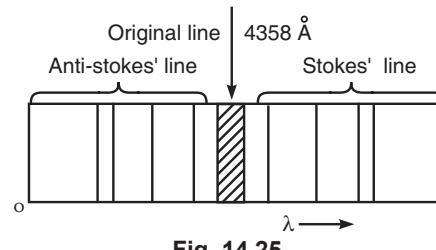


Fig. 14.25

#### Characteristics of Raman Lines

- (1) The Stokes lines are found to be more intense than the anti-Stokes lines.
- (2) The Raman lines are symmetrically displaced about the parent line.
- (3) The frequency difference between the modified and parent line represents the frequency of the corresponding infrared absorption line.

### 14.14 QUANTUM THEORY OF RAMAN EFFECT

Raman effect is due to the interaction between a light photon and a molecule of the scatterer. Suppose a photon of frequency  $v_1$  is incident on a molecule and there is a collision between the two. Let  $m$  = mass of the molecule,  $v_1$  and  $v_2$  its velocities before and after impact,  $E_1$  and  $E_2$  the intrinsic energies of the molecule before and after collision. Let  $v_2$  be the frequency of the scattered photon. Applying the principle of conservation of energy,

$$E_2 + \frac{1}{2}mv_2^2 + hv_2 = E_1 + \frac{1}{2}mv_1^2 + hv_1 \quad \dots(1)$$

We may assume that the K.E. of the molecule is unaltered during the process. Hence,

$$E_2 + hv_2 = E_1 + hv_1 \text{ or } v_2 - v_1 = \frac{E_1 - E_2}{h}$$

$$\text{or} \quad v_2 = v_1 + \frac{E_1 - E_2}{h} \quad \dots(2)$$

Three cases may arise:

- (1) When there is no change in the energy of the molecule,  $E_1 = E_2$ . Then  $v_1 = v_2$ . This represents the unmodified line.
- (2) If  $E_2 > E_1$ , then,  $v_2 < v_1$ . This represents the Stokes line. It means that the molecule has absorbed some energy from the incident photon. Consequently the scattered photon has lower energy or longer wavelength.
- (3) If  $E_2 < E_1$ , then,  $v_2 > v_1$ . This represents the anti Stokes line. It means that the molecule was previously in the excited state and it handed over some of its intrinsic energy to the incident photon. The scattered photon thus has greater energy or shorter wavelength.

Since the molecules possess quantised energy levels, we can write,

$$E_1 - E_2 = nhv_c \quad \dots(3)$$

where  $n = 1, 2, 3, \dots$  etc, and  $\nu_c$  = the characteristic frequency of the molecule.

In the simplest case  $n = 1$ , equation (2) reduces to

$$\nu_2 = \nu_1 \pm \nu_c \quad \dots(4)$$

Equation (4) shows that the frequency difference  $\nu_1 - \nu_2$  between the incident and scattered photon corresponds to the characteristic frequency  $\nu_c$  of the molecule.

### 14.15 CLASSICAL THEORY OF RAMAN EFFECT

When a molecule is placed in an electric field  $E$ , the field distorts the electronic structure and polarizes the molecule. The polarization induced ( $P$ ) is proportional to the applied electric field ( $E$ ).

$$P = \alpha E \quad \dots(1)$$

Here,  $\alpha$  is the polarizability of the molecule.

When radiation of frequency  $\nu_0$  is allowed to fall on the molecules, each molecule experiences a varying electric field  $E$  given by

$$E = E_0 \cos 2\pi\nu_0 t \quad \dots(2)$$

For a vibrating molecule, the polarizability along a direction will vary about an average value  $\alpha_0$  with a frequency equal to the vibrational frequency. We can write

$$\alpha = \alpha_0 + \beta \cos 2\pi\nu_m t \quad \dots(3)$$

Here,  $\alpha_0$  is the equilibrium polarizability, and  $\beta$  is the rate of change of polarizability with the vibration. Then the induced polarization is given by

$$P = (\alpha_0 + \beta \cos 2\pi\nu_m t) (E_0 \cos 2\pi\nu_0 t) \quad \dots(4)$$

$$\therefore P = \alpha_0 E_0 \cos 2\pi\nu_0 t + \frac{E_0 \beta}{2} [\cos 2\pi(\nu_0 + \nu_m)t + \cos 2\pi(\nu_0 - \nu_m)t] \quad \dots(5)$$

Thus, the induced polarization contains the three distinct frequency components  $\nu_0$ ,  $(\nu_0 + \nu_m)$  and  $(\nu_0 - \nu_m)$ .

$\nu_0$  is the frequency of incident radiation i.e., Rayleigh scattering.

The last two frequencies are more and less than the frequency of incident radiation and predict the existence of Raman scattering.

No Raman line will be observed, if the change in polarizability  $\beta$  is zero.

Thus we have the general rule:

*In order to be Raman active, a molecular rotation or vibration must cause some change in a component of the molecular polarizability.*

### 14.16 VIBRATIONAL RAMAN SPECTRA OF DIATOMIC MOLECULES

The transition energies of the vibrational Raman spectra for a diatomic molecule can be written using the energy expression of an anharmonic oscillator [cf. Eq. (2), Section 14.6].

$$\varepsilon = \bar{\nu}_e \left( V + \frac{1}{2} \right) - \bar{\nu}_e x_e \left( V + \frac{1}{2} \right)^2 m^{-1}, \text{ for } V = 0, 1, 2, \dots \quad \dots(1)$$

Here,  $\bar{\nu}_e$  is called the *equilibrium oscillation frequency* of the molecule expressed in wave numbers and is the value for small displacements.  $x_e$  is the corresponding anharmonicity constant.

The selection rule for Raman scattering is given by  $\Delta V = \pm 1, \pm 2, \dots$

Considering the  $V = 0 \rightarrow V = 1$  transition, we have

$$\Delta\epsilon = \bar{v}_e(1 - 2x_e) m^{-1} \quad \dots(2)$$

During collision, if the molecule gains energy from the exciting radiation  $\bar{v}_0$ , it gives rise to a line on the low frequency side of  $\bar{v}_0$  which is the Stokes Raman line.

In anti-Stokes Raman, the exciting radiation gains energy from the molecule. Hence, we may write

$$\bar{v} = \bar{v}_0 \pm \bar{v}_e(1 - 2x_e) \quad \dots(3)$$

Here the minus sign represents the Stokes' lines. The plus sign refers to the anti-Stoke's lines. Since the Raman scattering of light is of low intensity, we can ignore completely all the weaker effects such as overtones and 'hot' bands.

#### 14.16.1. Pure Rotational Raman Spectra (Linear Molecules)

The rotational energy levels of linear molecules are given by [cf. Eq. (6), Section 14.3]

$$\epsilon_J = BJ(J+1) - DJ^2(J+1)^2 \text{ cm}^{-1} (J=0, 1, 2, \dots)$$

But, in Raman spectroscopy, the precision of the measurements does not normally warrant the retention of the term involving  $D$ , the centrifugal distortion constant. Thus we take the simpler expression:

$$\epsilon_J = BJ(J+1) \text{ cm}^{-1} (J=0, 1, 2, \dots) \quad \dots(1)$$

to represent the energy levels.

Transitions between these levels follow the selection rule:

$$\Delta J = 0, \text{ or } \pm 2 \text{ only} \quad \dots(2)$$

We define  $\Delta J$  as  $(J_{\text{upper state}} - J_{\text{lower state}})$ . We can ignore the selection rule  $\Delta J = -2$  since, for a pure rotational change, the upper state quantum number must necessarily be greater than that in the lower state. Further, the 'transition'  $\Delta J = 0$  is trivial since this represents no change in the molecular energy and hence Rayleigh scattering only.

Combining, then,  $\Delta J = +2$  with the energy levels of Eq. (1) we have:

$$\begin{aligned} \Delta\epsilon &= \epsilon_{J'=J+2} - \epsilon_{J''=J} \\ &= B(4J+6) \text{ cm}^{-1} (J=0, 1, 2, \dots) \end{aligned} \quad \dots(3)$$

The wavenumbers of the rotational Raman lines are given by

$$\bar{v} = \bar{v}_0 \pm B(4J+6) \text{ cm}^{-1} \quad \dots(4)$$

Here, the plus sign refers to anti-Stokes' lines, the minus to Stokes' lines, and  $\bar{v}_0$  is the wavenumber of the exciting radiation.

Stokes' and anti-Stokes' lines have comparable intensity because many rotational levels are populated and hence downward transitions are approximately as likely as upward ones.

Fig. 14.26 shows the rotational energy levels of a diatomic molecule and the rotational Raman spectrum arising from transitions between them. Spectral lines are numbered according to their lower  $J$  values.

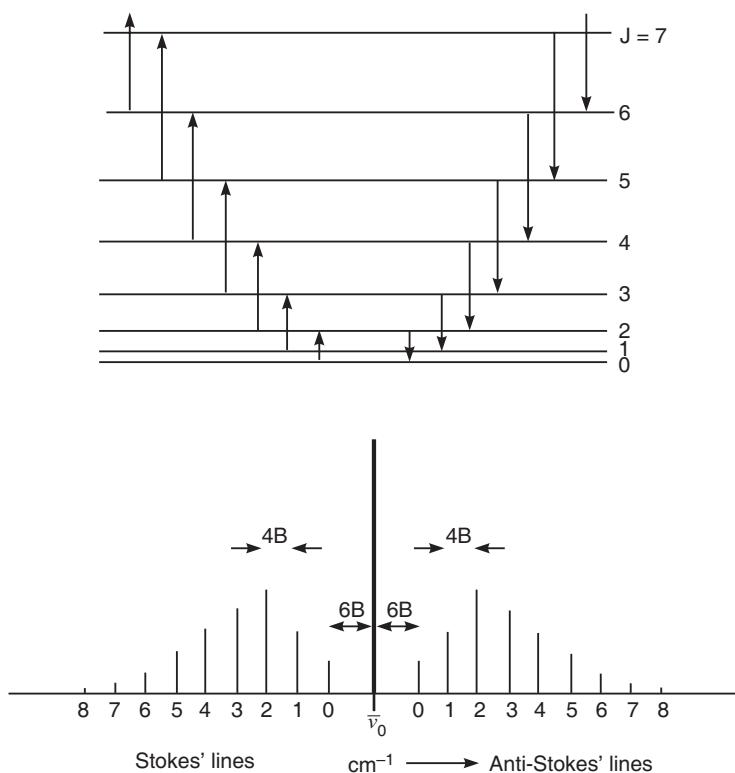


Fig. 14.26

Putting  $J = 0$  in Eq. (4), we find that the separation of the first line from the exciting line is  $6B \text{ cm}^{-1}$ , while the separation between successive lines is  $4B \text{ cm}^{-1}$ .

For diatomic and light triatomic molecules, the rotational Raman spectrum will normally be resolved and we can obtain a value of  $B$ . Hence the moment of inertia and bond length for such molecules can be found.

Homonuclear diatomic molecules (for example  $\text{O}_2$ ,  $\text{H}_2$ ) give no infra-red or microwave spectra since they possess no dipole moment, whereas they *do* give a rotational Raman spectrum. Therefore, structure of such molecules can be obtained from Raman technique. Raman technique is thus complementary to microwave and infra-red studies.

If a molecule has a centre of symmetry (*e.g.*,  $\text{H}_2$ ,  $\text{O}_2$ ,  $\text{CO}_2$ ), then the effects of nuclear spin will be observed in both the Raman and infra-red spectra. Thus for  $\text{O}_2$  and  $\text{CO}_2$  (since the spin of oxygen is zero) every alternate rotational level is absent.

For example, in the case of  $\text{O}_2$ , every level with *even*  $J$  values is missing, and thus every transition labelled  $J = 0, 2, 4, \dots$  in Fig. 14.26 is also completely missing from the spectrum. In the case of  $\text{H}_2$ , and other molecules composed of nuclei with non-zero spin, the spectral lines show an alternation of intensity.

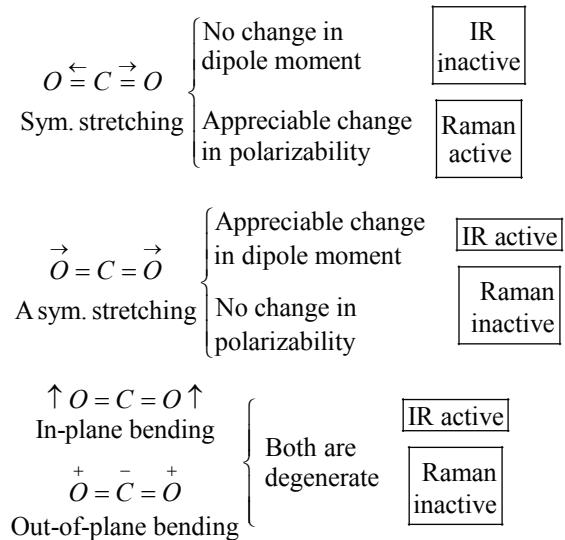
#### 14.16.2. Complementary Character of Raman and Infra-red Spectra

The differences between the Raman spectra and IR spectra are given below:

Raman Spectra	IR Spectra
<ol style="list-style-type: none"> <li>It appears due to the scattering of radiation by the vibrating molecules.</li> <li>The polarizability of the molecule changes.</li> <li>The intensity of the Raman line depends on the magnitude of the change of polarizability of the molecule.</li> <li>Homopolar molecules respond to Raman scattering.</li> </ol>	<ol style="list-style-type: none"> <li>It arises due to the absorption of radiation by the vibrating molecules.</li> <li>The dipole moment of the molecules changes.</li> <li>The intensity of an IR absorption band depends on the magnitude of the change in dipole moment of the molecule.</li> <li>Homopolar molecules cannot exhibit IR spectra.</li> </ol>

For IR spectra to occur, there must be a change in the dipole moment of the molecule. For the Raman spectra to occur, there must be a change in the polarizability of the molecule. Since these two requirements are somewhat different, lines may appear in one of the spectra or in both.

The symmetrical stretchings of the molecule which are usually missing in the IR spectra appear very prominently in Raman spectra, while the asymmetric vibrations exhibit opposite behaviour. It follows that the vibrational modes which are inactive in IR spectra, are sometimes active in Raman spectra. For example, in the case of  $\text{CO}_2$  (a linear molecule),



With small molecules, the less symmetric vibrations often produce intense IR-bands, while the (more) symmetric vibrations often give rise to strong Raman bands.

According to *Rules of Mutual Exclusion*, which states that if a molecule has a centre of symmetry (e.g., centrosymmetric molecule like  $\text{H}_2$ ,  $\text{CO}_2$ ,  $\text{C}_2\text{H}_2$ , etc.), the Raman active vibrations are IR-inactive and *vice versa*. The molecule for which the centre of symmetry is absent, may be simultaneously Raman and IR active. It is, therefore, the stretching vibrations of homonuclear diatomic molecules (such as  $\text{H}_2$ ,  $\text{N}_2$ ,  $\text{O}_2$ , etc.) which are IR-inactive, are Raman active.

Thus, it may be stated that the IR and Raman spectra are complementary to each other.

### 14.17 APPLICATIONS OF RAMAN EFFECT

#### (1) Structure determination from Raman and Infra-red Spectroscopy

**Triatomic molecules:** Dealing with triatomic molecules (of type  $AB_2$ ), the questions to be decided are: whether each molecule is linear or not and, if linear, whether it is symmetrical ( $B-A-B$ ) or unsymmetrical ( $B-B-A$ ). From the number and intensity of the observed lines in the Raman effect in conjunction with infrared data, it is possible to draw important conclusions about molecular structure. Theory leads to the following rule, known as the *rule of mutual exclusion*. It states that *for molecules with a centre of symmetry transitions that are allowed in the infrared are forbidden in Raman spectra and vice versa*. The rule does not imply that all transitions forbidden in one must occur in the other; *i.e.*, some transitions may be forbidden in both. On the other hand certain transitions can occur both in the infra-red and in the Raman spectra in the case of molecules without a centre of symmetry.

**Examples:** (a)  $CO_2$  has two very strong bands in its infrared absorption spectrum at 66800 and  $234900\text{ m}^{-1}$  while only one strong band in its Raman spectrum at  $138900\text{ m}^{-1}$ . None of these bands occur both in Raman and infrared spectra. Hence, it follows from the rule of mutual exclusion, that  $CO_2$  molecule has a centre of symmetry. This implies that the molecule is linear and symmetric and hence it should be represented by  $O - C - O$ .

(b) Nitrous oxide ( $N_2O$ ) has three absorption bands at 222400, 128500 and  $58900\text{ m}^{-1}$  of which the first two appear in the Raman spectrum. Thus the molecule cannot have a centre of symmetry though linear. Hence the molecule has the unsymmetrical structure  $N - N - O$ .

In a similar manner, the bent symmetric structure of a water molecule represented by Fig. 14.27 is revealed by Raman Effect. Thus the study of Raman spectra of different substances enables one to classify them according to their molecular structure.

**(2) Raman effect and molecular structure.** A qualitative identification of Raman spectrum consists in the evaluation of the wavelength of the lines, their intensities, and state of polarisation. Investigation of bond angles, bond stiffness, and other structural confirmation require Raman data in addition to infrared studies.

**Diatomc molecules.** From the analysis of Raman spectra of a diatomic molecule, we can have an idea about the nature of the chemical bond existing between the atoms. In a diatomic molecule, the frequency of vibration of the atoms is given by  $\nu = \frac{1}{2\pi} \sqrt{\frac{F}{\mu}}$  where  $F$  is the restoring force per

unit displacement and  $\mu$  is the reduced mass of the molecule. It is seen that a molecule in which the force binding the atoms is great should have higher characteristic frequency than one in which the force is weak. This force depends upon the nature of interatomic bonds. In covalent molecules polarisability is considerably changed by the nuclear oscillations due to the nuclei. This appreciable change in polarisability gives rise to intense Raman lines. In electrovalent molecules, the binding electrons definitely change over from one nucleus to the other in the formation of the molecule so that the polarisability of the molecule is little affected by nuclear oscillations and hence no Raman lines will appear.

(3) Raman effect in crystals is complementary to X-ray crystal study and provides information about the binding forces in crystals.

(4) Various chemical effects like strength of chemical bonds, electrolytic dissociation, hydrolysis, etc., have been understood through Raman effect. Specific heat capacities of solids, brilliance of metals and their molecular structure have been explained by Raman effect.

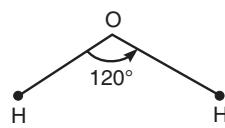
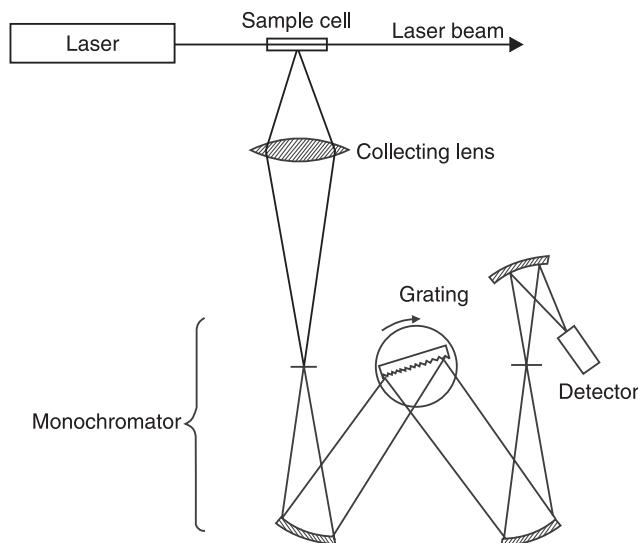


Fig. 14.27

### 14.18 TECHNIQUES AND INSTRUMENTATION

#### Raman Spectrometer

Fig. 14.28 shows the schematic diagram of a Raman spectrometer.



**Fig. 14.28**

**Working.** The laser beam is passed through a cell, usually a narrow glass or quartz tube filled with the sample. Here the major considerations are the effective illumination of the sample and collection of scattered radiation from the sample. Light scattered sideways from the sample is collected by a lens and passed into a grating monochromator. The signal is measured by a sensitive photomultiplier and after amplification, it is processed by a computer which plots the Raman spectrum.

We briefly discuss the components of the Raman spectrometer.

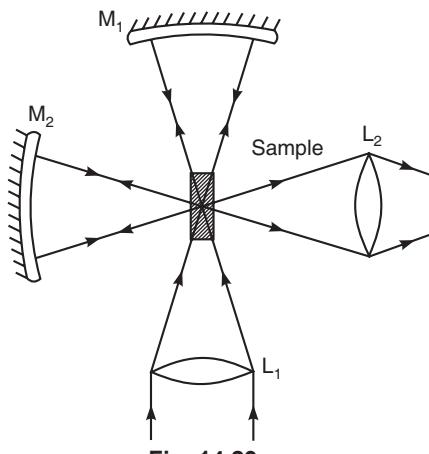
(i) **Source:** The highly intense, monochromatic, coherent, and directional properties of a laser allow one to obtain easily a high quality Raman spectrum. The fluorescence is expected to be low as the frequency of the radiation is less. The commonly used lasers are: He–Ne laser (632.8 nm), argon ion laser (488 and 514.5 nm) and krypton laser (647.1, 568.2, 530.8, 520.8, 482.5, 476.2 nm). In certain cases, red radiation is preferred to reduce fluorescence and decomposition of the sample.

(ii) **Sample illumination:** Here, the major considerations are the effective illumination of the sample and collection of scattered radiation from the sample.

Fig. 14.29 shows the arrangement for sample illumination.

- The lens  $L_1$  focuses the laser beam into the sample.
- The lens  $L_2$  collects the scattered radiation and focuses it to the dispersing system.
- Mirrors  $M_1$  and  $M_2$  together increase the intensity of the collected radiation.

(iii) **Dispersing medium:** The dispersing element allows only a narrow region to reach the detector at a



**Fig. 14.29**

time. Usually a diffraction grating is used as the dispersing element. As a single grating cannot give the required resolution, often two gratings are used in series. Then the system is called a double monochromator. For very high resolution work, a triple monochromator is preferred. Rotation of the grating allows successive regions of the spectrum to reach the detector.

(iv) **Detector:** A photomultiplier tube is used as the detecting device. To reduce thermal emission, it has to be cooled. The output of the photomultiplier tube is amplified and fed to the data processing system or recorder.

### Advantages

For vibrational measurements, the Raman technique has several advantages over infra-red.

(1) Because both the incident and scattered radiation are at ultraviolet or visible frequencies, conventional optics and sample cells (glass, quartz, etc.) can be used, so avoiding the problems inherent in NaCl windows, atmospheric absorption, etc.

(2) Because the beam can be focused extremely finely (diameters as small as 0.1 nm are possible), very small samples can be studied. This, combined with pulsed techniques which can give very short time resolutions, enables very small quantities of transient species to be studied.

(3) Water, which has strong infra-red absorptions, is a rather weak Raman scatterer. So aqueous solutions can be studied using Raman, because the sample signal is not swamped by that of the solvent.

- In addition to liquids and solutions, Raman spectra can be obtained from gas and solid samples.

### Electronic Spectra Molecules

## 14.19 ELECTRONIC SPECTRA OF DIATOMIC MOLECULES

### 14.19.1. Born-Oppenheimer Approximation

According to Born and Oppenheimer, the various forms of energies of a molecule are independent of each other.

The total energy  $E_{\text{total}}$  of a molecule is

$$E_{\text{total}} = E_{\text{electronic}} + E_{\text{vibration}} + E_{\text{rotation}} \quad \dots (1)$$

Eq. (1) implies that the electronic, vibrational, and rotational energies of a molecule are completely independent of each other.

A change in the total energy of a molecule is

$$\begin{aligned} \Delta E_{\text{total}} &= \Delta E_{\text{elec.}} + \Delta E_{\text{vib.}} + \Delta E_{\text{rot}} \text{ J} \\ \text{or} \quad \Delta \epsilon_{\text{elec.}} &= \Delta \epsilon_{\text{elec.}} + \Delta \epsilon_{\text{vib.}} + \Delta \epsilon_{\text{rot}} \text{ cm}^{-1} \end{aligned} \quad \dots (2)$$

The approximate orders of magnitude of these changes are:

$$\Delta \epsilon_{\text{elec.}} = \Delta \epsilon_{\text{vib.}} \times 10^3 \approx \Delta \epsilon_{\text{rot}} \times 10^6 \quad \dots (3)$$

So we see that *vibrational energy changes will produce a ‘coarse structure’ and rotational changes a ‘fine structure’ on the spectra of electronic transitions.*

- Pure rotation spectra are shown only by molecules possessing a permanent electric dipole moment.
- Vibrational spectra require a change of dipole moment during its normal mode of vibration.

- Electronic spectra are given by *all* molecules since changes in the electron distribution in a molecule are always accompanied by a dipole change. This means that homonuclear molecules (for example H<sub>2</sub> or N<sub>2</sub>), which show no rotation or vibration-rotation spectra, *do* give an electronic spectrum and show vibrational and rotational structure in their spectra. From this, rotational constants and bond vibration frequencies are derived.

#### 14.19.2. Vibrational Coarse Structure: Progressions and Sequences

An electronic transition involves a change in electronic, vibrational and rotational energies of the molecule. Ignoring rotational energy changes,

$$\begin{aligned} E_{\text{total}} &= E_{\text{elec.}} + E_{\text{vib. J}} \\ \varepsilon_{\text{total}} &= \varepsilon_{\text{elec.}} + \varepsilon_{\text{vib.}} \text{ cm}^{-1} \end{aligned} \quad \dots (1)$$

From Eq. (2) in section 14.6, we can write immediately

$$\varepsilon_{\text{total}} = \varepsilon_{\text{elec.}} + \left( v + \frac{1}{2} \right) \bar{\omega}_e - x_e \left( v + \frac{1}{2} \right)^2 \bar{\omega}_e \text{ cm}^{-1} (v = 0, 1, 2, \dots)$$

Fig. 14.30 shows the energy levels of Eq. (2) for two arbitrary values of  $\varepsilon'_{\text{elec.}}$

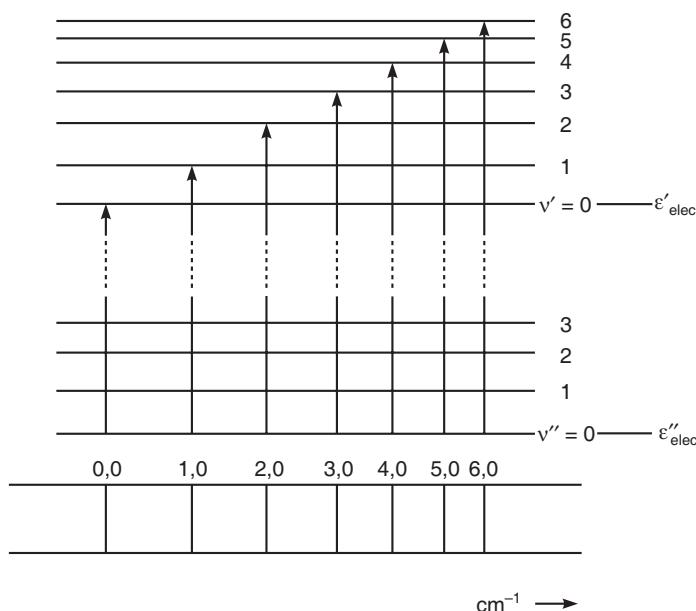


Fig. 14.30

- The lower states are distinguished by a *double prime* ( $v'', \varepsilon''_{\text{elec.}}$ )
- The upper states carry only a *single prime* ( $v', \varepsilon'_{\text{elec.}}$ ).
- An excited electronic state corresponds to a weaker bond in the molecule. So the spacing between the upper vibrational levels is smaller than that between the lower.

**Abosorption Spectrum.** There is no selection rule for  $v$  when a molecule undergoes an electronic transition, *i.e.*, every transition  $v'' \rightarrow v'$  has some probability. A great many spectral lines would, therefore, be expected. However, the situation is considerably simplified if the *absorption* spectrum is considered from the electronic ground state.

In this case, virtually all the molecules exist in the lowest vibrational state, that is  $v'' = 0$ .

So the only transitions to be observed with appreciable intensity are those indicated in Fig. 14.30.

- They are labelled according to their  $(v', v'')$  numbers (note: upper state first), that is  $(0, 0)$ ,  $(1, 0)$ ,  $(2, 0)$ ,  $(3, 0)$  etc. Such a set of transitions is called a *band* since, under low resolution, each line of the set appears somewhat broad and diffuse.
- The set of lines corresponding to these transitions is called a *v' progression*, since the value of  $v'$  increases by unity for each line in the set. The diagram shows that the lines in a band crowd together more closely at high frequencies; this is a direct consequence of the anharmonicity of the upper state vibration which causes the excited vibrational levels to converge.
- Fig. 14.30 shows the *vibrational 'coarse' structure of the band formed during electronic absorption from the ground ( $v'' = 0$ ) state to a higher state*.

We can write an analytical expression for this spectrum. From Eq. (2), we have

$$\Delta\epsilon_{\text{total}} = \Delta\epsilon_{\text{elec.}} + \Delta\epsilon_{\text{vib.}}$$

$$\bar{\nu}_{\text{spec}} = (\epsilon' - \epsilon'') + \left\{ \left( v' + \frac{1}{2} \right) \bar{\omega}_e' - x_e' \left( v' + \frac{1}{2} \right)^2 \bar{\omega}_e' \right\} - \left\{ \left( v'' + \frac{1}{2} \right) \bar{\omega}_e'' - x_e'' \left( v'' + \frac{1}{2} \right)^2 \bar{\omega}_e'' \right\} \text{ cm}^{-1} \quad \dots (3)$$

The observation of few lines in a band enables one to calculate the values for  $\bar{\omega}_e'$ ,  $x_e'$ ,  $\bar{\omega}_e''$ , and  $x_e''$ , as well the separation between electronic states,  $(\epsilon' - \epsilon'')$ . Thus the observation of a band spectrum leads not only to values of the vibrational frequency and anharmonicity constant in the ground state ( $\bar{\omega}_e'$  and  $x_e'$ ), but also to these parameters in the excited electronic state ( $\bar{\omega}_e''$  and  $x_e''$ ).

### Progressions and Sequences

- The vibrational transitions accompanying an electronic transition are called vibronic transitions. The vibronic transitions are divided into progressions and sequences.

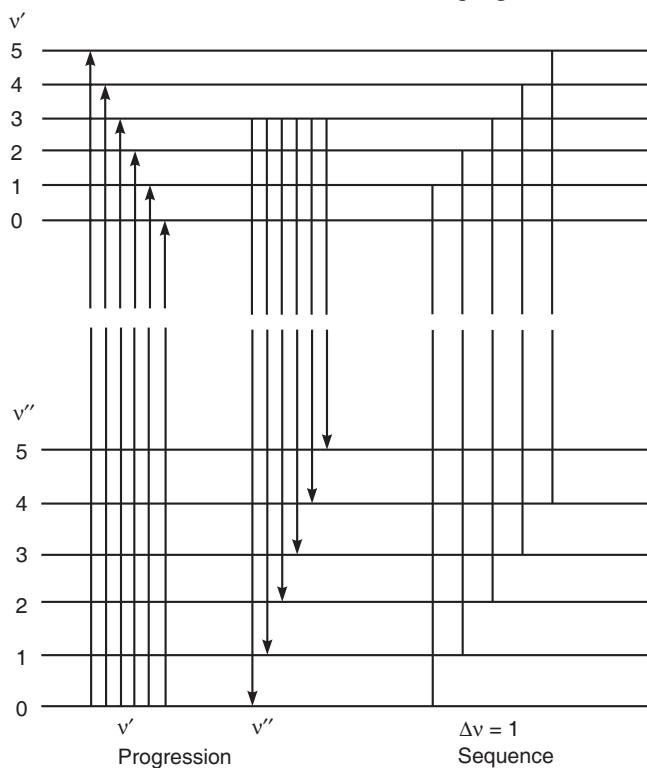


Fig. 14.31

- The first step in the vibrational analysis of a spectrum is the assignment of ( $v'$ ,  $v''$ ) values for a band. It would be easier to do the assignment if progressions and sequences are picked out first.

Fig. 14.31 shows vibrational progressions and sequence in the electronic spectrum of a diatomic molecule.

**Regression.** A progression involves a series of vibronic transitions with a common lower or upper level. For example, the  $v''$  progression members all have  $v' = 3$  level common and  $v' = 0$  progression members all have  $v'' = 0$  level common ( $v' = 0, 1, 2, 3, \dots$ ). A  $v''$  progression extends towards the lower wavenumber as  $v''$  increases and terminates in a continuum where the lower electronic state dissociates. A  $v'$  progression extends towards higher wavenumber as  $v'$  increases and terminates in a continuum where the upper electronic state dissociates.

**Sequence.** A group of transitions with the same value of  $\Delta v$  is called a sequence (Fig. 14.31). Bands such as (0, 0), (1, 1), (2, 2), (3, 3), ..., (0, 1), (1, 2), (2, 3), (3, 4), ..., (0, 2), (1, 3), (2, 4), ... are called **sequences**. The wave numbers of the bands in a sequence do not differ considerably.

#### 14.19.3. Intensity of Vibrational-Electronic Spectra: The Franck-Condon Principle

The intensity distribution of the lines in the electronic spectra of molecules is not the same. The vibrational lines in a progression are not all observed to be of the same intensity.

- In some spectra, the (0, 0) transition is the strongest.
- In others, the intensity increases to a maximum at some value of  $v'$ .
- In some cases, only a few vibrational lines with high  $v'$  are seen, followed by a continuum.

All these types of spectrum are explained on the basis of the *Franck-Condon principle*.

**Statement.** *An electronic transition takes place so rapidly that a vibrating molecule does not change its internuclear distance appreciably during the transition.*

**Morse potential energy curve.** The anharmonic model of a diatomic molecule has a potential energy curve represented by the Morse function. Figure 14.32 shows the Morse potential energy curve along with the probability densities  $|\Psi|^2$  as a function of the internuclear distance for the lowest five vibrational states.

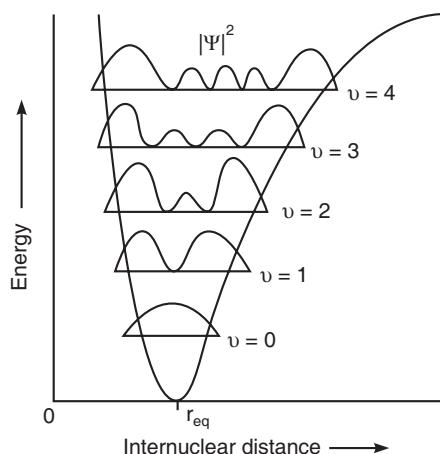


Fig. 14.32

- For  $v = 0$ , the atom is most likely to be found at the *center* of its motion, i.e., at the equilibrium internuclear distance  $r_{eq}$ .

- For  $v = 1, 2, 3, \dots$  the most probable positions steadily approach the extremities. This behaviour is shown in Fig. 14.32 where we plot the probability distribution in each vibrational state against internuclear distance.
- Fig. 14.32 shows the variation of  $\psi^2$  with internuclear distance, where  $\psi$  is the vibrational wave function. *The nuclei are most likely to be found at distances apart given by the maxima of the curve for each vibrational state.*

### Explanation of Franck Condon Principle

In a molecule, the individual atoms vibrate about a mean position. Therefore, the internuclear distance changes continuously. According to Franck-Condon principle, during an electronic transition, the internuclear distance remains the same.

- Electrons move and rearrange themselves much faster than the nuclei of molecules. For example, the time for an electron to circle a hydrogen nucleus is of the order of  $10^{-16}$ s, which is thousand times less than the typical period of vibration  $10^{-13}$ s of the molecule. Thus the internuclear distance ( $R$ ) of a vibrating molecule does not change appreciably in course of transition between two electronic states. Hence *the transition should be represented by a vertical line*, that is a line of constant  $R$ .

Franck-Condon principle can be stated thus: transition between electronic states occur vertically in a potential energy diagram.

- The straight line representing a transition between electronic states will be vertical.

### The operation of Franck-Condon principle

Consider a diatomic molecule undergoing transition from  $E_e''$  to  $E_e'$ .

There are four possibilities.

#### Case (i) Internuclear distances equal in the upper and lower states

In Fig. 14.33, we show the upper electronic state having the same equilibrium internuclear distance as the lower. Now the Franck-Condon principle suggests that a transition occurs *vertically* on this diagram, since the internuclear distance does not change. So if we consider the molecule to be initially in the ground state both electronically ( $\epsilon''$ ) and vibrationally ( $v'' = 0$ ), then the most probable transition is that indicated by the vertical line in Fig. 14.33. Thus the strongest spectral line of the  $v'' = 0$  progression will be the (0, 0). However, the quantum theory only says that the *probability* of finding the oscillating atom is greatest at the equilibrium distance in the  $v = 0$  state—it allows some, although small, chance of the atom being near the extremities of its vibrational motion. Hence there is some chance of the transition starting from the ends of the  $v'' = 0$  state and finishing in the  $v' = 1, 2, \dots$  states. The (1, 0), (2, 0) etc., lines diminish rapidly in intensity, however, as shown at the foot of Fig. 14.33.

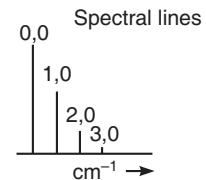
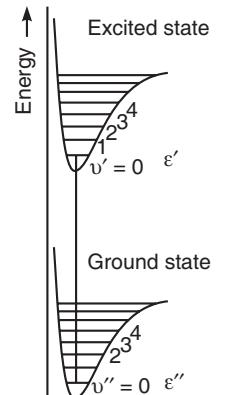


Fig. 14.33

**Case (ii) Upper state internuclear distance a little less than that in the lower state.** In Fig. 14.34, the excited electronic state has a *slightly* smaller internuclear separation than the ground state.

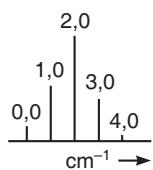
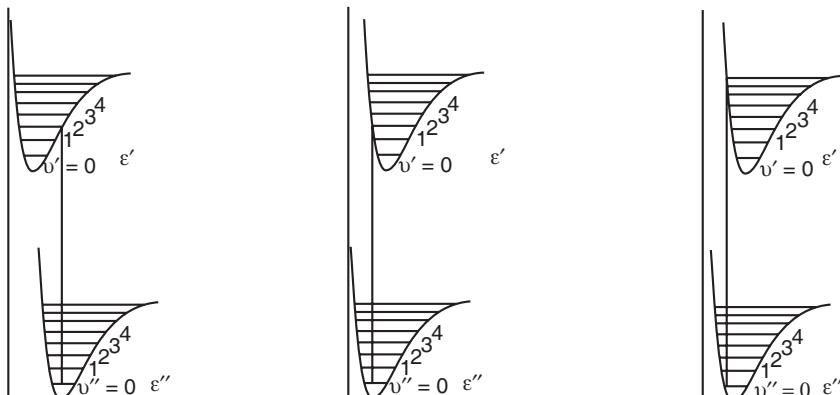


Fig. 14.34

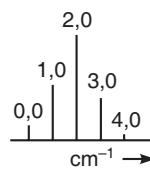


Fig. 14.35

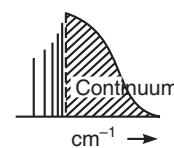


Fig. 14.36

- A vertical transition from the  $v'' = 0$  level will be most likely to occur into the upper vibrational state  $v' = 2$ .
- Transitions to lower and higher  $v'$  states are less likely to occur.
- The most intense will be the (2, 0) transition.

**Case (iii) Upper state distance a little greater than in the lower.** In Fig. 14.35, the excited electronic state has a slightly larger internuclear separation than the ground state, but the resulting transitions and spectrum are similar to case (ii).

**Case (iv) Upper state distance considerably greater.** In Fig. 14.36, the upper state separation is drawn as *considerably* greater than that in the lower state.

The vibrational level to which a transition takes place has a high  $v'$  value.

The vertical line from the maximum  $v'' = 0$  goes straight to the continuum of states.

- The transitions are not quantized and a continuum results. This is shown at the foot of the figure.

Thus, Franck–Condon principle is able to account for the intensities of lines in vibrational electronic spectra.

#### 14.19.4. Dissociation Energy and Dissociation Products

- The energy required to separate a stable diatomic molecule  $AB$  in the  $v = 0$  vibrational state into to the two unexcited atoms  $A$  and  $B$  is called the **dissociation energy**  $D_0$ .

Fig. 14.37 illustrates dissociation by excitation into a stable upper state. The equilibrium nuclear separation in the upper state is considerably greater than that in the lower.

- The dashed line limits of the Morse curves represent the dissociation of the normal and excited molecule into atoms.

The dissociation energies are  $D_0''$  and  $D_0'$  from the  $v = 0$  state in each case.

- The total energy of the dissociation products (i.e., atoms) from the upper state is greater by an amount called  $E_{ex}$ , than that of the products of dissociation in the lower state. This energy is the *excitation energy* of one (or rarely both) of the atoms produced on dissociation.
- The spectrum of this system consists of some vibrational transitions (quantized) followed by a continuum (non-quantized) transitions representing dissociation. The lower wavenumber limit of this continuum must represent just sufficient energy to cause dissociation and no more (i.e., the dissociation products separate with virtually zero kinetic energy).

$$\therefore \bar{v}_{cont\ limit} = D_0'' + E_{ex} \text{ cm}^{-1} \quad \dots (1)$$

### Determination of dissociation energy

- Generally a molecule in the state corresponding to the continuum limit dissociates into a normal atom and an excited atom with an excitation energy  $E_{ex}$ . This  $E_{ex}$  is the total energy of the dissociation products in the upper state minus the total energy of the dissociation products in the lower state. When the molecule dissociates from one of the continuum states, the energy in excess of continuum limit appears as kinetic energy of the separated atoms. The energy balance requires

$$\bar{v}_{cont\ limit} = D_0'' + E_{ex}$$

Value of  $\bar{v}_{cont\ limit}$  is known from electronic spectroscopy.

The value of  $E_{ex}$  is known from atomic spectroscopy.

Using Eq. (1), we can calculate the dissociation energy  $D_0''$ .

- In many electronic spectra no continua appear at all—the internuclear distances in the upper and lower states are such that transitions near to the dissociation limit are of negligible probability—but it is still possible to derive a value for the dissociation energy by noting how the vibrational lines converge.

The vibrational energy  $\varepsilon_v$  of a diatomic molecule is

$$\varepsilon_v = \left( v + \frac{1}{2} \right) \bar{\omega}_e - x_e \left( v + \frac{1}{2} \right)^2 \bar{\omega}_e \text{ cm}^{-1} \quad \dots (2)$$

The separation between neighbouring levels,  $\Delta\varepsilon$ , is

$$\begin{aligned} \Delta\varepsilon &= \varepsilon_{v+1} - \varepsilon_v \\ &= \bar{\omega}_e [1 - 2x_e(v + 1)] \text{ cm}^{-1} \end{aligned} \quad \dots (3)$$

This separation obviously decreases linearly with increasing  $v$  and the dissociation limit is reached when  $\Delta\varepsilon \rightarrow 0$ . Thus the maximum value of  $v$  is given by  $v_{max}$ , where

$$\bar{\omega}_e [1 - 2x_e(v_{max} + 1)] = 0$$

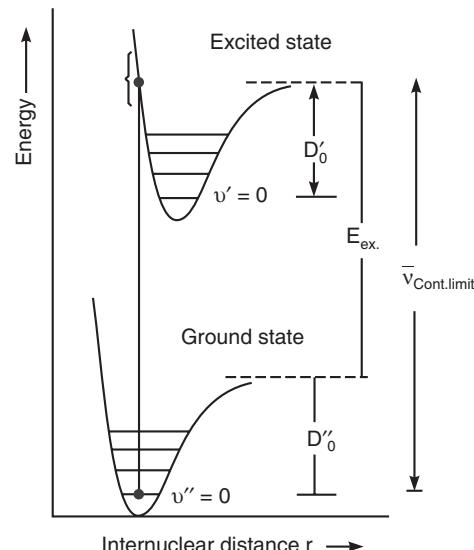


Fig. 14.37

$$\therefore v_{\max} = \frac{1}{2x_e} - 1 \quad \dots (4)$$

The anharmonicity constant,  $x_e$ , is of the order of  $10^{-2}$ ; hence  $v_{\max}$  is about 50.

The vibrational energy for  $v = v_{\max}$  is the dissociation energy  $D_e$ . Thus, we have

$$\begin{aligned} D_e &= \left( \frac{1}{2x_e} - \frac{1}{2} \right) \bar{\omega}_e - \left( \frac{1}{2x_e} - \frac{1}{2} \right)^2 \bar{\omega}_e x_e \\ &= \frac{\bar{\omega}_e}{4x_e} - \frac{1}{4} \bar{\omega}_e x_e \approx \frac{\bar{\omega}_e}{4x_e} \text{ cm}^{-1} \end{aligned}$$

For the known values of  $\bar{\omega}_e$  and  $\bar{\omega}_e x_e$ , the value of  $D_e$  can be obtained.

- The equilibrium dissociation energy  $D_e$  is different from  $D_0$  and is given by

$$D_e = D_0 + \varepsilon_0 \text{ cm}^{-1} \quad \dots (5)$$

where  $\varepsilon_0$  is the zero point energy.

From the known value  $D_e$ , the value of  $D_0''$  can be evaluated.

## EXERCISES

### Part A: Choose the Correct Answer

1. The rotational spectroscopy is in the
    - (a) microwave region
    - (b) infrared region
    - (c) visible region
    - (d) radio frequency region
  2. Which of the following molecule is microwave inactive?
    - (a) HCl
    - (b) H<sub>2</sub>
    - (c) CO
    - (d) none of these
  3. Rotational spectra are always obtained in
    - (a) emission
    - (b) absorption
    - (c) dispersion
    - (d) deviation
  4. The assumption that the electronic, vibrational and rotational modes of excitation of molecules are non-interacting is known as
    - (a) Frank-condon approximation
    - (b) first order approximation
    - (c) second order approximation
    - (d) Born-Oppenheimer approximation
  5. Pure rotational spectrum of a diatomic molecule consists of
    - (a) two equally spaced lines
    - (b) three equally spaced lines
    - (c) many equally spaced lines
    - (d) no regular pattern
- (B.U. M.Sc. 2014)
6. When compared to the energy levels of rigid rotator, the levels of non-rigid rotator are
    - (a) slightly displaced to low frequencies
    - (b) such that displacement increases with increasing value of  $(J+1)^3$
    - (c) such that the displacement between successive lines is not constant
    - (d) all the above
  7. In Raman spectrum, the Raman shift is
    - (a) a function of wavelength of incident light
    - (b) dependent on the nature of scattering substance
    - (c) independent of the nature of scattering substance
    - (d) dependent on temperature of the substance
  8. The pure rotational levels of a molecule in the far-infrared region follows the formula  $F(J) = BJ(J+1)$ , where  $F(J)$  is the energy of the rotational level with quantum number  $J$  and  $B$  is the rotational constant.

The lowest rotational energy gap in rotational Raman spectrum is



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

11. Describe the general features of molecular spectra.
  12. Give the elementary theory of the origin of pure rotational spectrum of a molecule.
  13. Discuss rotational spectra of a diatomic molecule, treated as a non-rigid rotator.
  14. Derive an expression for energy of rotational states of non-rigid rotator and show how force constant of the molecule can be determined.
  15. Give the elementary theory of the origin of the vibrational spectrum of a molecule.
  16. Discuss how the study of vibrational spectrum of a diatomic molecule enables us to determine the anharmonicity constant and equilibrium frequency of vibration.
  17. Explain, giving theory, the method of finding the force constant of a diatomic molecule from its anharmonic vibration spectrum.
  18. Explain the rotational-vibrational spectra of diatomic molecules. *(Osmania 1992)*
  19. What parameters can one get from a study of the rotation-vibration spectrum of a heteronuclear diatomic molecule? How are they estimated?
  20. Give a brief account of any method of producing and studying the infra-red spectrum. What is the importance of such a study?
  21. Give the theory of Raman effect and describe an experimental arrangement for studying it.
  22. Give the classical and quantum explanation of Raman effect. *(B.U.M.Sc. 2014)*
  23. Discuss the main features of the vibrational and rotational Raman spectra of diatomic molecules. Give the necessary theory. What light does it throw on the structure of the molecules? *(P.U. 93)*
  24. Discuss Raman spectra of diatomic molecules and point out the similarity and difference in infra-red and Raman spectra. How are Raman spectra studied in the laboratory?
  25. The first rotational Raman line of  $H_2$  appears at  $34600\text{ m}^{-1}$  from the exciting line. Calculate the bond length of  $H_2$  molecule, Given  $^1H = 1.673 \times 10^{-27}\text{ kg}$ .

**Sol.** The first rotational Raman line appears at a distance of  $6B$  from the exciting line.

$$6 B = 34600 \text{ m}^{-1}; B = 5766.7 \text{ m}^{-1} = \frac{h}{8\pi^2 Ic}$$

$$I = \frac{h}{8\pi^2 c B} = \frac{6.626 \times 10^{-34} \text{ Js}}{8\pi^2 (3 \times 10^8 \text{ m/s})(5766.7 \text{ m}^{-1})} = 4.8514 \times 10^{-48} \text{ kg m}^2$$

$$\text{Reduced mass, } \mu = \frac{m_1 m_2}{m_1 + m_2} = \frac{(1.673 \times 10^{-27} \text{ kg})(1.673 \times 10^{-27} \text{ kg})}{2 \times 1.673 \times 10^{-27} \text{ kg}} = 0.8365 \times 10^{-27} \text{ kg}$$

$$r^2 = \frac{I}{\mu} = \frac{4.8514 \times 10^{-48} \text{ kg m}^2}{0.8365 \times 10^{-27} \text{ kg}} = 0.5800 \times 10^{-20} \text{ m}^2$$

$$\therefore r = 0.762 \times 10^{-10} \text{ m}$$

26. What is an electronic band system? Explain the origin of progression, sequences and formation of band heads. (P.U. 90)  
27. Describe in detail the Franck Condon Principle in absorption. Discuss its importance. (P.U. 90 93, 94)

## SPECTRA OF ALKALI AND ALKALINE EARTH ELEMENTS

After reading this chapter, you should be able to

- ◆ Explain why the spectra of alkali metals are identical with the spectrum of H atom.
- ◆ Explain the origin of quantum defect.

### 15.1 DIFFERENT SERIES IN ALKALI SPECTRA: MAIN FEATURES

The main examples of alkali metals are lithium, sodium, potassium, rubidium and cesium. In case of alkali metals, the outermost shell contains a single electron. This electron is called the ‘optical’ or ‘valence’ electron. The spectral lines of an alkali atom arise due to the transitions of this single valence electron. The alkali spectrum is called ‘one-electron’ spectrum.

The alkali spectrum consists of spectral lines which can be classified into four series : principal series, sharp series, diffuse series and fundamental series. The principal series is the most prominent and can be observed in emission as well as in absorption spectrum. The other series are observed in emission spectrum only.

**1. Principal series.** The series arises from the transitions between various  $P$ -levels and lowest  $S$ -level. The lowest  $S$ -level has the lowest possible value for the energy and represents the ground state of the atom. The wave number of the series is given by the relation,

$$\bar{v}_p = \frac{R}{(1+\mu_s)^2} - \frac{R}{(m+\mu_p)^2}, \quad m \geq 2.$$

**2. Sharp series.** This type of series comes from the transitions from  $S$ -levels (exclusive of the lowest) to the lowest of the  $P$ -levels.

The wave number of the series is given by the relation

$$\bar{v}_s = \frac{R}{(2+\mu_p)^2} - \frac{R}{(m+\mu_s)^2}, \quad m \geq 2.$$

**3. The diffuse series.** It arises from the transitions between the various  $D$ -levels and the lowest  $P$ -level.

The wave number of the series is given by the relation

$$\bar{v}_D = \frac{R}{(2+\mu_p)^2} - \frac{R}{(m+\mu_D)^2}, \quad m \geq 3.$$

**4. Fundamental series or Bergmann series.** The fundamental series arises from the transitions from various  $F$ -levels to the lowest  $D$ -level.

The wave number of the series is expressed by the relation

$$\bar{v}_F = \frac{R}{(3+\mu_D)^2} - \frac{R}{(m+\mu_F)^2}, \quad m \geq 4.$$

Here  $\mu_s$ ,  $\mu_p$ ,  $\mu_D$  and  $\mu_F$  are the characteristic constants of sharp, principal, diffuse and fundamental series, respectively. The abbreviated forms for writing the wave number of each series are as follows:

- |   |                     |
|---|---------------------|
| 1. $(1, \mu_s) \leftarrow (m, \mu_p)$ with $m \geq 2$ | Principal series.   |
| 2. $(2, \mu_p) \leftarrow (m, \mu_s)$ with $m \geq 2$ | Sharp series.       |
| 3. $(2, \mu_p) \leftarrow (m, \mu_D)$ with $m \geq 3$ | Diffuse series.     |
| 4. $(3, \mu_D) \leftarrow (m, \mu_F)$ with $m \geq 4$ | Fundamental series. |

The other important features of the series are:

1. The sharp and diffuse series have a common limit.
2. The wave number interval between this common limit and the limit of the principal series is equal to that of first line of principal series.

## 15.2 ENERGY LEVELS IN ALKALI SPECTRA AND QUANTUM DEFECT

The emission of alkali spectral lines can be fairly explained on the same lines as the Bohr-Sommerfeld theory for hydrogen atom. An atom has a number of discrete energy states. Each state is characterised by a total quantum number  $n (= 1, 2, 3, \dots \infty)$ . For each value of  $n$ , there are component levels labelled by an additional quantum number  $l$ , called the ‘orbital’ quantum number.  $l$  can take values,  $0, 1, 2, \dots (n-1)$ . Thus  $n = 1$  state has only one level ( $l=0$ );  $n = 2$  state has two levels ( $l=0, 1$ ), and so on. The levels corresponding to  $l=0, 1, 2, 3, \dots$  are called  $s, p, d, f, \dots$  levels respectively. Thus  $n = 1$  state has a level called  $1s$ ; the  $n = 2$  state has two levels called  $2s$  and  $2p$ ; the  $n = 3$  state has three levels called  $3s, 3p$  and  $3d$ ; the  $n = 4$  state has four levels  $4s, 4p, 4d, 4f$ , and so on.

The energies of these levels are given by

$$E_{n,l} = -\frac{Rhc}{(n-\Delta)^2}.$$

Here,  $\Delta$  is called ‘quantum defect’ and depends on  $l$ . Thus the energies of levels with same  $n$  but different  $l$  are *different*.

## 15.3 NONPENETRATING AND PENETRATING ORBITS

(i) **Nonpenetrating Orbits.** We assume that the charge due to the electrons in the closed core is distributed uniformly over a hollow sphere of radius  $\rho$  with the nucleus ( $N$ ) at its centre (Fig. 15.1). Suppose in the course of its motion round the nucleus, the elliptic orbit of the single optical electron in an alkali atom of atomic number  $Z$  does not penetrate the closed orbits or core of  $(Z-1)$  electrons. We expect the energy of the optical electron in the orbit to be nearly the same as that of the electron in the corresponding orbit in a hydrogen atom, on the Bohr-Sommerfeld model. It is found that the energies of the  $F$  states of the sodium atom agree to within one percent with the corresponding energy values of the hydrogen atom. Such orbits for which the energies are nearly the same as in hydrogen atoms are called *nonpenetrating orbits*.

*Nonpenetrating orbits are defined as those orbits for which the observed energies are very nearly equal to those of the corresponding hydrogen-like orbits.*

*The orbits in all of the alkali metals are good examples of nonpenetrating orbits.*

The resultant field under which the optical electron describes the orbit varies slowly as this electron moves in its orbit about the nucleus and the closed electron core. So the elliptic Bohr-Sommerfeld orbits in these cases precess slowly round the nucleus.

(ii) **Penetrating Orbits.** If the optical electron orbit in an alkali atom penetrates the completed electron shells or sub-shells, more and more of the closed electron core is left behind as the optical electron moves in its orbit from aphelion to perihelion. This results in an increase in the field under which this electron traces its path as its distance from the nucleus goes from the maximum to the minimum value (Fig. 15.2). So the electron traverses a path of increasing eccentricity. The increase in the force of attraction on the optical electron makes its energy in the orbit less than that for a corresponding Bohr-Sommerfeld state in a hydrogen atom. Thus the energies of the  $S$  and  $P$  states in the sodium atom are less than those of the corresponding states of the hydrogen atom. Such orbits for which the energies are considerably different from the corresponding values in the hydrogen atom are called *penetrating orbits*.

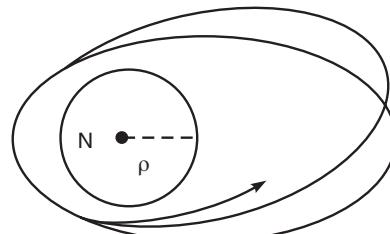


Fig. 15.1

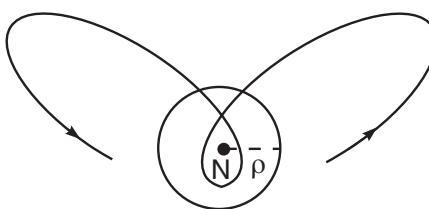


Fig. 15.2

### 15.4 EXPLANATION OF SALIENT FEATURES OF ALKALI SPECTRA

The electronic configuration of alkali metals is such that the core of an inert gas is surrounded by an  $s$  electron.

$$\begin{aligned} \text{Li} &= [\text{He}] + 2s^1 \\ \text{Na} &= [\text{Ne}] + 3s^1 \\ \text{K} &= [\text{Ar}] + 4s^1 \\ \text{Rb} &= [\text{Kr}] + 5s^1 \\ \text{Cs} &= [\text{Xe}] + 6s^1 \end{aligned}$$

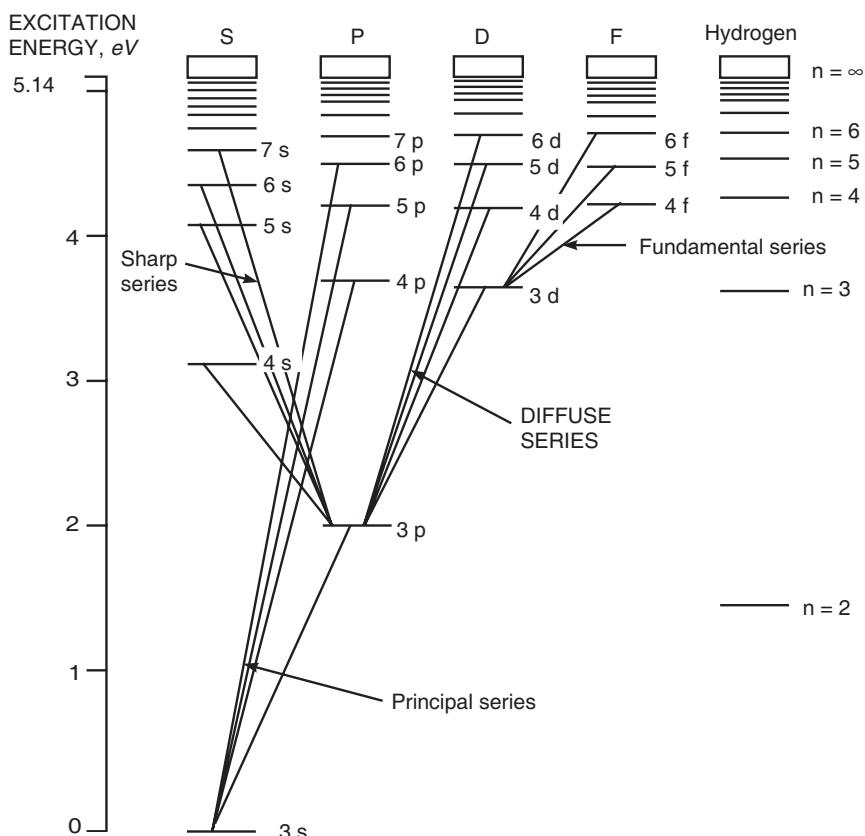
The core electrons do not play any part in optical spectra. Only valence electrons, also called *optical electron* determine the characteristics of optical spectra. The configuration is thus identical to hydrogen.

Let us discuss the characteristics of alkali spectra taking sodium as a typical example of an alkali element. The optical  $3s$  electron of sodium when excited jumps to higher energy states, such as  $3P$ ,  $4S$ ,  $3D$ ,  $4S$ ,  $5S$ ,  $4D$  ... etc. From the higher level the electron jumps back to the lower levels but only such that the  $l$  value changes by  $\pm 1$  i.e.,

$$\Delta l = \pm 1.$$

This is called the '*selection rule*'.

Fig. 15.3 gives the energy level diagram of the sodium atom and shows the allowed transitions of the optical electron.



**Fig. 15.3**

- (i) When the electron jumps from any  $p$ -level to the lowest  $s$ -level ( $3s$ ), it emits a line of principal series.
- (ii) When the electron jumps from any  $s$ -level to the lowest  $p$ -level ( $3p$ ), it emits a line of sharp series.
- (iii) When the electron jumps from any  $d$ -level to the lowest  $p$ -level ( $3p$ ), it emits a line of diffuse series.
- (iv) When the electron jumps from any  $f$ -level to the lowest  $d$ -level ( $3d$ ), it emits a line of fundamental series.

Thus emission of all the spectral series is explained.

The different series in the spectrum arise from the following transitions of the optical electron.

Principal series	$n \quad ^2P \rightarrow 3^2S, \quad n = 3, 4, 5\dots$
Sharp series	$n \quad ^2S \rightarrow 3^2P, \quad n = 4, 5, 6\dots$
Diffuse series	$n \quad ^2D \rightarrow 3^2P, \quad n = 3, 4, 5\dots$
Fundamental series	$n \quad ^2F \rightarrow 3^2D, \quad n = 4, 5, 6\dots$

For a given set of values of the principal quantum number, we get several series here, unlike in the case of hydrogen.

According to Bohr-Sommerfeld theory and quantum mechanical theory, all substates belonging to a given  $n$  in hydrogen have same value of energy. But this  $l$  degeneracy is removed in a multi-electron system because of shield of nuclear charge and penetration of atomic core. The term value is written as

$$T = \frac{RZ_0^2}{n_{\text{eff}}^2}$$

Here,  $Z_0$  is effective nuclear charge outside the core.  $n_{\text{eff}}^2$  is effective quantum number. As  $Z_0$  is greater than one and  $n_{\text{eff}} < n$ , the term value is increased. As a result, energy levels in alkali atoms lie lower than the corresponding hydrogen levels. However, for large  $n$ ,  $Z_0$  is equal to unity and  $n_{\text{eff}} \rightarrow n$ . The corresponding levels approach the hydrogen levels.

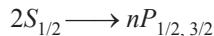
## 15.5 THE SPECTRUM OF LITHIUM AND OTHER HYDROGEN-LIKE SPECIES

The energy levels of the lithium atom are sketched in Fig. 15.4. It shows the difference in energy of  $s$ ,  $p$ , and  $d$  states with the same value of  $n$ . Some allowed transitions are also shown.

**Selection Rules :** The selection rules for alkali metals are the same as for hydrogen.

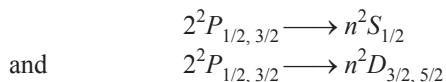
$$\Delta n = \text{anything}, \Delta l = \pm 1, \Delta j = 0, \pm 1$$

- (i) The transitions from the ground state ( $1s^2 2s$ ) can occur to  $p$  levels:



A series of doublets similar to the Lyman series will be formed.

- (ii) From the  $2p$  state, two separate series of lines will be seen:



The former will be doublets, the latter compound doublets, but their frequencies will differ because the  $s$  and  $d$  orbital energies are no longer the same.

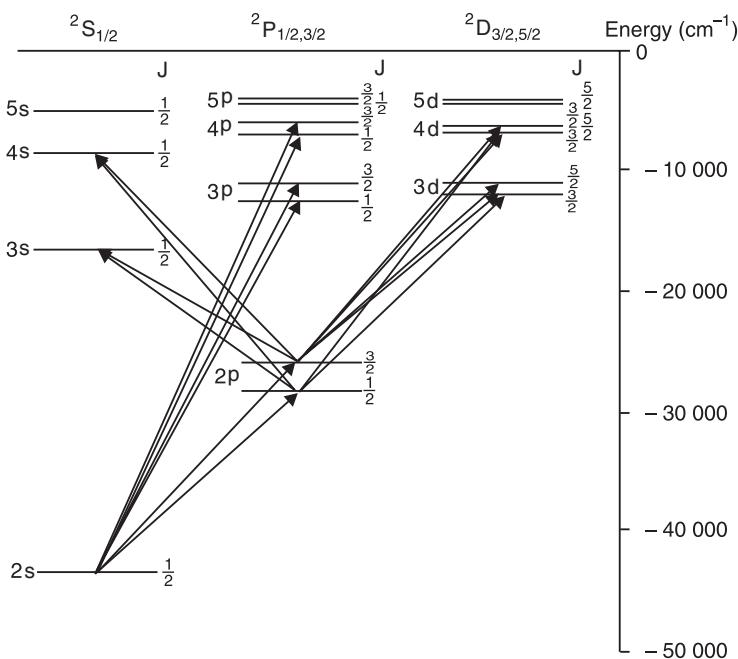


Fig. 15.4

The same remarks apply to the other alkali metals, the differences between their spectra and that of lithium being a matter of scale only. For instance the *j*-splitting due to coupling between *l* and *s* increases markedly with the atomic number. Thus the doublet separation of lines in the spectral series, which is scarcely observable for hydrogen, is less than 1 cm<sup>-1</sup> for the 2*p* level of lithium, about 17 cm<sup>-1</sup> for sodium, and over 5000 cm<sup>-1</sup> for cesium.

Any atom which has a single electron moving outside a closed shell will exhibit a spectrum of the type discussed above. Thus ions of the type He<sup>+</sup>, Be<sup>+</sup>, B<sup>2+</sup>, etc., show 'hydrogen-like spectra'.

## 15.6 THE SPECTRUM OF HELIUM AND THE ALKALINE EARTHS

Helium consists of a central nucleus and two outer electrons. There are only two possibilities for the relative spins of the two electrons:

### Case 1. Singlet states

The spins of the two electrons are paired ( $\uparrow\downarrow$ ). If  $S_{1z}$  is  $+\frac{1}{2}$ ,  $S_{2z}$  must be  $-\frac{1}{2}$ .

Hence  $S_z = S_{1z} + S_{2z} = 0$ . So  $S = 0$  and we have *singlet* states.

The lowest possible energy state of this atom is when both electrons occupy the 1*s* orbital. By Pauli's principle, this is possible only if their spins are paired. So the ground state of helium must be a singlet state. Further,  $L = l_1 + l_2 = 0$ , and hence  $J$  can only be zero. The ground state of helium, therefore, is  $^1S_0$ .

The selection rules for many-electron systems are:

$$\Delta S = 0 \quad \Delta L = \pm 1 \quad \Delta J = 0, \pm 1 \quad \dots(1)$$

Since  $S$  cannot change during a transition, the singlet ground state can undergo transitions only to other singlet states.

We assume that only one electron undergoes transitions, leaving the other in the 1*s* orbital.

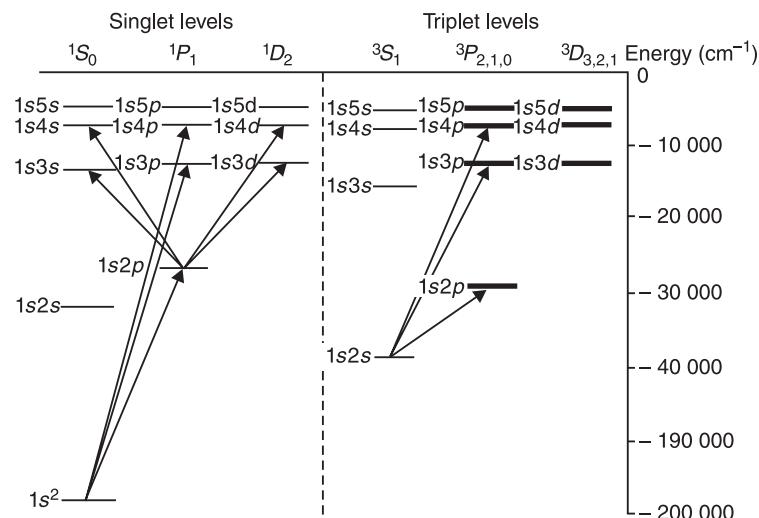


Fig. 15.5

The left-hand side of Fig. 15.5 shows the energy levels of the electrons and the allowed transitions.

- (i) Initially the  $1s^2 \ ^1S_0$  state can undergo a transition only to  $1s^1 \ np^1$  states (abbreviated to  $1snp$ ). In the  $1snp$  state  $L = 1$ ,  $S = 0$ , and hence  $J = 1$  only.

So the transition may be symbolized :

$$\begin{aligned} 1s^2 \ ^1S_0 &\longrightarrow 1snp \ ^1P_1 \\ \text{or, briefly: } & \quad ^1S_0 \longrightarrow ^1P_1 \end{aligned}$$

- (ii) From the  $^1P_1$  state the system could either revert to  $^1S_0$  states, or undergo transitions to the higher  $^1D_2$  states (for these  $S = 0$ ,  $L = 2$  and hence  $J = 2$  only).

In general, then, all these transitions will give rise to spectral series very similar to those of lithium except that here transitions are between *singlet* states only and all the spectral lines will be single.

### Case 2. Triplet states

The electron spins are parallel ( $\uparrow\uparrow$ ). Now  $S_{1z} = S_{2z} = +\frac{1}{2}$ . So  $S = 1$  and the states are *triplet*.

Since the electrons are now forbidden by Pauli's principle from occupying the same orbital, the lowest energy state is  $1s \ 2s$ . This and other triplet energy levels are shown on the right of Fig. 15.5. The  $1s \ 2s$  state has  $S = 1$ ,  $L = 0$ , and hence  $J = 1$  only, and so it is  $^3S_1$ . By the selection rules of Eq. (1) it can only undergo transitions into the  $1snp$  triplet states. These, with  $S = 1$ ,  $L = 1$ , have  $J = 2, 1$ , or  $0$ . So, the transitions may be written:

$$^3S_1 \rightarrow ^3P_2, ^3S_1 \rightarrow ^3P_1 \text{ and } ^3S_1 \rightarrow ^3P_0$$

All three transitions are allowed, since  $\Delta J = 0$  or  $\pm 1$ , so the resulting spectral lines will be *triplets*.

### Transitions between $^3P$ and $^3D$ Levels

Transitions from the  $^3P$  states may take place to  $^3D$  states. For  $^3D$  we have  $S = 1$ ,  $L = 2$ , and hence  $J = 3, 2$ , or  $1$ . Fig. 15.6, shows the 'compound triplet', spectrum arising from transitions between  $^3P$  and  $^3D$  states, bearing in mind the selection rule  $\Delta J = 0, \pm 1$ .

- (i)  $^3P_2$  can go to each of  $^3D_{3,2,1}$ ,
- (ii)  $^3P_1$  can go only to  $^3D_{2,1}$ , and
- (iii)  $^3P_0$  can go only to  $^3D_1$

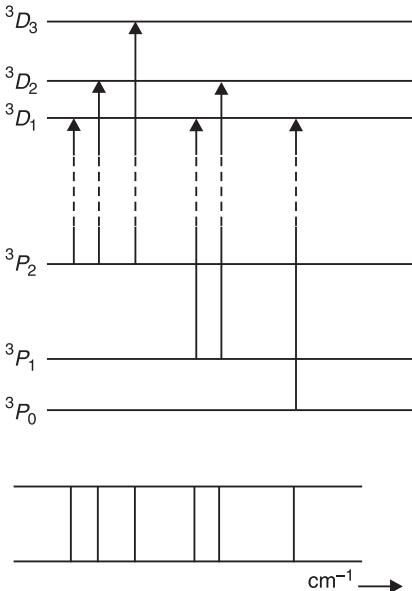
Thus the complete spectrum should consist of six lines. Normally, however, the very close spacing is not resolved, and only three lines are seen. For this reason, the spectrum is referred to as a *compound triplet*.

We see, then, that the spectrum of helium consists of spectral series grouped into two types which overlap each other in frequency.

- (i) In one type, involving transitions between singlet levels, all the spectral lines are themselves singlets.
- (ii) In the other type, the transitions are between triplet states and each 'line' is at least a close triplet and possibly even more complex.

Because of the selection rule  $\Delta S = 0$  there is a strong prohibition on transitions between singlet and triplet states, and transitions *cannot occur* between the right- and left-hand sides of Fig. 15.5.

Other atoms containing two outer electrons exhibit spectra similar to that of helium. Thus the alkaline earths, beryllium, magnesium, calcium, etc., fall into this category, as do ionized species with just two remaining electrons, for example,  $B^+$ ,  $C^{2+}$ , etc.



**Fig. 15.6**

### EXERCISE

1. Describe the general features of the spectra of alkali-like atoms. How are they explained? In what respect do they differ from the spectra of  $H_2$ ? *(Purvanchal University, 90, 91)*
2. What are penetrating and non-penetrating orbits ? *(P.U. 91)*
3. Describe the broad features of the spectrum of helium atom and show how the observations have been accounted for.
4. Discuss with necessary theory the origin of singlet and triplet terms in the spectra of elements with two valence electrons. Why do the triplet terms not arise in the ground state configuration of alkaline earths? *(P.U. 91, 94)*

## X-RAY SPECTRA

**After reading this chapter, you should be able to**

- ◆ Compare Optical and X-ray spectra.
- ◆ Explain why discrete lines are observed in optical atomic absorption whereas X-ray absorption shows continua with edges.
- ◆ Discuss X-ray Absorption Spectrum

## 16.1 FINE STRUCTURE OF X-RAY LEVELS

The resolution of  $L$  level into 3 sublevels, of  $M$  level into 5 sublevels and of  $N$  level into 7 sub-levels is called *fine structure* of X-ray levels. To account for the fine structure of X-ray levels, we consider the effect of electrons which screen the nuclear charge and the effect of spin of electrons.

The  $L$  electrons move in an approximately Coulomb field. The two  $K$  electrons act very much as if they had coalesced with the nucleus. This screening of the nucleus by the  $K$  electrons reduces the nuclear charge by 2.

This reducing of nuclear charge by the electrons which are nearer to the nucleus is called screening effect. In a similar way, for an electron in  $M$  shell,  $K$  and  $L$  electrons screen the nuclear charge.

Therefore, the energy of the system is written as

$$T = \frac{R(Z-b)^2}{n^2}.$$

Here,  $b$  is screening constant.

$(Z-b)$  may be called the effective atomic number of the element.

It is to be noted that the constant  $b$  also includes the effect of outer electrons on the inner ones. This effect is known as external screening. Further, when an electron moves in an orbit of low quantum number in the field of the nucleus of high atomic number, the velocity of the electron is high enough to take into account the relativistic variation of mass. With this argument, Sommerfeld and Wentzel showed that the energy of the system is calculated by

$$\begin{aligned} T = & \frac{R(Z-b)^2}{n^2} + \frac{R\alpha^2}{n^4}(Z-d)^4 \left( \frac{n}{k} - \frac{3}{4} \right) \\ & + \frac{R\alpha^4}{n^6}(Z-d)^6 \left( \frac{1}{4} \frac{n^3}{k^3} + \frac{3n^2}{4k^2} - \frac{3n}{2k} + \frac{5}{8} \right) \\ & + \frac{R\alpha^6}{n^8}(Z-d)^8 \left\{ \frac{1}{8} \frac{n^5}{k^5} + \frac{3}{8} \frac{n^4}{k^4} + \frac{1}{8} \frac{n^3}{k^3} - \frac{15}{8} \left( \frac{n}{k} \right)^2 + \frac{15}{8} \frac{n}{k} - \frac{35}{64} \right\}. \end{aligned}$$

Here,  $\alpha^2 = 4\pi^2 e^4/h^2c^2 = 5.30 \times 10^{-5}$  is the square of Sommerfeld's fine-structure constant,  $R$  is the Rydberg constant,  $b$  and  $d$  are screening constants,  $n$  is the total quantum number, and  $k$  is Sommerfeld's original azimuthal quantum number;  $k = 1, 2, 3, \dots$  for  $s, p, d, \dots$  electrons, respectively.

The first screening constant  $b$  depends on the distribution of electrons in shells both outside and inside the one containing the electron considered.

The second screening constant  $d$  is due to spin-orbit interaction and relativity.

The effect of two screening constants  $b$  and  $d$  in shifting the energy level from those of the hydrogen-like atom is illustrated in Fig. 16.1.

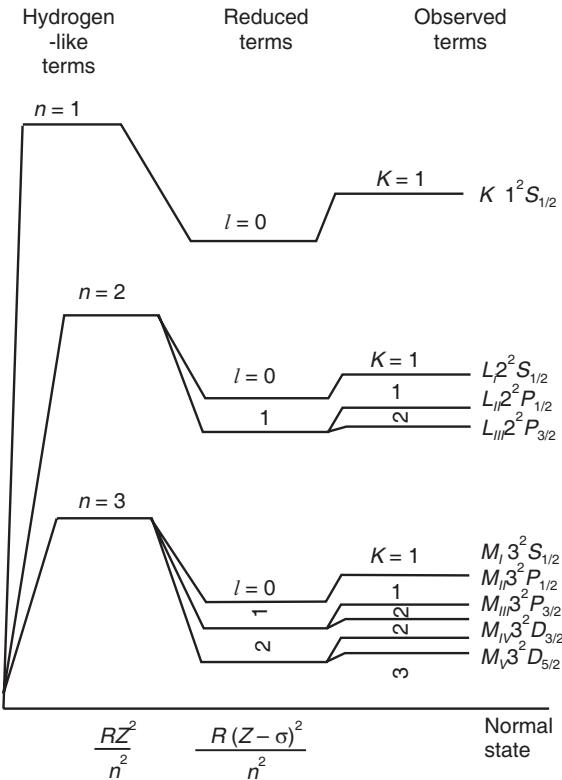


Fig. 16.1

The effect of screening constant  $b$  is to break one level into  $n$  sub-levels and the corresponding term is known as reduced term. If now to this effect, we add the effect of second screening constant  $d$ , each sub-level again breaks up into two with the exception of ground level.

## 16.2 SPIN-RELATIVITY DOUBLETS (REGULAR DOUBLETS)

The term *spin-relativity doublet* refers in general to each pair of energy levels having the same  $n$ ,  $S$ , and  $L$  values but different  $J$  values. In x-ray notation, these are the intervals  $L_{II} - L_{III}$ ,  $M_{II} - M_{III}$ ,  $M_{IV} - M_{V}$ ,  $N_{II} - N_{III}$ ,  $N_{IV} - N_{V}$ ,  $N_{VI} - N_{VII}$ ,  $O_{II} - O_{III}$ ,  $O_{IV} - O_{V}$ , etc.

The term values for  $L_{II}$  and  $L_{III}$  are

$$T_{L_{II}} = \frac{R(Z-d)^2}{2^2} + \frac{R\alpha^2}{2^4}(Z-d)^4 \\ \left\{ \frac{5}{4} + \frac{21}{8} \frac{\alpha^2}{2^2}(Z-d)^2 + \frac{429}{64} \frac{\alpha^4}{2^4}(Z-d)^4 + \dots \right\}$$

$$T_{L_{III}} = \frac{R(Z-d)^2}{2^2} + \frac{R\alpha^2(Z-d)^4}{2^4} \\ \left\{ \frac{1}{4} + \frac{1}{8} \frac{\alpha^2}{2^2}(Z-d)^2 + \frac{5}{64} \frac{\alpha^4}{2^4}(Z-d)^4 + \dots \right\}.$$

$$\Delta\nu = T_{L_{II}} - T_{L_{III}} = \frac{R\alpha^2}{2^4}(Z-d)^4 \left\{ 1 + \frac{5\alpha^2}{8}(Z-d)^2 + \frac{53\alpha^4}{128}(Z-d)^4 + \dots \right\}.$$

The value of  $d$  is constant for a particular value of  $L$  and it is independent of  $Z$ . For example, for  $L_{\text{II}} - L_{\text{III}}$ , the value of  $d$  is 3.50.

Neglecting higher terms,

$$\Delta v = \frac{R\alpha^2(Z-d)^4}{2^4}.$$

#### 16.2.1. The Regular-doublet Law

The regular-doublet law refers to the law governing the spin-relativity doublets  $L_{\text{II}} - L_{\text{III}}$ ,  $M_{\text{II}} - M_{\text{III}}$ ,  $M_{\text{IV}} - M_{\text{V}}$ ,  $N_{\text{II}} - N_{\text{III}}$ , etc.

**Statement.** *The doublet separation,  $\Delta v$  in wave numbers, for any spin-relativity doublet is approximately proportional to the fourth power of the effective atomic number,  $(Z-d)$ .*

For  $L_{\text{II}} - L_{\text{III}}$ , we have,

$$\Delta v = \frac{R\alpha^2}{2^4}(Z-3.5)^4.$$

#### 16.2.2. Screening Doublets

The doublet intervals  $(L_{\text{I}} - L_{\text{II}})$ ,  $(M_{\text{I}} - M_{\text{II}})$ ,  $(M_{\text{III}} - M_{\text{IV}})$ ,  $(N_{\text{I}} - N_{\text{II}})$ ,  $(N_{\text{III}} - N_{\text{IV}})$ ... etc., are called screening doublets. These are the levels having same  $n$ ,  $S$ , and  $J$  values but different  $L$  values.

Taking only the effect of screening into account, we write

$$T = \frac{R(Z-b)^2}{n^2}$$

or 
$$\left(\frac{T}{R}\right)^{1/2} = \frac{Z-b}{n}.$$

Hence, we write the value  $\left(\frac{T}{R}\right)^{1/2}$  for  $L_{\text{I}}$  and  $L_{\text{II}}$  as

$$\left(\frac{T}{R}\right)_{L_1}^{1/2} = \frac{Z-b_1}{n}$$

$$\left(\frac{T}{R}\right)_{L_2}^{1/2} = \frac{Z-b_2}{n}$$

$$\therefore \Delta \sqrt{\left(\frac{T}{R}\right)} = \frac{b_2 - b_1}{n} = \frac{\Delta b}{n} = \text{constant.}$$

#### 16.2.3. Irregular-doublet Law

**Statement.** *The difference between the square roots of the term values of a given doublet is a constant independent of the atomic number  $Z$ .*

### 16.3 X-RAY ABSORPTION SPECTRUM

A beam of continuous X-rays is passed through the film of a substance. The transmitted beam is examined by an X-ray spectrograph. Then, the X-ray absorption spectrum of that substance is obtained. The nature of the X-ray absorption spectrum is quite different from the X-ray emission spectrum of the same substance.

- (i) The emission lines  $K_{\alpha}, K_{\beta}, K_{\gamma}, \dots$  do not appear in the absorption spectrum.
- (ii) The absorption spectrum shows a continuous region bounded by a sharp edge (known as absorption edge) in the position of the limit of the  $K$ -series.
- (iii) The absorption edges are characteristic of the absorber element. When successive elements in the periodic table are used as absorbers, the edge shifts toward higher frequency (Fig. 16.2). Thus these edges are characteristic of the absorber element and are called 'absorption edges'.

**(i) Explanation of Absence of Absorption lines.** The  $K_{\alpha}, K_{\beta}, \dots$  emission lines cannot be observed as absorption lines. For example,  $K_{\alpha}$  line in emission occurs when an electron drops from the  $L$ -shell to the  $K$ -shell. The line can occur in absorption if electron from  $K$ -shell may jump to  $L$ -shell after absorbing necessary energy from continuous beam. But this does not happen because the  $L$ -shell has already its full quota of 8 electrons and so it cannot receive any further electron (Pauli's exclusion principle). Hence the  $K_{\alpha}$ -line cannot be observed as absorption line. This is true for all the  $X$ -ray emission lines that are commonly observed.

**(ii) Explanation of Appearance of Absorption Edges.** An  $X$ -ray photon, capable of removing an inner electron to infinity (where no shells exist), can be absorbed. Let  $E_K$  be the energy required to remove a  $K$ -electron. Then, a photon of frequency  $\nu \geq \nu_K$  can eject a  $K$ -electron where

$$h\nu_K = E_K.$$

Photons of frequency less than  $\nu_K$  are transmitted by the absorbing element.

At the frequency  $\nu_K$ , absorption in  $K$ -shell suddenly starts and continues for all frequencies greater than  $\nu_K$ . Therefore, absorption spectrum is abruptly discontinued at  $\nu_K$  and photographic plate is not blackened beyond  $\nu_K$ .

As  $Z$  increases,  $E_K$  also increases. Hence  $\nu_K$  also increases.

**Multiplicity of Atomic Shells.** When the  $X$ -ray absorption spectrum is taken in the  $L$ -region, again a continuum is observed but now with three absorption edges. This means that the  $L$ -shell has three energy levels from which electrons can be ejected to infinity. Similarly,  $M$ -shell has five energy levels, and so on. Thus  $X$ -ray absorption spectrum reveals the multiplicity of atomic shells.

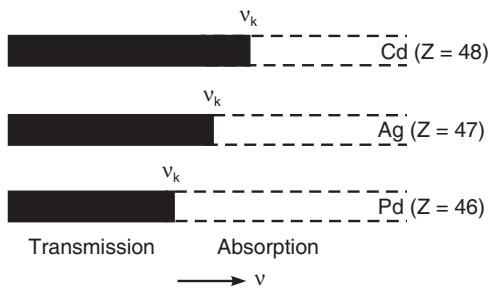


Fig. 16.2

### EXERCISE

1. Analyse the fine structures of  $X$ -ray energy levels and give an account of screening and spin relativity doublets.
2. Discuss  $X$ -ray absorption spectrum and the importance of sharp edges. (Gorakhpur 87, Kanpur 85)

## PART - V

# NUCLEAR PHYSICS

**17. Introduction to the Nucleus**

**18. Detectors of Nuclear Radiations**

**19. Particle Accelerators**

**20. Radioactivity**

**21. Artificial Transmutation of Elements**

**22. Nuclear Fission and Fusion**

**23. Cosmic Rays**

**24. Elementary Particles**

## INTRODUCTION TO THE NUCLEUS

**After reading this chapter, you should be able to**

- ◆ Identify properties of the nucleus and other sub-atomic particles
- ◆ Explain the theory and the experimental observations related to the nucleus.
- ◆ Demonstrate an understanding of the factors affecting the stability of the nucleus
- ◆ Discuss the problem of the constitution of the nucleus
- ◆ Demonstrate an understanding of nuclear forces, interactions and models
- ◆ Name the methods to find the nuclear radius and interpret experimental data
- ◆ Compare various nuclear models
- ◆ Explain why stable nuclei never have more protons than neutrons.

## 17.1 INTRODUCTION

The atomic nucleus was discovered in 1911 by Rutherford. Rutherford's  $\alpha$ -particle scattering experiments showed that the atom consists of a very small nucleus ( $\approx 10^{-14}$  m in diameter) surrounded by orbiting electrons. It is the purpose of this book to explain all aspects of the nucleus, its structure and its behaviour under various conditions. In this chapter we introduce some of its most basic characteristics, its mass, size, shape, and other externally observable properties. We also consider some deeper questions, such as the force that holds the nucleus together.

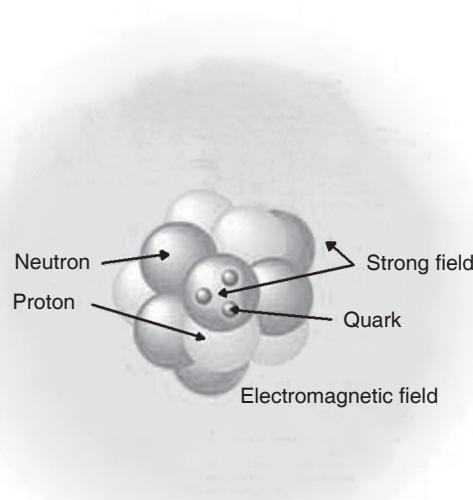
Let us begin by reviewing a few fundamental facts that are probably already familiar. All atomic nuclei are made up of elementary particles called *protons* and *neutrons*. A proton has a positive charge of the same magnitude as that of an electron.

A neutron is electrically neutral. The proton and the neutron are considered to be two different charge states of the same particle which is called a *nucleon*.

A species of nucleus, known as a *nuclide*, is represented schematically by  ${}_Z^A X$  where  $Z$ , the *atomic number* indicates the number of protons,  $A$ , the *mass number*, indicates the total number of protons plus neutrons and  $X$  is the chemical symbol of the species.

$$N = \text{Number of neutrons} = A - Z.$$

As an example, the chlorine nucleus  ${}_{17}^{35}\text{Cl}$  has  $Z = 17$  protons,  $A = 35$  nucleons and  $N = 35 - 17 = 18$  neutrons.



*The Nucleus*

## 17.2 CLASSIFICATION OF NUCLEI

Atoms of different elements are classified as follows :

(i) *Isotopes* are nuclei with the same atomic number  $Z$  but different mass numbers  $A$ . The nuclei  ${}_{14}^{28}\text{Si}$ ,  ${}_{14}^{29}\text{Si}$ ,  ${}_{14}^{30}\text{Si}$ , and  ${}_{14}^{32}\text{Si}$  are all isotopes of silicon. The isotopes of an element all contain the same number of protons but have different number of neutrons. Since the nuclear charge is what is ultimately responsible for the characteristic properties of an atom, all the isotopes of an element have identical chemical behaviour and differ physically only in mass.

(ii) Those nuclei, with the same mass number  $A$ , but different atomic number  $Z$ , are called *isobars*. The nuclei  ${}_8^{16}\text{O}$  and  ${}_{17}^{35}\text{Cl}$  are examples of isobars. The isobars are atoms of different elements and have different physical and chemical properties.

(iii) Nuclei, with an equal number of neutrons, that is, with the same  $N$ , are called *isotones*. Some isotones are  ${}_6^{14}\text{C}$ ,  ${}_{15}^{30}\text{N}$  and  ${}_8^{16}\text{O}$  ( $N = 8$  in each case).

(iv) There are atoms, which have the same  $Z$  and same  $A$ , but differ from one another in their nuclear energy states and exhibit differences in their internal structure. These nuclei are distinguished by their different life times. Such nuclei are called *isomeric nuclei* or *isomers*.

(v) Nuclei, having the same mass number  $A$ , but with the proton and neutron number interchanged (that is, the number of protons in one is equal to the number of neutrons in the other) are called *mirror nuclei*.

**EXAMPLE.**  ${}_4Be^7$  ( $Z = 4$  and  $N = 3$ ) and  ${}_3Li^7$  ( $Z = 3$  and  $N = 4$ ).

### Notes

Nuclei, consisting of an even (odd) number of protons and even (odd) number of neutrons, are said to be even-even (odd-odd). Nuclei consisting of an even (odd) number of protons and an odd (even) number of neutrons are said to be even-odd (odd-even).

## 17.3 GENERAL PROPERTIES OF NUCLEUS

**Nuclear size.** Rutherford's work on the scattering of  $\alpha$ -particles showed that the mean radius of an atomic nucleus is of the order of  $10^{-14}$  to  $10^{-15}$  m while that of the atom is about  $10^{-10}$  m. Thus the nucleus is about 10000 times smaller in radius than the atom.

The empirical formula for the nuclear radius is

$$R = r_0 A^{1/3}$$

where  $A$  is the mass number and  $r_0 = 1.3 \times 10^{-15}$  m =  $1.3\text{ fm}$ . Nuclei are so small that the fermi ( $\text{fm}$ ) is an appropriate unit of length.  $1\text{ fm} = 10^{-15}$  m. From this formula we find that the radius of the  ${}_6C^{12}$  nucleus is  $R \approx (1.3)(12)^{1/3} = 3\text{ fm}$ . Similarly, the radius of the  ${}_47Ag^{107}$  nucleus is  $6.2\text{ fm}$  and that of the  ${}_92U^{238}$  nucleus is  $8.1\text{ fm}$ .

The nuclear radius may be estimated from the scattering of neutrons and electrons by the nucleus, or by analysing the effect of the finite size of the nucleus on nuclear and atomic binding energies.

Fast neutrons of about 100 MeV energy, whose wavelength is small compared to the size of the nucleus, are scattered by nuclear targets. The fraction of neutrons scattered at various angles can be used to deduce the nuclear size. The results of these experiments indicate that the radius of a nucleus is given by  $R \approx r_0 A^{1/3}$  where  $r_0 \approx 1.3 - 1.4\text{ fm}$ . The scattering can be done with proton beams as well. In this case, however, the effects due to Coulomb interaction have to be separated out. The observations are in agreement with the equation  $R \approx r_0 A^{1/3}$  with  $r_0 \approx 1.3 - 1.4\text{ fm}$ .

The scattering of fast electrons of energy as high as  $10^4$  MeV, with a wavelength of about  $0.1\text{ fm}$ , has the advantage that it can directly measure the charge density inside a nucleus. The results of the experiment are in agreement with the equation  $R \approx r_0 A^{1/3}$  but with a somewhat smaller value of  $r_0 \approx 1.2\text{ fm}$ . The slight difference in the value of  $r_0$  may be ascribed to the fact that the electron scattering measures the charge density whereas the neutron and proton scattering experiments measure the region of large nuclear potential, which may be expected to be somewhat larger than the size of the nucleus.

**EXAMPLE.** The radius of  $Ho^{165}$  is  $7.731\text{ fm}$ . Deduce the radius of  $He^4$ .

**SOL.** Let  $R_1, A_1$  and  $R_2, A_2$  be the radius and mass number of  $Ho^{165}$  and  $He^4$  respectively. Then  $R_1 = r_0 A_1^{1/3}$  and  $R_2 = r_0 A_2^{1/3}$ .

$$\therefore \frac{R_1}{R_2} = \frac{A_1^{1/3}}{A_2^{1/3}} \text{ or}$$

$$R_2 = \frac{R_1 A_2^{1/3}}{A_1^{1/3}} = \frac{7.731 \times 4^{1/3}}{(165)^{1/3}} = 2.238\text{ fm}.$$

**Nuclear mass.** We know that the nucleus consists of protons and neutrons. Then the mass of the nucleus should be

$$\text{assumed nuclear mass} = Zm_p + Nm_n$$

Here,  $m_p$  and  $m_n$  are the respective proton and neutron masses and  $N$  is the neutron number. Nuclear masses are experimentally measured accurately by mass spectrometers. Measurements by mass spectrometer, however, show that

$$\text{real nuclear mass} < Zm_p + Nm_n$$

The difference in masses

$$Zm_p + Nm_n - \text{real nuclear mass} = \Delta m$$

is called the *mass defect*.

**Nuclear density.** The nuclear density  $\rho_N$  can be calculated from  $\rho_N = \frac{\text{Nuclear mass}}{\text{Nuclear volume}}$ .

Nuclear mass =  $Am_N$  where  $A$  = mass number and  $m_N$  = mass of the nucleon =  $1.67 \times 10^{-27}$  kg.

$$\text{Nuclear volume} = \frac{4}{3} \pi R^3 = \frac{4}{3} \pi (r_0 A^{1/3})^3 = \frac{4}{3} \pi r_0^3 A$$

$$\begin{aligned} \therefore \rho_N &= \frac{Am_N}{\frac{4}{3} \pi r_0^3 A} = \frac{m_N}{\frac{4}{3} \pi r_0^3} = \frac{(1.67 \times 10^{-27})}{\frac{4}{3} \pi (1.3 \times 10^{-15})^3} \\ &= 1.816 \times 10^{17} \text{ kg m}^{-3}. \end{aligned}$$

Note the high value of the density of the nucleus. This shows that the nuclear matter is in an extremely compressed state. Certain stars (the “white dwarfs”) are composed of atoms whose electron shells have collapsed owing to enormous pressure, and the densities of such stars approach that of pure nuclear matter.

**Nuclear charge.** The charge of the nucleus is due to the protons contained in it. Each proton has a positive charge of  $1.6 \times 10^{-19}$  C. The nuclear charge is  $Ze$  where  $Z$  is the atomic number of the nucleus. The value of  $Z$  is known from X-ray scattering experiments, from the nuclear scattering of  $\alpha$ -particles, and from the X-ray spectrum.

**Spin angular momentum.** Both the proton and neutron, like the electron, have an intrinsic spin. The *spin angular momentum* is computed by  $L_s = \sqrt{l(l+1)} \hbar$ . Here the quantum number  $l$ , commonly called the spin, is equal to  $1/2$ . The spin angular momentum, then has a value  $L_s = \frac{\sqrt{3}}{2} \hbar$

**Resultant angular momentum.** In addition to the spin angular momentum, the protons and neutrons in the nucleus have an *orbital angular momentum*. The resultant angular momentum of the nucleus is obtained by adding the spin and orbital angular momenta of all the nucleons within the nucleus. The total angular momentum of a nucleus is given by  $L_N = \sqrt{l_N(l_N+1)} \hbar$ . This total angular momentum is called *nuclear spin*.

**Nuclear magnetic dipole moments.** We know that the spinning electron has an associated magnetic dipole moment of 1 Bohr magneton. i.e.,  $\mu_e = \frac{e\hbar}{2m_e}$ . Proton has a positive elementary

charge and due to its spin, it should have a magnetic dipole moment. According to Dirac's theory,  $\mu_N = \frac{e\hbar}{2m_p}$ . Here,  $m_p$  is the proton mass.

- $\mu_N$  is called a *nuclear magneton* and is the unit of nuclear magnetic moment.

- $\mu_N$  has a value of  $5.050 \times 10^{-27}$  J/T. Since  $m_p = 1836 m_e$ , the nuclear magneton is only 1/1836 of a Bohr magneton. For nucleons, however, measurements give  $\mu_p = 2.7925 \mu_N$  and  $\mu_n = -1.9128 \mu_N$ . Physicists have found that it is especially hard to understand how the neutron, which is a neutral particle, can have a magnetic moment.

**Electric quadrupole moment.** In addition to its magnetic moment, a nucleus may have an electric quadrupole moment. An electric dipole moment is zero for atoms and nuclei in stationary states. This is a consequence of the symmetry of nuclei about the centre of mass. However, this symmetry does not need to be spherical; there is nothing precluding the nucleus from assuming the shape of an ellipsoid of rotation, for instance. Indeed most nuclei do assume approximately such a shape, and the deviation from spherical symmetry is expressed by a quantity called the *electric quadrupole moment*. It is defined as

$$Q = \left(\frac{1}{e}\right) \int (3z^2 - r^2) \rho \, d\tau$$

Here,  $\rho$  is the charge density in the nucleus.

- $Q$  is actually a measure of the eccentricity of the ellipsoidal nuclear surface. Evidently  $Q = 0$  for a spherically symmetric charge distribution. A charge distribution stretched in the z-direction (prolate) will give a positive quadrupole moment, and an oblate distribution will give a negative quadrupole moment (Fig. 17.1). Since the expression is divided by the electronic charge, the dimension of the quadrupole moment is that of an area. In nuclear physics, area is measured in barns. (1 barn =  $10^{-28} \text{ m}^2$ ).

### Parity of Nuclei

The total spin of a nucleus consists of the sum of orbital angular momentum of nucleons and the sum of their spins. The orbital angular momentum  $\sum_n l_n$  actually defines the parity of the nuclei. The angular momentum eigen function can be expressed as a Spherical Harmonics  $Y_m$  which is an even function for even  $I_n$  and an odd function for an odd  $I_n$ . This shows that  $\sum_n l_n$  is either even or odd for nuclei. According to this, the nuclear wave functions are said to have either an even parity or odd parity.

For positive or even parity,  $\psi(x, y, z) = \psi(-x, -y, -z)$ .

For negative or odd parity,  $\psi(x, y, z) = -\psi(-x, -y, -z)$ .

If  $\sum_n l_n$  is even the parity of nuclide is positive and if  $\sum_n l_n$  is odd, the parity is negative.

As an example, the Deuteron nucleus contains a neutron and a proton in *S*-state with  $l = 0$  and the parity of deuteron is positive.

Nuclei of various atoms in a ground state have a definite parity which is either positive or negative. When the nuclei are in the excited state, their parities are not always the same as in the ground state. Parity in nuclear transformations is conserved quantity but is not conserved in weak interactions like Beta decay. It is conserved in nuclear reactions and gamma decay.

### Isospin quantum number ( $T$ )

Neutrons and protons are similar in all respects except charge. We may regard a *nucleon* as a single entity having two states, the *proton* and the *neutron*. To describe their quantum state, quantum number used is *isospin quantum number ( $T$ )*. A nucleon is assigned an isospin of  $1/2$ . In an electromagnetic field, two charge states with isospin components of  $1/2$  and  $-1/2$  can be

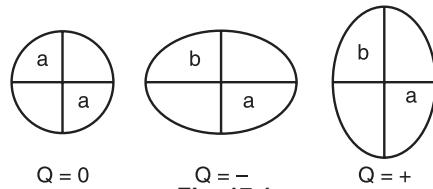


Fig. 17.1

distinguished as the proton and neutron respectively. The isospin is a vector in a three dimensional space, called isospin space. The component that is measured is said to be  $T_3$ , or  $T_z$ , the component of  $T$  in the direction of the third axis in isospin space.

### 17.4 BINDING ENERGY

The theoretical explanation for the mass defect is based on Einstein's equation  $E = mc^2$ . When the  $Z$  protons and  $N$  neutrons combine to make a nucleus, some of the mass ( $\Delta m$ ) disappears because it is converted into an amount of energy  $\Delta E = (\Delta m) c^2$ . This energy is called the *Binding energy (B.E.)* of the nucleus. To disrupt a stable nucleus into its constituent protons and neutrons, the minimum energy required is the binding energy. The magnitude of the B.E. of a nucleus determines its *stability against disintegration*. If the B.E. is large, the nucleus is stable. A nucleus having the least possible energy, equal to the B.E., is said to be in the *ground state*. If the nucleus has an energy  $E > E_{\min}$ , it is said to be in the *excited state*. The case  $E = 0$  corresponds to dissociation of the nucleus into its constituent nucleons.

If  $M$  is the experimentally determined mass of a nuclide having  $Z$  protons and  $N$  neutrons,

$$\text{B.E.} = \{(Zm_p + Nm_n) - M\}c^2.$$

If B.E.  $> 0$ , the nucleus is stable and energy must be supplied from outside to disrupt it into its constituents. If B.E.  $< 0$ , the nucleus is unstable and it will disintegrate by itself.

**Illustration:** Let us illustrate the calculation of B.E. by taking the example of the deuteron. A deuteron is formed by a proton and a neutron.

$$\text{Mass of proton} = 1.007276u.$$

$$\text{Mass of neutron} = 1.008665u.$$

$$\therefore \text{Mass of proton + neutron in free state} = 2.015941u$$

$$\text{Mass of deuteron nucleus} = 2.013553u$$

$$\therefore \text{Mass defect} = \Delta m = 0.002388u$$

$$\text{B.E.} = 0.002388 \times 931 = 2.23 \text{ MeV } (\because 1u = 931 \text{ MeV})$$

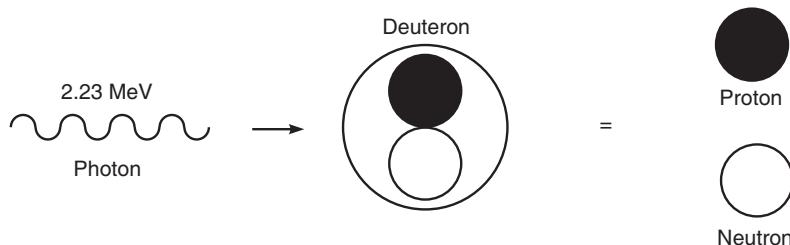
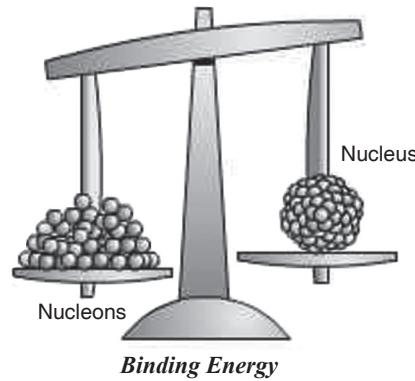


Fig. 17.2

Thus, when a deuteron is formed from a free proton and neutron, 2.23 MeV of energy is liberated. Conversely, 2.23 MeV must be supplied from an external source to break a deuteron up into a proton and a neutron. This is confirmed by experiments that show that a gamma-ray photon with a minimum energy of 2.23 MeV can split a deuteron into a free neutron and a free proton (Fig. 17.2).



**EXAMPLE 1.** Calculate the binding energy of an  $\alpha$ -particle and express the result both in MeV and joules.

$$\begin{aligned}\text{SOL. } \text{Mass of 2 protons} + 2 \text{ neutrons} &= (2 \times 1.007276 + 2 \times 1.008665)u \\ &= 4.031882u.\end{aligned}$$

$$\text{Mass of the } \alpha\text{-particle} = 4.001506u.$$

$$\begin{aligned}\text{Mass defect} &\quad \Delta m = (4.031882 - 4.001506) u = 0.030376 u \\ \therefore &\quad \text{B.E.} = (0.030376 \times 931.3) \text{ MeV} = 28.29 \text{ MeV.} \\ &\quad = 45.32 \times 10^{-13} \text{ J}\end{aligned}$$

**EXAMPLE 2.** Given the following isotope masses:

$${}_3Li^7 = 7.016004, {}_3Li^6 = 6.015125 \text{ and } {}_0n^1 = 1.008665 u$$

Calculate the B.E. of a neutron in the  ${}_3Li^7$  nucleus. Express the result in u, MeV and joules.

$$\begin{aligned}\text{SOL. B.E. of a neutron in } u &= M({}_3Li^6 + {}_0n^1) - M({}_3Li^7) \\ &= 6.015125 + 1.008665 - 7.016004 \\ &= 0.007786 u.\end{aligned}$$

In MeV this becomes  $0.007786 \times 931 \text{ MeV} = 7.35 \text{ MeV}$ .

In joules,  $7.35 \times 1.6 \times 10^{-13} \text{ J} = 1.18 \times 10^{-12} \text{ J}$ .

#### Stability of Nucleus and Binding Energy

$$\text{B.E per nucleon} = \frac{\text{Total B.E. of a nucleus}}{\text{The number of nucleons it contains}}.$$

The Binding Energy per nucleon is plotted as a function of mass number  $A$  in Fig. 17.3.

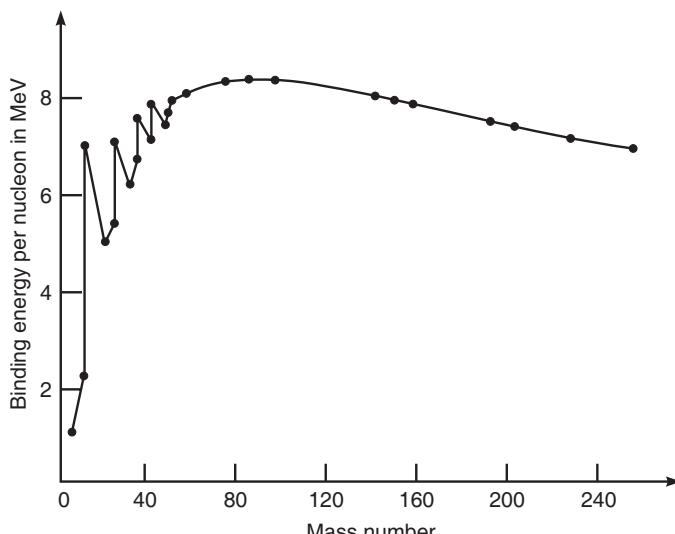


Fig. 17.3

- The curve rises steeply at first and then more gradually until it reaches a maximum of 8.79 MeV at  $A = 56$ , corresponding to the iron nucleus  ${}_{26}Fe^{56}$ .

- The curve then drops slowly to about 7.6 MeV at the highest mass numbers.

**Nuclear Fission and Fusion.** Evidently, nuclei of intermediate mass are the most stable, since the greatest amount of energy must be supplied to liberate each of their nucleons. This fact suggests that a large amount of energy will be liberated if heavier nuclei can

somehow be split into lighter ones or if light nuclei can somehow be joined to form heavier ones. The former process is known as *nuclear fission* and the latter as *nuclear fusion*. Both the processes indeed occur under proper circumstances and do evolve energy as predicted.

**Packing fraction.** The ratio between the mass defect ( $\Delta m$ ) and the mass number ( $A$ ) is called the packing fraction ( $f$ ):

$$f = \Delta m/A.$$

Packing fraction means the mass defect per nucleon. Since atomic masses are measured relative to C-12, the packing fraction for this isotope is zero. Packing fraction is a measure of the comparative stability of the atom.

Packing fraction is defined as

$$\text{Packing fraction} = \frac{\text{Isotopic mass} - \text{Mass number}}{\text{Mass number}} \times 10^4.$$

Packing fraction may have a *negative* or a *positive* sign. If packing fraction is negative, the isotopic mass is less than the mass number. In such cases, some mass gets transformed into energy in the formation of that nucleus, in accordance with Einsein's equation  $E = mc^2$ . Such nuclei, therefore, are more stable. A positive packing fraction would imply a tendency towards instability. But this is not quite correct, especially for elements of low atomic masses.

A plot of packing fraction against the corresponding mass numbers of the various elements is shown in Fig. 17.4. It is seen that helium, carbon and oxygen atoms of mass numbers 4, 12 and 16 respectively, do not fall on this curve. Their packing fractions have small values. These elements are, therefore, stable. The transition elements, with mass numbers in the neighbourhood of 45, have lowest packing fractions with a negative sign, which indicates their high stability. The packing fraction beyond mass number 200 becomes positive and increases with increase in mass number. This indicates increasing instability of these elements. Elements with mass numbers beyond 230 are radioactive and undergo disintegration spontaneously.

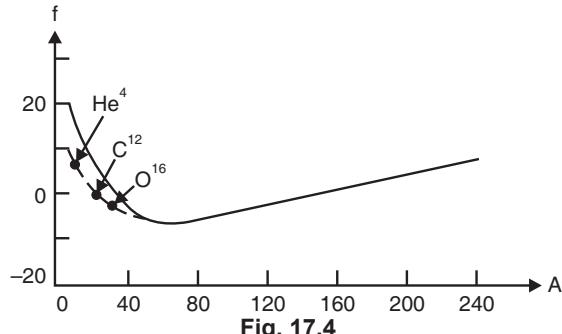


Fig. 17.4

## 17.5 NUCLEAR STABILITY

Table 17.1 shows how the 272 stable nuclei found in nature are classified according to even and odd numbers of protons and neutrons.

Table 17.1.

Protons	Neutrons	Stable Nuclides
even	even	160
even	odd	56
odd	even	52
odd	odd	4
		272

The combination of an even number of protons and an even number of neutrons, composing the nucleus, is evidently preferred by nature for stable nuclides. The odd-odd combination of stable nuclides is found only in the light elements. The number of even-odd combinations is about the same.

A plot of the number of neutrons versus the number of protons for the stable nuclides is shown in Fig. 17.5. Notice that for  $Z < 20$ , the stability line is a straight line with  $Z = N$ . For the heavier nuclides  $Z > 20$ ,  $N > Z$ , the stability curve bends in the direction of  $N > Z$ . For example,  $^{20}_{20}\text{Ca}^{48}$  has  $N = 28$ ,  $Z = 20$ ; for larger values of  $Z$ , the tendency is more pronounced, as in the case of  $^{91}_{91}\text{Pa}^{232}$  which has  $N = 141$ ,  $Z = 91$ .

Evidently, for large values of  $Z$ , the coulomb electrostatic repulsion becomes important, and the number of neutrons must be greater to compensate this repulsive effect.

Thus the curve of Fig. 17.5 departs more and more from the  $N = Z$  line as  $Z$  increases. For maximum stability, there is an optimum value of neutron/proton ratio. The number of neutrons  $N (= A - Z)$  required for maximum stability is plotted as a function of proton number  $Z$  in Fig. 17.6. All the stable nuclei fall within the shaded region. Nuclei above and below the shaded region are unstable. Artificial radioactive nuclei lie at the fringe of the region of stability. All nuclei with  $Z > 83$ , and  $A > 209$  spontaneously transform themselves into lighter ones through the emission of  $\alpha$  and  $\beta$  particles.  $\alpha$  and  $\beta$  decays enable an unstable nucleus to reach a stable configuration.

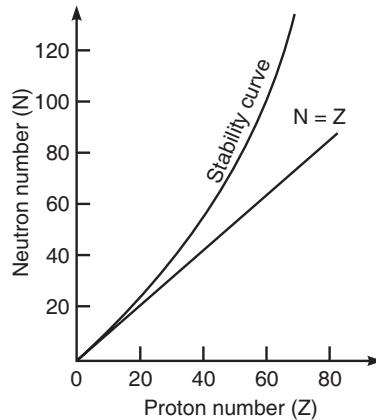


Fig. 17.5

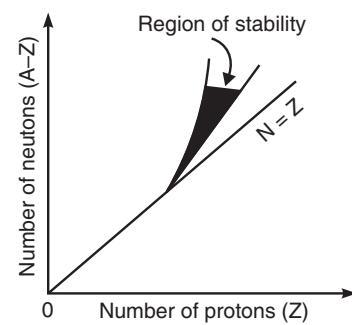


Fig. 17.6

## 17.6 THEORIES OF NUCLEAR COMPOSITION

Several theories were proposed about the nuclear constitution. Prominent among these are proton-electron theory and proton-neutron theory.

**(1) Proton-electron hypothesis.** Before the discovery of neutron by Chadwick (1932), it was assumed that the nucleus consists of protons and electrons. This concept arose because electrons were found to be emitted by radioactive nuclei in  $\beta$ -decay process. According to this hypothesis, a nucleus of mass number  $A$  and atomic number  $Z$  contains  $A$  protons and  $A-Z$  electrons. The nucleus is surrounded by  $Z$  electrons, so that the atom is electrically neutral.

**EXAMPLE.** The nucleus of  $^{11}_{11}\text{Na}^{23}$  should consist of 23 protons and 12 electrons. The total nuclear charge is 11 which is balanced by —11 units of electron charge in the extranuclear space.

There was very strong theoretical evidence for the fact that electrons cannot exist in the nucleus. We discuss below some very strong arguments which forbid the existence of electrons inside the nucleus.

### Why electrons cannot be present inside the nucleus?

**(i) Nuclear size.** Typical nuclei are less than  $10^{-14}$  m in radius. If an electron exists inside a nucleus, the uncertainty in its position ( $\Delta x$ ) may not exceed  $10^{-14}$  m. According to Heisenberg's uncertainty principle, uncertainty in the electron's momentum is

$$\Delta p \geq \frac{\hbar}{\Delta x} \geq \frac{1.054 \times 10^{-34}}{10^{-14}} \geq 1.1 \times 10^{-20} \text{ kg ms}^{-1}.$$

If this is the uncertainty in the electron's momentum, the momentum itself must be at least comparable in magnitude.

$\therefore$  Approximate momentum of the electron =  $p = 1.1 \times 10^{-20} \text{ kg ms}^{-1}$ . An electron whose momentum is  $1.1 \times 10^{-20} \text{ kg ms}^{-1}$  has a K.E. ( $T$ ) many times greater than its rest energy  $m_0 c^2$  i.e.,  $T \gg m_0 c^2$ . Hence we can use the extreme relativistic formula  $T = pc$  to find  $T$ .

$$\begin{aligned}\therefore T &= (1.1 \times 10^{-20}) (3 \times 10^8) = 3.3 \times 10^{-12} \text{ J} \\ &= \frac{3.3 \times 10^{-12}}{1.6 \times 10^{-13}} \text{ MeV} = 20.63 \text{ MeV}.\end{aligned}$$

This shows that if an electron exists in the nucleus, the K.E. of the electron must be more than  $20 \text{ MeV}$ . Electrons of such large energy are never found to be emitted during  $\beta$ -decay. The maximum energy of a  $\beta$ -particle emitted is only 2 to  $3 \text{ MeV}$ . Hence we conclude that electrons cannot be present within nuclei.

(ii) **Nuclear spin.** Electrons and protons have a spin of  $1/2$ . Thus nuclei with an even number of protons and electrons should have integral spins, while those with an odd number of protons and electrons should have half-integral spins. Let us consider deuteron as an example. Deuteron nucleus has 3 particles (two protons and one electron). Hence the nuclear spin of deuteron should be  $1/2$  or  $3/2$ . But experiment shows that the spin of the deuteron is 1. Thus the experimental result is in contradiction to the hypothesis.

(iii) **Magnetic moment.** Protons and electrons are endowed with magnetic properties. The magnetic moment of an electron is about one thousand times that of a proton. If electrons exist inside the nucleus, the magnetic moment of electrons will have a dominating influence and so nuclear magnetic moments ought to be of the same order of magnitude as that of the electron. However, the observed magnetic moments of nuclei are comparable with that of the proton. This experimental fact goes against the electrons existing inside the nucleus.

Due to these reasons, it is concluded that electrons cannot exist in the nucleus. Hence the proton-electron hypothesis regarding the constitution of the nucleus has been given up.

(2) **Proton-Neutron hypothesis.** After the discovery of neutron by Chadwick in 1932, proton-neutron theory has been gaining support. According to this, a nucleus of atomic number  $Z$  and mass number  $A$  consists of  $Z$  protons and  $A-Z$  neutrons. The number of electrons in the extra nuclear space is also  $Z$ , so that the atom as a whole is neutral.

**Merits.** (i) This assumption is able to explain the observed value of nuclear spin and nuclear magnetic moment.

(ii) There is no difficulty in explaining the existence of isotopes on this hypothesis. Isotopes of a given element differ only in the number of neutrons they contain.

(iii) The  $\beta$ -ray emission of radioactive elements is explained as follows: The electron does not pre-exist in the nucleus. The electron is formed just at the instant of emission, caused by the transformation of a neutron into a proton. i.e.,  $n \rightarrow p + e^-$ . The positron emission is due to the converse process, viz., when a proton transforms itself into a neutron. i.e.,  $p \rightarrow n + e^+$ .

(iv) The emission of  $\alpha$ -particles from the nuclei of radioactive elements is due to the combination of 2 protons and 2 neutrons at the instant of emission. Thus, this hypothesis explains  $\beta$ -decay and  $\alpha$ -decay.

## 17.7 NUCLEAR FORCES

Since stable nuclei exist, it follows that there must be certain forces acting between their nucleons that bind them into the nucleus. These are called *nuclear forces*. The nuclear force must be strongly attractive, in order to overcome the electrostatic repulsion between protons. Only three kinds of attractive forces can be conceived in the nucleus, viz., neutron-neutron ( $n-n$ ), neutron-proton ( $n-p$ ) and proton-proton ( $p-p$ ) interactions.

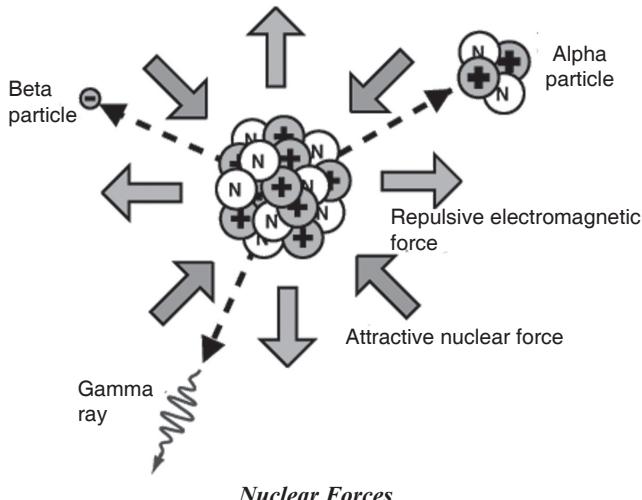
Yukawa attributed the following characteristics to the nuclear forces.

(1) **Nuclear forces are effective only at short ranges.** Nuclear forces are appreciable only when the distance between nucleons is of the order of  $10^{-15}$  m or less. The force vanishes for all practical purposes at distances greater than a few times  $10^{-15}$  m. These distances are called the *action radii* or *range* of the nuclear forces. In the up-to-date version of the exchange theory of nuclear forces, it is supposed that interaction between nucleons is accomplished by the exchange of  $\pi$ -mesons. The exchange version of nuclear forces explains their short range action. Let  $m$  be the rest mass of the  $\pi$ -meson.

Rest energy of the  $\pi$ -meson =  $\Delta E$   
 $= mc^2$ . According to Heisenberg's uncertainty principle, the time required for nucleons to exchange  $\pi$ -mesons cannot exceed  $\Delta t$ , for which  $\Delta E \Delta t \geq \hbar$ . The distance that a  $\pi$ -meson can travel away from a nucleon in the nucleus during the time  $\Delta t$ , even at a velocity  $\approx c$ , is

$$R_0 \approx \frac{\hbar}{mc} \approx 1.2 \times 10^{-15} \text{ m.}$$

This approximately coincides with the value of the nuclear radius and is of the order of magnitude of the nuclear force range.



(2) **Nuclear forces are charge independent.** The nuclear forces acting between two protons, or between two neutrons, or between a proton and a neutron, are the same. It follows that nuclear forces are of a non-electric nature.

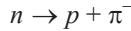
(3) **Nuclear forces are the strongest known forces in nature.**

(4) **Nuclear forces have saturation property.** Nuclear forces are limited in range. As a result, each nucleon interacts with only a limited number of nucleons nearest to it. This effect is referred to as the *saturation* of nuclear forces.

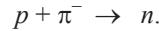
### 17.7.1. Meson Theory of Nuclear Forces

According to the meson theory of nuclear forces, all nucleons consist of identical cores surrounded by a "cloud" of one or more mesons. Mesons may be neutral or may have a positive or negative charge. The sole difference between neutrons and protons is supposed to lie in the composition of their respective meson clouds. Yukawa assumed that  $\pi$  meson is exchanged between the nucleons and that this exchange is responsible for the nuclear binding forces. The forces that act, between one neutron and another, and between one proton and another, are the result of the exchange of neutral mesons ( $\pi^0$ ) between them. The force between a neutron and a proton is the result of the exchange of charged mesons ( $\pi^+$  and  $\pi^-$ ) between them.

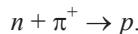
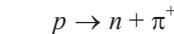
Thus a neutron emits a  $\pi^-$  meson and is converted into a proton:



The absorption of the  $\pi^-$  meson by the proton (with which the neutron was interacting) converts it into a neutron:



In the reverse process, a proton emits a  $\pi^+$  meson, becoming a neutron and the neutron, on receiving this  $\pi^+$  meson, becomes a proton :



Thus in the nucleus of an atom, attractive forces exist between (1) proton and proton (2) proton and neutron and (3) neutron and neutron. These forces of attraction are much larger than the electrostatic force of repulsion between the protons, thus giving a stability to the nucleus.

Just as a photon is a quantum of electromagnetic field, a meson is a quantum of nuclear field. Yukawa considered the equation for particle of mass  $m$  as,

$$\left( \nabla^2 - \frac{m^2 c^2}{\hbar^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \phi = 0 \quad \dots(1)$$

This is a relativistic equation valid only for spinless particles.

Separating the time dependent part, the equation for the radial part is

$$(\nabla^2 - \mu^2) \phi(r) = 0 \quad \dots(2)$$

where  $\mu = mc/\hbar$ .

$$\text{The solution of Eq. (2) is } \phi(r) = -g \frac{e^{-\mu r}}{r} \quad \dots(3)$$

Here  $g$  is a constant, which plays the same role as the charge  $q$  in electromagnetic theory. In analogy with electromagnetism, the potential between two nucleons is then given by

$$V(r) = -g^2 \frac{e^{-\mu r}}{r} \quad \dots(4)$$

Here  $g^2$  is called the '*Coupling constant*'. This argument made Yukawa predict the existence of *pion* as a quantum of nuclear force field.

The range of the pion field is  $\frac{\hbar}{m_\pi c} \approx 1.4 \text{ fm}$ .

The form of  $V(r)$  given by Eq. (4) is known as the *one-pion-exchange potential (OPEP)*.

### Mass of the meson

On the basis of the range of nuclear force and the uncertainty principle, it is possible to estimate mass of the meson. According to uncertainty principle,  $\Delta E \times \Delta t = \hbar$ .

Here,  $\Delta E$  and  $\Delta t$  are the uncertainties in energy and time.

The range of nuclear force is  $R \approx 1.4 \times 10^{-15} \text{ m}$ .

Assume that the meson travels between nuclei at approximately the speed of light  $c$ . Let  $\Delta t$  be the time interval between the emission of meson from one nucleon and the absorption by the other nucleon.

Then

$$\Delta t = R/c. \quad \therefore \Delta E = \frac{\hbar}{\Delta t} = \frac{\hbar c}{R}$$

$\therefore$  The minimum meson mass is specified by  $m \geq \frac{\hbar}{Rc}$ .

In terms of the electronic mass  $m_e$ , the mass of the meson is

$$\frac{m}{m_e} = \frac{\hbar}{m_e R c} = \frac{1.054 \times 10^{-34}}{(9.108 \times 10^{-31})(1.4 \times 10^{-15})(3 \times 10^8)} = 275$$

i.e., mass of the meson  $\approx 275 \times$  mass of electron.

In 1947, Powell discovered  $\pi$  meson of mass about  $273 m_e$ . This particle showed strong interaction with nucleons and was recognized as the Yukawa particle.

### Merits

The discovery of the meson of mass of about 273 electron mass and the existence of positive, negative and neutral mesons, lends some support to this theory.

- The experimental values of the magnetic moments of a free proton and of a free neutron also lend some support to the “Yukawa’s meson field theory” of nuclear forces.

A free proton is, for a part of its life-time, a neutron with a closely bound meson. Hence the magnetic moment of a free proton can be the resultant of the true magnetic moment of the proton and the magnetic moment of the meson. Thus the net magnetic moment of a free proton will exceed that given by the simple theory.

Similarly, a neutron is for a fraction of its life-time dissociated into a proton and a negative meson. This combination will have a negative magnetic moment. It follows that, though uncharged, a neutron will have a negative magnetic moment.

#### 17.7.2. Experimental Evidence of Short Range, Saturation, Charge Independence, Charge Symmetry, State Dependence, Tensor Nature

The most straightforward techniques to study nuclear forces are based on scattering experiments. The study is mainly about the angular distribution of neutrons and protons elastically scattered by protons. There are two essential differences between the scattering by neutron-proton system and proton-proton system.

- (i) In  $p-p$  scattering, in addition to the specific nuclear forces, Coulomb repulsive forces are also present.
- (ii) Another difference lies in that the scatterer and the scattered particles are *identical*.

#### [I] Neutron-proton scattering

Neutrons are chargeless particles. Hence  $n-p$  scattering is purely controlled by the nuclear force. To produce a monoenergetic beam of neutrons, protons are accelerated in a cyclotron. These energetic protons knock out neutrons from a hydrogenous substance in the form of a monoenergetic beam. The neutrons are scattered by the protons present in the molecules having low chemical binding say  $\sim 0.1$  eV. Thus, even for the neutrons having energy of 1 eV, their scattering by the molecules is essentially the scattering of neutrons by free protons.

From these studies, we conclude that the *nuclear force depends on the orientations of the spins of the two nucleons*.

The studies indicate that the interaction between a neutron and a proton having opposite spins is stronger than that between triplet proton-neutron system.

#### [II] Proton-proton scattering

The proton-proton scattering is different from the proton-neutron scattering.

- Since protons are charged nucleons, in the  $p-p$  scattering two type of forces, namely, nuclear and Coulom forces are involved.

- In this case, incident protons and target protons are identical particles and they obey Fermi-Dirac statistics. Hence, their total wave function, containing space and spin co-ordinates, has to be anti-symmetric. This gives rise to a new type of scattering called the ***exchange scattering***. In this scattering, the target proton enters into the detector and the incident proton becomes a target proton.

At low energies, the experimental results can be explained by considering Coulomb scattering along with  $S$  wave ( $l = 0$ ) nuclear force scattering. However, at higher energies, higher partial waves along with non-central nuclear forces are also to be included.

- Experimentally, a beam of high energy protons is extracted from a particle accelerator and made to strike a material containing hydrogen, like polyethylene. The scattered protons are counted by using suitable counter.

### Advantages

(1) Experimental study of  $p$ - $p$  scattering is easier than the  $n$ - $p$  scattering because well collimated monoenergetic beam of protons over a wide range of energy can be easily obtained.

(2) Detection of protons is easier due to their ionising properties. From these studies, we learn the following properties of nuclear forces:

**Short range.** The study reveals that the internuclear potential has a hard core that prevents the nucleons from approaching each other much closer than about 0.4 fm. Two nucleons within about 2 fm of each other feel an *attractive nuclear force*. The nuclear force is essentially zero outside to about 3 fm. Since the nuclear force falls to zero very abruptly, it is a *short-range force*.

**Saturation property.** The nuclear force is quite strong but of short range ( $\sim$  fermi). Hence nucleons interact strongly only with their nearest neighbours. This effect is referred to as the saturation of the nuclear forces.

**Charge independence.** The nuclear force does not depend upon the charge of the nucleons. Hence the force between  $p$  and  $p$  is same as those between  $p$  and  $n$  and  $n$  and  $n$ .

**Spin dependence.** The nuclear force is known to be *spin dependent* because the bound state of the deuteron has the neutron and proton spins aligned. But there is no bound state with the spins antialigned (that is coupled to total spin 0). Nuclear forces depend on relative orientations of spin and displacement vector.

**Tensor force (Non-central Force).** The force is not exactly central. It has a non-central component also. This non-central force, which depends upon  $r$  and the direction of  $r$  is called ***tensor force***.

### 17.7.3. Interpretation of High Energy Nucleon-Nucleon Scattering—Exchange Forces

The ordinary (non-exchange) central forces are introduced by the more familiar type of potentials:

(i) the square-well (ii) exponential (iii) Gaussian and (iv) Yukawa. They are all purely radial potentials and have no exchange character whatsoever. Such potentials give rise to what are called *Wigner forces* in nuclear physics.

- The exchange forces are unknown to classical physics and are purely quantum mechanical in nature. Scattering experiments indicate that nuclear forces are strong and short-range and that they depend not only on ' $r$ ' – the separation between the nucleons but also upon the orientation of  $r$  in space and the spin of the nucleons. If  $r_1, \sigma_1$  stand for the space and spin co-ordinates of the first nucleon respectively and  $r_2, \sigma_2$  for those of the second nucleon, then the wave function for the nucleon pair may be written as  $\psi(r_1\sigma_1, r_2\sigma_2)$ . Then it is possible to introduce only three different type of *quantum mechanical exchange operators*, which operating upon the wave-function  $\psi(r_1\sigma_1, r_2\sigma_2)$ , exchange:

- (i) the position vectors  $r_1$  and  $r_2$  of the nucleons (space exchange) only,
- (ii) the spin co-ordinates  $\sigma_1$  and  $\sigma_2$  (spin exchange) only,
- (iii) the position vector as well as the spin co-ordinates (space-spin exchange).

These give rise to (i) Majorana, (ii) Bartlett, (iii) Heisenberg forces respectively. The explicit results of these operators on the wave function  $\Psi$  are tabulated below:

**Table 17.2. Possible exchange operators**

	Name	Definition of the Operator
1.	Majorana exchange operator ( $P^M$ )	$P^M \psi(r_1\sigma_1, r_2\sigma_2) - \psi(r_2\sigma_1, r_1\sigma_2)$
2.	Bartlett exchange operator ( $P^B$ )	$P^B \psi(r_1\sigma_1, r_2\sigma_2) - \psi(r_1\sigma_2, r_2\sigma_1)$
3.	Heisenberg exchange operator ( $P^H$ )	$P^H \psi(r_1\sigma_1, r_2\sigma_2) - \psi(r_2\sigma_2, r_1\sigma_1)$

Three types of exchange forces

- (i) The space exchange or Majorana force.
- (ii) The spin exchange or Bartlett force.
- (iii) The space-spin exchange or Heisenberg force.

#### 17.7.4. Deuteron-Ground State Properties

A deuteron has one proton and one neutron. Table 17.3 gives various properties of the deuteron.

**Table 17.3. Deuteron – Ground State Properties**

Mass	2.013553 $u$
Charge	$1.602 \times 10^{-19} C$
Binding energy	2.224 MeV
Average radius	4.2 fm
Angular momentum	$1\hbar$
Magnetic dipole moment	$0.8574 \mu_N$
Electric quadrupole moment	$0.282 \text{ fm}^2$

Let us discuss some of the properties.

(i) **Binding energy.** Deuterium nucleus is the simplest among the atomic nuclei. It contains one proton and one neutron. These two particles are bound together with a binding energy of 2.225 MeV. There are essentially two direct methods of experimental measurement of the binding energy of deuteron. One is to measure the energy of gamma radiation emitted when a neutron and proton combine to form a nucleus of deuteron ( $n - p$  capture reaction). The second method consists of measuring the energy of gamma radiation that will break the bond between the neutron and proton (photo disintegration of deuteron).

(ii) **Angular momentum and parity.** The angular momentum quantum number of the ground state of the deuteron is  $I = 1$ . This is obtained as the resultant of the spin of proton ( $1/2$ ) and of neutron ( $1/2$ ). The orbital angular momentum for both the nucleons will be zero. The parity is taken as +ve, which is corroborated by experimental results.

(iii) **Magnetic Dipole moment.** The magnetic dipole moment of the deuteron is  $+ 0.85739 \mu_N$  as given by N.M.R. data. The magnetic dipole moment of a proton is  $+ 2.79275 \mu_N$  and that of a neutron is  $- 1.91315 \mu_N$ . The magnetic dipole moment of deuteron is slightly smaller than the resultant obtained by the algebraic sum of the magnetic moments of proton and of neutron.

(iv) **Electric quadrupole moment.** The electric quadrupole moment of the deuteron is  $Q = +0.00282 \text{ barn} = 0.282 \text{ fm}^2$ . The quantity of the quadrupole moment is an indication that the charge distribution is not spherically symmetrical in deuteron. The +ve sign for  $Q$  indicates that the charge distribution is prolate rather than oblate.

(v) **Radius of deuteron.** The size of the deuteron is found by studying the scattering of high energy electrons. The average radius of the deuteron is found to be 4.2 fermi. [It is the root mean square distance between the neutron and proton].

(vi) **The force between a neutron and proton is spin-dependent.** I, Angular momentum of deuteron = 1 in units of  $\hbar$ . Since the spins of both, the neutron and the proton, is  $\frac{1}{2}$  in the units of  $\hbar$ , the deuteron must be formed with the spin vectors of the neutron and the proton parallel. The other possibility of the two spin vectors antiparallel giving a net spin = 0 does not lead to a bound state for a deuteron. Deuteron excited states do not exist and so the state to be studied is the ground state. For the ground state,  $l = 0$ . It is  $1s$  state.

∴

$$I = L + S = S = 1$$

The  $1s$  state (ground state) with  $S = 0$  does not give a bound state for deuteron. The energy of the ground state of the deuteron is different if the spins of the neutron and proton are parallel than if they are anti-parallel. Otherwise why should the spin parallel system be stable and spin anti-parallel not stable? The important conclusion that can be drawn from this is that the *force between a neutron and a proton is spin-dependent*. The nuclear force keeping the two nucleons system bound, thus depends on whether the spins of the two nucleons are parallel or antiparallel i.e., whether the total spin  $S = 1$  or 0. This feature is quite different than, for example, the Coulomb force which goes as  $1/r^2$  and so is just a function of  $r$ .

### MEASUREMENT OF NUCLEAR RADIUS

## 17.8 HIGH ENERGY ELECTRON SCATTERING EXPERIMENT

**Experimental set up.** Fig. 17.16 shows Hofstadter's experimental arrangement for nuclear radius measurement.

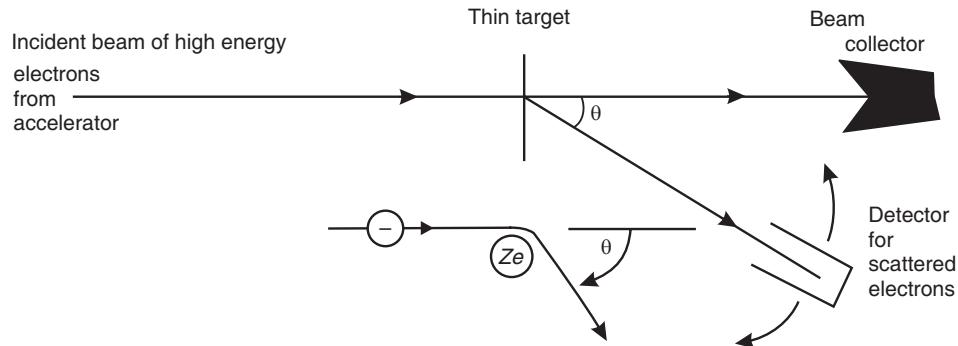
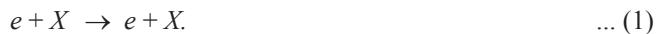


Fig. 17.6

- This method consists of measurement of nuclear charge distribution and assumes that charge distribution and matter distribution within nuclei are essentially the same. The charge distribution in nuclei is obtained by analysing scattering of high-energy electrons from various nuclei. The elastic-scattering process is



*Electron* is chosen as the bombarding particle because its electro-magnetic interaction with the nucleus is well-known.

- The equipment includes an electron accelerator to prepare the high energy electron beam, a scattering target of species  $X$ , and a spectrometer to detect electrons scattered elastically in directions given by the indicated scattering angle  $\theta$ .
- The de Broglie wavelength of 200 MeV electrons is about  $10^{-14}$  m which is the size of a nucleus. Thus to get information about nuclear charge distribution, the incident electron beam energy has to be of the order of 200 MeV.

**Method.** The electron beam is incident on a thin foil of gold called the target and gets scattered. The elastically scattered electrons are observed by a detector as a function of **scattering angle**  $\theta$ .

- Portions of electron wave incident on different parts of the nucleus will be scattered in a particular direction, with phase differences resulting in constructive or destructive interference at some angle.

Fig. 17.7 (a) shows the angular distribution of elastically scattered 200 MeV electrons from nuclei assumed to have spherical uniform charge distribution (having uniform density up to radius  $r$ ). The actual nuclear charge distribution is not exactly given by a step function, but is shown by the dotted line [Fig. 17.7 (b)].

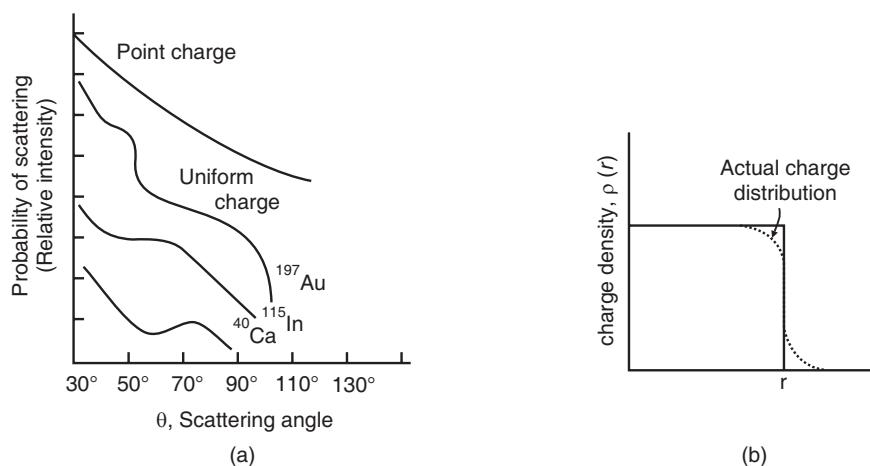


Fig. 17.7

Hofstadter assumed that the nuclear charge density has the form

$$\rho(r) = \frac{\rho_0}{1 + e^{(r - R)/b}} \quad \dots(2)$$

Here,  $\rho_0$  is density at the nuclear centre,  $R$  is the radius at which  $\rho$  falls to  $\rho_0/2$  and  $b$  measures how rapidly  $\rho$  falls to zero at the nuclear surface.

Nuclear charge distribution found in some nuclei using high-energy electrons as probes is shown in Fig. 28.3. This analysis determines a nuclear radius  $R$  that grows with mass number  $A$  according to the formula

$$R = R_0 A^{1/3} \quad \dots(3)$$

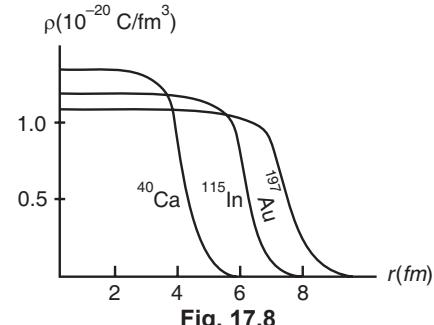
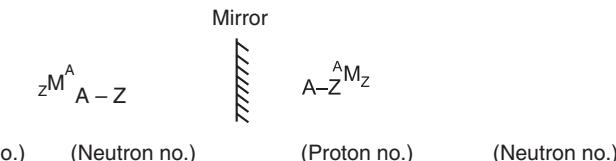


Fig. 17.8

## 17.9 DETERMINATION OF NUCLEAR RADIUS—MIRROR NUCLEI METHOD

Two nuclei having same number of nucleons (protons + neutrons) but the number of protons in one of them being equal to the number of neutrons in the other are called the mirror-nuclei (Fig. 17.9.)



**Fig. 17.9**

**EXAMPLE.**  ${}^6C^{II}$  and  ${}^5B^{II}$  are mirror nuclei.  ${}^6C^{II}$  contains 6 protons and 5 neutrons while  ${}^5B^{II}$  contains 5 protons and 6 neutrons.

The other examples of mirror nuclei are:  $({}^3H, {}^3He)$ ;  $({}^7Li, {}^7Be)$ ;  $({}^{13}C, {}^{13}N)$ ;  $({}^{39}K, {}^{39}Ca)$ .

Bethe suggested that the difference of Coulomb energies between two neighbouring mirror nuclei can be used to calculate nuclear radii.

Consider a spherical nucleus of radius  $R$  and atomic number  $Z$  (Fig. 17.10). Each proton carries a charge  $+e$ . Thus the total charge is  $+Ze$ .

$$\therefore \text{Charge density}, \quad \rho = \frac{\text{Charge on nucleus}}{\text{Volume of nucleus}} = \frac{Ze}{\frac{4}{3}\pi R^3} \quad \dots(1)$$

The electrostatic energy of nucleus is the work done against electrostatic forces in assembling such a sphere.

Suppose at any instant we have a spherical core of radius  $x$ .

$$\text{Charge on this spherical core} = q_1 = \frac{4}{3}\pi x^3 \rho$$

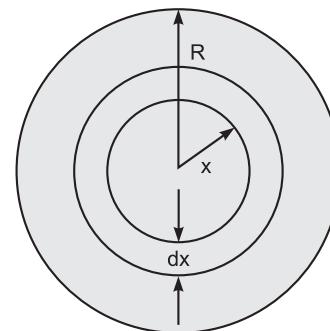
Now we bring a thin spherical shell of radius  $x$  and thickness  $dx$  from infinity in the process of assembling the charge.

The charge on this shell is

$$q_2 = \text{volume of the shell} \times \text{charge density} = (4\pi x^2 dx) \rho$$

The electrostatic potential energy of this spherical core and thin spherical shell

$$\begin{aligned} dE_c &= \frac{1}{4\pi \epsilon_0} \frac{q_1 q_2}{x} \\ &= \frac{1}{4\pi \epsilon_0} \frac{\left(\frac{4}{3}\pi x^3 \rho\right) (4\pi x^2 dx \rho)}{x} = \frac{1}{4\pi \epsilon_0} \left(\frac{16}{3}\pi^2 \rho^2 x^4 dx\right) \end{aligned}$$



**Fig. 17.10**

Total electrostatic potential energy of nucleus is obtained by integrating this expression between limits 0 and  $R$ .

$$E_c = \int_0^R \frac{1}{4\pi \epsilon_0} \left(\frac{16}{3}\pi^2 \rho^2 x^4 dx\right) = \frac{1}{4\pi \epsilon_0} \frac{16}{3}\pi^2 \rho^2 \left[\frac{x^5}{5}\right]_0^R$$

Substituting value of  $\rho$  from Eq. (1), we get

$$E_c = \frac{1}{4\pi\epsilon_0} \frac{16}{3} \pi^2 \cdot \left( \frac{Ze}{\frac{4}{3}\pi R^3} \right)^2 \cdot \frac{R^5}{5} = \frac{1}{4\pi\epsilon_0} \cdot \frac{3}{5} \frac{Z^2 e^2}{R} \quad \dots(2)$$

The difference in Coulomb energies between two neighbouring nuclei of charges ( $Ze$ ) and  $(Z+1)e$  having the same radius  $R$  is

$$\begin{aligned} \Delta E_c &= \frac{1}{4\pi\epsilon_0} \frac{3}{5} \frac{(Z+1)^2 e^2}{R} - \frac{1}{4\pi\epsilon_0} \frac{3}{5} \frac{Z^2 e^2}{R} = \frac{1}{4\pi\epsilon_0} \frac{3}{5} \frac{e^2}{R} [(Z+1)^2 - Z^2] \\ \Delta E_c &= \frac{1}{4\pi\epsilon_0} \frac{3}{5} \frac{e^2}{R} (2Z+1) \end{aligned} \quad \dots(3)$$

In the mirror nuclei  $(2Z+1) = A$ , where  $A$  is mass number. Eq. (3) becomes

$$\Delta E_c = \frac{1}{4\pi\epsilon_0} \frac{3}{5} \frac{e^2 A}{R} \quad \dots(4)$$

Taking experimental observations of mirror nuclei ( $^{14}_{14}Si^{29}$ ,  $^{15}_{15}P^{29}$ ), the difference in Coulomb-energies comes out to be 4.96 MeV. Substituting this value,  $R = 4.96$  fermi.

From the relation  $R = R_0 A^{1/3}$ , we obtain the radial parameter,  $R_0 \approx 1.5$  fermi.

#### MODELS OF NUCLEAR STRUCTURE

The precise nature of the forces acting in the nucleus is unknown. Hence, nuclear models are resorted to for investigation and theoretical prediction of its properties. Such models are based on (i) the extrinsic analogy between the properties of atomic nuclei and those of a liquid drop (ii) the electron shell of an atom etc. The corresponding models are called the liquid-drop model, shell model, etc.

### 17.10 THE LIQUID DROP MODEL

In the liquid-drop model, the forces acting in the nucleus are assumed to be analogical to the molecular forces in a droplet of some liquid. This model was proposed by Neils Bohr who observed that there are certain marked similarities between an atomic nucleus and a liquid drop. The similarities between the nucleus and a liquid drop are the following:

- (i) The nucleus is supposed to be spherical in shape in the stable state, just as a liquid drop is spherical due to the symmetrical surface tension forces.
- (ii) The force of surface tension acts on the surface of the liquid-drop. Similarly, there is a potential barrier at the surface of the nucleus.
- (iii) The density of a liquid-drop is independent of its volume. Similarly, the density of the nucleus is independent of its volume.
- (iv) The intermolecular forces in a liquid are short range forces. The molecules in a liquid drop interact only with their immediate neighbours. Similarly, the nuclear forces are short range forces. Nucleons in the nucleus also interact only with their immediate neighbours. This leads to the saturation in the nuclear forces and a constant binding energy per nucleon.
- (v) The molecules evaporate from a liquid drop on raising the temperature of the liquid due to their increased energy of thermal agitation. Similarly, when energy is given to a nucleus by

bombarding it with nuclear projectiles, a compound nucleus is formed which emits nuclear radiations almost immediately.

(vi) When a small drop of liquid is allowed to oscillate, it breaks up into two smaller drops of equal size. The process of nuclear fission is similar and the nucleus breaks up into two smaller nuclei.

**Semi-empirical mass formula.** The liquid-drop model can be used to obtain an expression for the binding energy of the nucleus. Weizacker proposed the semi-empirical nuclear binding energy formula for a nucleus of mass number  $A$ , containing  $Z$  protons and  $N$  neutrons. It is written as

$$\text{B.E.} = aA - bA^{2/3} - \frac{cZ(Z-1)}{A^{1/3}} - \frac{d(N-Z)^2}{A} \pm \frac{\delta}{A^{3/4}}$$

where  $a, b, c, d$  and  $\delta$  are constants.

**Explanation of the terms.** (1) The first term is called the *volume energy* of a nucleus ( $E_v = aA$ ). The larger the total number of nucleons  $A$ , the more difficult it will be to remove the individual protons and neutrons from the nucleus. The B. E. is directly proportional to the total number of nucleons  $A$ .

(2) The nucleons, at the surface of the nucleus, are not completely surrounded by other nucleons. Hence energy of the nucleon on the surface is less than that in the interior. The number of surface nucleons depends upon the surface area of the nucleus. A nucleus of radius  $R$  has an area of  $4\pi R^2 = 4\pi r_o^2 A^{2/3}$ . Hence the *surface effect* reduces the B.E. by  $E_s = b A^{2/3}$ . The negative energy  $E_s$  is called the *surface energy* of a nucleus. It is most significant for the lighter nuclei, since a greater fraction of their nucleons are on the surface.

(3) The *electrostatic repulsion* between each pair of protons in a nucleus also contributes towards decreasing its B.E. The Coulomb energy  $E_c$  of a nucleus is the work that must be done to bring together  $Z$  protons from infinity into a volume equal to that of the nucleus. Hence  $E_c \propto Z(Z-1)/2$  (the number of proton pairs in a nucleus containing  $Z$  protons) and  $E_c$  is inversely proportional to the nuclear radius  $R = r_o A^{1/3}$ .  $E_c$  is negative because it arises from a force that opposes nuclear stability.

(4) The fourth term  $E_a = \frac{d(N-Z)^2}{A}$  originates from the *lack of symmetry* between the

number of protons ( $Z$ ) and the number of neutrons ( $N$ ) in the nucleus. The maximum stability of a nucleus occurs when  $N = Z$ . Any departure from this introduces an asymmetry  $N-Z$ , which results in a decrease in stability. The decrease in the B.E. arising from this is called the asymmetric energy ( $E_a$ ). This is also negative.

(5) The final correction term  $\delta$  allows for the fact that *even-even* nuclei are more stable than *odd-odd* nuclei.  $\delta$  is positive for even-even nuclei, is negative for odd-odd nuclei and  $\delta = 0$  for an odd  $A$ .

The best values of the constants, expressed in MeV, are  $a = 15.760$ ;  $b = 17.810$ ,  $c = 0.711$ ,  $d = 23.702$ ,  $\delta = 34$ .

The contributions of the various effects in Weizacker's empirical formula are represented schematically in the graph of Fig. 17.11.

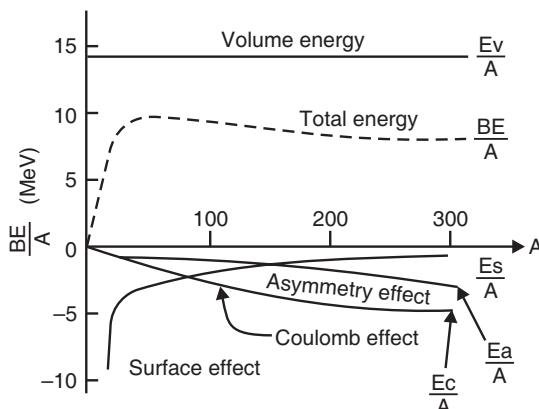


Fig. 17.11

**Merits.** (1) The liquid drop model accounts for many of the salient features of nuclear matter, such as the observed binding energies of nuclei and their stability against  $\alpha$  and  $\beta$  disintegration as well as nuclear fission.

(2) The calculation of atomic masses and binding energies can be done with good accuracy with the liquid drop-model.

However, this model fails to explain other properties, in particular the magic numbers. It fails to explain the measured spins and magnetic moments of nuclei.

**EXAMPLE 1.** Calculate the atomic number of the most stable nucleus for a given mass number  $A$ .

**SOL.** The most stable nucleus with a given mass number  $A$  is that which has the maximum value of the B.E. Thus we have to compute  $\frac{\partial (B.E.)}{\partial Z}$  with  $A$  constant, and equate it to zero. In the formula for B.E., we can write  $Z(Z-1) = Z^2$  and  $N-Z = A-2Z$ . Then

$$B.E. = aA - bA^{2/3} - \frac{cZ^2}{A^{1/3}} - \frac{d(A-2Z)^2}{A} \pm \frac{\delta}{A^{3/4}}$$

$$\frac{\delta E_b}{\delta Z} = -2cZA^{-1/3} + 4d(A-2Z)A^{-1} = 0$$

or, introducing the numerical values of  $c$  and  $d$ , we have

$$Z = \frac{A}{2 + 0.0157 A^{2/3}}.$$

For light nuclei, having small  $A$ , we can neglect the second term in the denominator.  
 $\therefore Z \approx A/2$ . This result is confirmed experimentally.

**EXAMPLE 2.** Obtain an expression for energy released in symmetric fission.

(Mumbai University, October 2010)

**SOL.** In a 'symmetric fission', a nucleus  $(Z, A)$  breaks into two equal halves  $(Z/2, A/2)$ . Semi-empirical mass formula gives a satisfactory account of fission energetics.

The energy released in a symmetric fission is:

$$Q = M(Z, A) - 2M(Z/2, A/2) \quad \dots(1)$$

From the semi-empirical mass formula, binding energy for a nucleus  $(Z, A)$  is:

$$B(Z, A) = a_v A - a_s A^{2/3} - a_c \frac{Z(Z-1)}{A^{1/3}} - a_a \frac{(A-2Z)^2}{A} - a_p A^{-3/4} \quad \dots(2)$$

Binding energy for two fission fragments  $(Z/2, A/2)$  is:

$$2B(Z/2, A/2) = a_v A - 2a_s \left(\frac{A}{2}\right)^{2/3} - 2a_c \frac{\frac{Z}{2} \left(\frac{Z}{2}-1\right)}{\left(\frac{A}{2}\right)^{1/3}} - 2a_a \frac{(A/2-Z)^2}{A/2} - 2a_p \left(\frac{A}{2}\right)^{-3/4} \quad \dots(3)$$

$$Q = a_s A^{2/3} \left(1 - 2^{1/3}\right) + a_c \frac{Z(Z-1)}{A^{1/3}} \left(1 - \frac{1}{2^{2/3}}\right) + \text{pairing energy difference.}$$

The difference in pairing energy is small and thus can be neglected as an approximation.

$$\therefore Q = A^{2/3} \left(-0.260a_s + 0.370 \frac{Z(Z-1)}{A} a_c\right) \quad \dots(4)$$

Using the values of  $a_s = 13 \text{ MeV}$  and  $a_c = 0.60 \text{ MeV}$ , we get

$$Q \sim A^{2/3} \left( -3.38 \text{ MeV} + 0.22 \text{ MeV} \frac{Z(Z-1)}{A} \right) \quad \dots(5)$$

From this we see that  $Q$  becomes  $> 0$  for

$$\frac{Z^2}{A} \geq \frac{3.38}{0.22} \sim 15 \quad \dots(6)$$

Thus for all nuclei having,  $Z > 35$ ,  $A > 80$ , fission is possible and would release energy.

For the symmetric fission of  $^{238}\text{U}$ ,

$$Q = (238)^{2/3} \left( -3.38 + 0.22 \frac{92(92-1)}{238} \right) \text{ MeV} \sim 170 \text{ MeV} \quad \dots(7)$$

### 17.11 APPLICATIONS OF SEMI-EMPIRICAL MASS-FORMULA

#### 1. Mass Parabolas: Prediction of stability of nuclei against $\beta$ -decay.

The isobaric nuclei have the same mass number  $A$  but different atomic number  $Z$ .

The semi-empirical mass formula can be written as,

$$\begin{aligned} {}_Z M^A &= Z(M_p - M_n) + A(M_n - a_v) + a_c \frac{Z^2}{A^{1/3}} + a_s A^{2/3} + a_a A - a_a 4Z + a_a \frac{4Z^2}{A} \pm \delta \\ {}_Z M^A &= \alpha A + \beta Z + \gamma Z^2 \pm \delta \end{aligned} \quad \dots(1)$$

$$\text{Here, } \alpha = M_n - \left( a_v - a_a - \frac{a_s}{A^{1/3}} \right)$$

$$\beta = -4a_a - (M_n - M_p)$$

$$\gamma = \left( \frac{4a_a}{A} + \frac{a_c}{A^{1/3}} \right)$$

$$\delta = 0 \text{ for odd-}Z, \text{ even } N$$

$$= 0 \text{ for even-}Z, \text{ odd-}N$$

$$= +\delta \text{ of odd-}Z, \text{ odd-}N$$

$$= -\delta \text{ for even } Z, \text{ even-}N$$

$\pm \delta$  is the pairing energy and does not contain terms in  $Z$ .

When  $A$  is constant, Eq. (1) is the equation of a *parabola*. It is known as the *mass parabola*. Eq. (1) gives the dependence of nuclear mass on the nuclear charge with *constant A*. This dependence is *parabolic*.

The most stable nucleus has the *minimum* mass.

Differentiating Eq. (1) for constant  $A$  and equating it to zero, we have

$$\begin{aligned} \frac{\partial}{\partial Z} ({}_Z M^A) &= 0 = \beta + 2\gamma Z_0 \text{ at the point } Z = Z_0. \\ \therefore Z_0 &= \text{nuclear charge of the "most stable" isobar} = -\frac{\beta}{2\gamma} \end{aligned} \quad \dots(2)$$

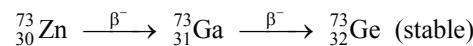
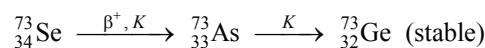
**(i) Mass parabolas of odd  $A$  isobars**

For Odd  $A$  nuclei,  $\delta = 0$  and thus there is only one parabola, implying that there is only one stable nucleus.

The isobar  $Z = Z_0 + 1$  lying on the right hand side of the parabola (Fig. 17.12), has greater mass. It will undergo  $\beta^+$  decay, giving rise to a stable isobar with  $Z = Z_0$ .

Similarly, the isobar lying on the left of the curve will undergo  $\beta^-$  decay.

The situation is similar for two isobars  $(A, Z_0 + 2)$  and  $(A, Z_0 - 2)$  which can decay to  $(A, Z_0 + 1)$  and  $(A, Z_0 - 1)$ , respectively; and so on.



**Transition Energies**

$$\begin{aligned} {}_{Z_0}^A M^A &= \alpha A + \beta Z_0 + \gamma Z_0^2 = \alpha A - \gamma Z_0^2 && \text{(using Eq. 2)} \\ {}_Z^A M^A - {}_{Z_0}^A M^A &= (\alpha A + \beta Z + \gamma Z^2) - (\alpha A - \gamma Z_0^2) = \gamma (Z - Z_0)^2 && \dots(3) \end{aligned}$$

Energy released in  $\{Z \rightarrow (Z + 1)\}$  transition is

$$Q_\beta^- = {}_Z^A M^A - {}_{Z+1}^A M^A = \gamma [(Z - Z_0)^2 - \{(Z + 1) - Z_0\}^2] = 2\gamma (Z_0 - Z - 1/2)$$

The  $Q$  value for  $Z \rightarrow (Z - 1)$  transition is given by

$$Q_\beta^+ = {}_Z^A M^A - {}_{Z-1}^A M^A = \gamma \{(Z - Z_0)^2 - (Z - 1 - Z_0)^2\} = 2\gamma [Z - Z_0 - 1/2]$$

**(ii) Mass parabolas of even  $A$  Isobars**

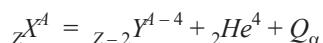
Here the pairing term  $\delta \neq 0$ . Since both odd-odd and even-even nuclei are included, we have two parabolas. The lower parabola corresponds to more stable nuclei with even  $Z$ , while the upper parabola to less stable nuclei with odd  $Z$ . The vertical separation between two parabolas is  $2\delta$  due to the opposite sign of  $\delta$  in two cases.

Fig. 17.13 shows the decay scheme. The reaction energy  $Q$  for transition  $Z \rightarrow Z \pm 1$  for isobars of mass number  $A$  is

$$Q_\beta = 2\gamma \{\pm [Z_0 - Z] - 1/2\} \pm 2\delta \quad \begin{cases} + 2\delta & \text{for odd } Z \\ - 2\delta & \text{for even } Z \end{cases}$$

**2. Alpha Decay:**

The  $\alpha$ -decay process is conventionally written as



The  $Q_\alpha$ -disintegration energy involved in  $\alpha$ -decay process is written as

$$Q_\alpha = {}_Z^A M^A - {}_{Z-2}^A M^{A-4} - {}_2^4 M^A ({}_2^4 \text{He})$$

In terms of B.E. of nuclei involved, this expression for large  $Z$  and  $A$  becomes

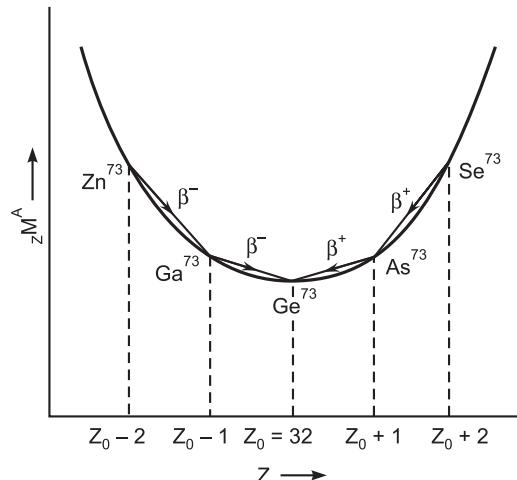


Fig. 17.12

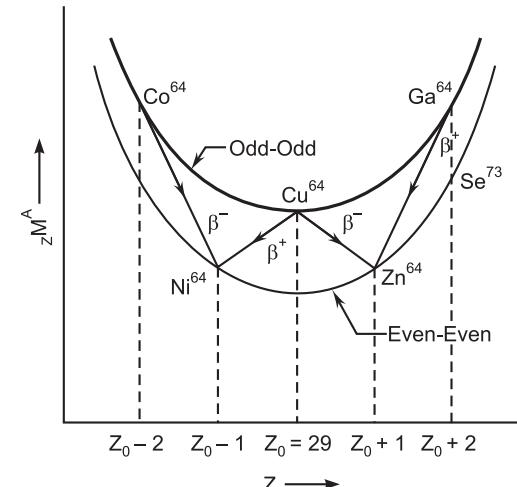


Fig. 17.13

$$\begin{aligned}
Q_\alpha(A, Z) &= \text{B.E.}(A-4, Z-2) + \text{B.E.}({}_2^4He) - \text{B.E.}(A, Z) \\
&= a_v(A-4) - a_s(A-4)^{2/3} - a_c \frac{(Z-2)^2}{(A-4)^{1/3}} - a_a \frac{(A-2Z)^2}{(A-4)} \\
&\quad - a_v A + a_s A^{2/3} + a_c \frac{Z^2}{A^{1/3}} + a_a \frac{(A-2Z)^2}{A} + \text{B.E.}({}_2^4He) \\
&= \text{B.E.}({}_2^4He) - 4a_v + a_s \{A^{2/3} - (A-4)^{2/3}\} \\
&\quad + a_c \left\{ \frac{Z^2}{A^{1/3}} - \frac{(Z-2)^2}{(A-4)^{1/3}} \right\} + a_a (A-2Z)^2 \left( \frac{1}{A} - \frac{1}{(A-4)} \right) \\
&= 28.3 - 4a_v + \frac{8}{3} a_s A^{-1/3} + \frac{4a_c Z}{A^{1/3}} \left( 1 - \frac{Z}{3A} \right) - \frac{4a_a (A-2Z)^2}{A(A-4)}
\end{aligned}$$

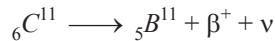
In the above formulation, pairing energy term has been neglected. If numerical values of  $a_v$ ,  $a_s$ ,  $a_c$  and  $a_a$  are employed, it is found that  $Q_\alpha > 0$  for  $A > 160$ . Therefore, the nuclei  $A > 160$  should be  $\alpha$ -disintegrating. Practically  $\alpha$ -disintegration is observed in nuclei with  $A > 200$ . For nuclei  $A < 200$ , the energy release is so small that barrier penetration becomes small.

**EXAMPLE 3.** What are mass parabolas? Define mirror nuclei and obtain an expression for  $\beta$  disintegration energy of mirror nuclei. (Mumbai University, October 2010)

#### Beta Disintegration Energy of Mirror Nuclei:

Mirror nuclei are pairs of isobaric nuclei in which the proton and neutron numbers are interchanged and differ by one unit.  ${}_6^C{}^{11}$  and  ${}_5^B{}^{11}$  are mirror nuclei. For such pairs  $A = 2Z - 1$ .

The members of the pairs with higher  $Z$  are usually found  $\beta^+$  emitters such as



The maximum energy of the  $\beta^+$  emitted in this case is 0.96 MeV. No  $\gamma$ -ray energy is emitted in the transition.

We neglect pairing energy term  $\delta$  (for odd  $A$ ).

The semi-empirical mass formula is written as

$$\begin{aligned}
M(A, Z) &= Z M_p + N M_n - a_v A + a_s A^{2/3} + a_c \frac{Z^2}{A^{1/3}} + a_a \frac{(A-2Z)^2}{A} \\
&= Z M_p + (Z-1) M_n - a_v A + a_s A^{2/3} + a_c \frac{Z^2}{A^{1/3}} + \frac{a_a}{A}
\end{aligned}$$

For the daughter term

$$\begin{aligned}
M(A, Z-1) &= (Z-1) M_p + (Z-2) M_n - a_v A + a_s A^{2/3} + a_c \frac{(Z-1)^2}{A^{1/3}} + \frac{a_a}{A} \\
M(A, Z) - M(A, Z-1) &= M_p - M_n + \frac{a_c (2Z-1)}{A^{1/3}} \\
\therefore Q_{\beta^+} &= M(A, Z) - M(A, Z-1) - 2m_e \\
&= M_p - M_n + \frac{a_c (2Z-1)}{A^{1/3}} - 2m_e \\
&= a_c A^{2/3} - (M_n - M_p + 2m_e)
\end{aligned}$$

$$Q_{\beta^+} = a_c A^{2/3} + 1.804 \text{ MeV}$$

The equation shows that plot of disintegration energy  $Q_{\beta^+}$  against  $A^{2/3}$  should be straight line with a slope  $a_c = \frac{3e^2}{20\pi\epsilon_0 r_0}$ . The slope gives nuclear radius of the order 1.44 fermi.

### 17.12 THE SHELL MODEL

The shell model of the nucleus assumes that the energy structure (energy levels of the nucleons) of the nucleus is similar to that of an electron shell in an atom. According to this model, the protons and neutrons are grouped in shells in the nucleus, similar to extra-nuclear electrons in various shells outside the nucleus. The shells are regarded as "filled" when they contain a specific number of protons or neutrons or both. The number of nucleons in each shell is limited by the *Pauli exclusion principle*. The shell model is sometimes referred to as the *independent particle model* because it assumes that each nucleon moves independently of all the other nucleons and is acted on by an average nuclear field produced by the action of all the other nucleons.

**Evidence for shell model.** It is known that a nucleus is stable if it has a certain definite number of either protons or neutrons. These numbers are known as *magic numbers*. The magic numbers are 2, 8, 20, 50, 82 and 126. Thus nuclei containing 2, 8, 20, 50, 82 and 126 nucleons of the same kind form some sort of closed nuclear shell structures. The main points in favour of this inference are:

- (i) The inert gases with closed electron shells exhibit a high degree of chemical stability. Similarly, nuclides whose nuclei contain a magic number of nucleons of the same kind exhibit more than average stability.
- (ii) **Relative abundance.** Isotopes of elements having an isotopic abundance greater than 60% belong to the magic number category.
- (iii) **Stable isotopes.** Tin ( $_{50}\text{Sn}$ ) has ten stable isotopes, while calcium ( $_{20}\text{Ca}^{40}$ ) has six stable isotopes. So elements with  $Z = 50, 20$  are more than usually stable.
- (iv) **Stable end products of three radioactive series.** The three main radioactive series (viz., the uranium series, actinium series and thorium series) decay to  $_{82}\text{Pb}^{208}$  with  $Z = 82$  and  $N = 126$ . Thus lead  $_{82}\text{Pb}^{208}$  is the most stable isotope. This again shows that the numbers 82 and 126 indicate stability.
- (v) **Neutron capture cross section.** Nuclei having a number of neutrons equal to the magic number, cannot capture a neutron because the shells are closed and they cannot contain an extra neutron.
- (vi) Some isotopes are spontaneous neutron emitters when excited above the nucleon binding energy by a preceding  $\beta$ -decay. These are  $_{8}\text{O}^{17}$ ,  $_{36}\text{Kr}^{87}$  and  $_{54}\text{Xe}^{137}$  for which  $N = 9, 51$  and  $83$  which can be written as  $8 + 1$ ,  $50 + 1$ , and  $82 + 1$ . If we interpret this loosely bound neutron, as a valency neutron, the neutron numbers 8, 50, 82 represent greater stability than other neutron numbers.

It is apparent from the above conclusions that nuclear behaviour is often determined by the excess or deficiency of nucleons with respect to closed shells of nucleons corresponding to the magic numbers. It was, therefore, suggested that nucleons revolve inside the nucleus just as electrons revolve outside in specific permitted orbits.

#### Main Assumptions of the Shell Model

- (1) The protons and neutrons move in two separate systems of orbits round the centre of mass of all the nucleons.

- (2) The extra-nuclear electrons revolve in the Coulomb field of a relatively distant heavy nucleus. But the nucleons move in orbits around a common centre of gravity of all the constituents of the nucleus.
- (3) Each nucleon shell has a specific maximum capacity. When the shells are filled to capacity, they give rise to particular numbers (the magic numbers) characteristic of unusual stability.

#### Predictions of the Shell Model

The shell model is able to account for several nuclear phenomena in addition to magic numbers.

(i) It is observed that even-even nuclei are, in general, more stable than odd-odd nuclei. This is obvious from the shell model. According to Pauli's principle, a single energy sublevel can have a maximum of two nucleons (one with spin up and other with spin down). Therefore, in an even-even nucleus only *completed* sublevels are present which means greater stability. On the other hand, an odd-odd nucleus contains incomplete sublevels for both kinds of nucleon which means lesser stability.

(ii) The shell model is able to predict the total angular momenta of nuclei. In even-even nuclei, all the protons and neutrons should pair off so as to cancel out one another's spin and orbital angular momenta. Thus even-even nuclei ought to have zero nuclear angular momenta, as observed. In even-odd and odd-even nuclei, the half-integral spin of the single "extra" nucleon should be combined with the integral angular momentum of the rest of nucleus for a half-integral total angular momentum. Odd-odd nuclei each have an extra neutron and an extra proton whose half-integral spins should yield integral total angular momenta. Both these predictions are experimentally confirmed.

#### Nuclear Energy Level Scheme and Explanation of Magic Numbers.

To account for the observed magic numbers, Mayer and Jensen postulated a strong *nuclear spin-orbit interaction*. The magnitude of the spin-orbit interaction is such that the consequent splitting of energy levels into sublevels is many times larger than the analogous splitting of atomic energy levels. The nuclear spin-orbit splitting of a single-nucleon energy level is assumed to be *large* and also *inverted* (Fig. 17.14). We ascribe this behaviour to a nuclear interaction of the form

$$V_{SL}(\text{nucleus}) = -\frac{a_{SL}^2}{r} \frac{dV}{dr} \mathbf{S} \cdot \mathbf{L}$$

The minus sign accomplishes the required inversion of the split levels.

The constant  $a_{SL}^2$  produces the desired amount of energy splitting. The central-field function  $V(r)$  appears along with the orbital and spin angular momenta of the nucleon. The exact form of the potential-energy function is not critical, provided that it more or less resembles a square well.

The shell theory assumes that *LS coupling* holds only for the very lightest nuclei, in which the  $l$  values are necessarily small in their normal configurations. In this scheme, the intrinsic spin angular momenta  $S_i$  of the particles concerned are coupled together into a total spin momentum  $\mathbf{S}$ . The orbital angular momenta  $\mathbf{L}_i$  are separately coupled together into a total orbital momentum  $\mathbf{L}$ . Then  $\mathbf{S}$  and  $\mathbf{L}$  are coupled to form a total angular momentum  $\mathbf{J}$  of magnitude  $\sqrt{\mathbf{J}(\mathbf{J}+1)}\hbar$ .

After a transition region in which an intermediate coupling scheme holds, the heavier nuclei exhibit ***jj coupling***. In this case, the  $\mathbf{S}_i$  and  $\mathbf{L}_i$  of each particle are first coupled to form a  $\mathbf{J}_i$  for that particle of magnitude  $\sqrt{\mathbf{J}(\mathbf{J}+1)}\hbar$ . The various  $\mathbf{J}_i$  then couple together to form the total angular momentum  $\mathbf{J}$ . The *jj* coupling scheme holds for the great majority of nuclei.

Fig. 17.15 shows the nucleon energy levels according to the shell model. The levels are designated by a prefix equal to the total quantum number  $n$ , a letter that indicates  $l$  for each particle

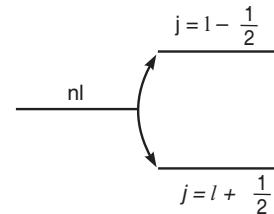


Fig. 17.14

in that level, and a subscript equal to  $j$ . The spin-orbit interaction splits each state of given  $j$  into  $2j + 1$  substates. The accumulated population of nucleons corresponds to a magic number at every one of the larger energy gaps. Hence shells are filled when there are 2, 8, 20, 28, 50, 82 and 126 neutrons or protons in a nucleus.

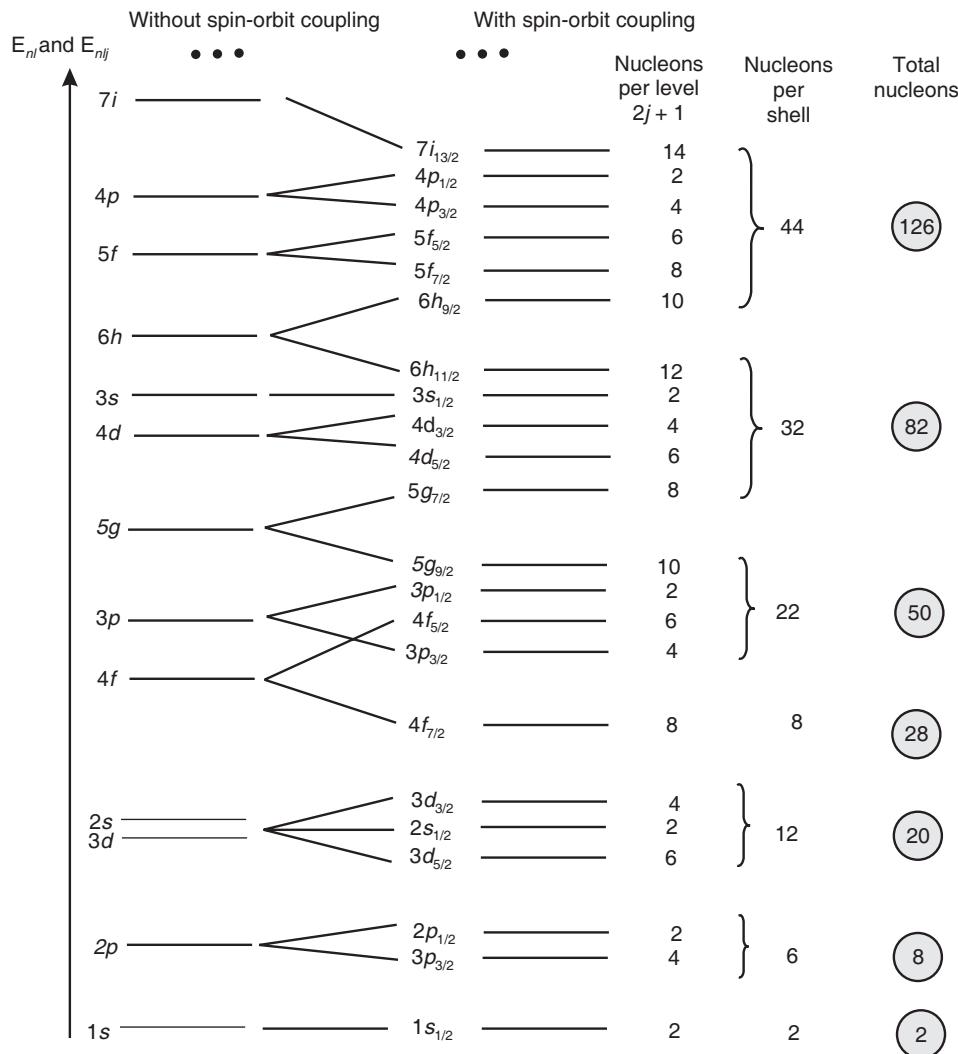


Fig. 17.15

### 17.13 THE COLLECTIVE MODEL

The collective model was proposed by A. Bohr, B. R. Mottleson and James Rainwater. The model combines the best features of the liquid drop model and the shell model. In this model, it is assumed that the particles within the nucleus exert a centrifugal pressure on the surface of the nucleus. That results in the permanent deformation to non-spherical shape. As a result, the surface may undergo periodic oscillations.

The particles within the nucleus move in a non-spherical potential. Thus the nuclear distortion reacts on the particles and modifies the independent particle aspect. Thus the nucleus is considered

as a shell structure capable of performing oscillations in shape and size. Thus the model can easily describe the drop like properties such as nuclear fission and at the same time it can retain the shell model characteristics. It is capable of explaining not only the large electric quadrupole moments but it can also predict the fine structure of nuclear level spectrum due to the energies associated with the vibrational and rotational motion of the core.

The total energy is expressed as

$$W = E_{rot} + E_{vib} + E_n$$

where  $E_{rot}$  is the energy due to rotational motion of the core,  $E_{vib}$  is the energy due to the vibrational coordinates and  $E_n$  is the energy due to the nucleonic coordinates. So the wave function is the product of three wave functions each containing the respective coordinates.

According to the collective model, all even-even nuclei such as  ${}_2\text{He}^4$ ,  ${}_8\text{O}^{16}$ ,  ${}_{20}\text{Ca}^{40}$ ,  ${}_{82}\text{Pb}^{208}$  will have spherical shapes and zero electric quadrupole moment, while the even-odd, odd-even or odd-odd nuclei will have non-spherical shapes and finite electric quadrupole moment. The nuclear energy levels predicted by the model agree closely with the ones given by  $\gamma$ -ray spectra of the nuclei.

## EXERCISE

1. The relation between nuclear radius and mass number is given by

$$(a) R = R_0 A^{\frac{1}{2}} \quad (b) R = R_0 A^{\frac{1}{3}} \quad (c) R = R_0 A^{\frac{1}{4}} \quad (d) R = (R_0 A)^{\frac{1}{3}}$$

(B.U. 2013)

2. The nuclear radius of a nucleus with mass number 27 is

$$(a) 3.9 \text{ fm} \quad (b) 9.3 \text{ fm} \quad (c) 3.1 \text{ fm} \quad (d) 1.3 \text{ fm}$$

(B.U. 2013)

3. Nuclear magneton is

$$(a) \frac{e\hbar}{2M_p} \quad (b) \frac{\hbar}{2eM_p} \quad (c) \frac{M_p}{2e\hbar} \quad (d) \frac{2e\hbar}{M_p}$$

(B.U. M.Sc. 2014)

4. Electric quadrupole moment is exhibited by

$$(a) spherical nuclei \quad (b) ellipsoidal nuclei \quad (c) both (a) and (b) \quad (d) neither (a) nor (b)$$

5. Exchange forces requires that they are mixture of

$$(a) Wigner and Majorana forces \quad (b) Wigner and Barlett forces \\ (c) Heisenberg and Majorana forces \quad (d) Heisenberg and Wigner forces$$

(B.U. M.Sc. 2015)

6. The binding forces between nucleons are due to the exchange of

$$(a) Proton \quad (b) \pi meson \quad (c) Neutrino \quad (d) Antineutrino$$

(B.U. M.Sc. 2014)

7. Nuclear forces are appreciable only when the distance between nucleons is of the order of

$$(a) 10^{-14} \text{ m} \quad (b) 10^{-15} \text{ m or less} \quad (c) 10^{-13} \text{ m} \quad (d) 10^{-10} \text{ m}$$

(B.U. M.Sc. 2015)

## GATE QUESTIONS

8. The volume of a nucleus in an atom is proportional to the

$$(a) mass number \quad (b) proton number \quad (c) neutron number \quad (d) electron number$$

9. The following gives a list of pairs containing (i) a nucleus (ii) one of its properties. Which pair is inappropriate?

$$(a) (i) {}_{10}^{20}N \text{ nucleus; } (ii) \text{ stable nucleus}$$

- (b) (i) A spheroidal nucleus; (ii) an electric quadrupole moment  
 (c) (i)  $^{16}_8 O$  nucleus; (ii) nuclear spin  $J = 1/2$   
 (d) (i)  $^{238}_{92} U$  nucleus; (ii) B.E. = 1785 MeV (app.)
10. As one moves along the line of stability from  $^{56}\text{Fe}$  to  $^{235}\text{U}$  nucleus, the nuclear B.E./particle decreases from about 8.8 MeV to 7.6 MeV. This trend is mainly due to the  
 (a) short range nature of nuclear forces      (b) long range nature of coulomb forces  
 (c) tensor nature of the nuclear forces      (d) spin dependence of the nuclear forces  
 [Hint. According to liquid drop model of the nucleus, the long-range nature of the coulomb force is responsible for the decrease in B.E. of heavy nuclei from  $^{56}\text{Fe}$  to  $^{235}\text{U}$ . ]
11. Consider a nucleus with  $N$  neutrons and  $Z$ -protons. If  $m_p$ ,  $m_n$  and B.E. represent the mass of  $p$ , the mass of  $n$  and the binding energy of the nucleus respectively and  $c$  is the velocity of light in free space, the mass of the nucleus is given by  
 (a)  $Zm_n + Zm_p$       (b)  $Nm_n + Zm_n$       (c)  $Nm_n + Zm_p + BE/c^2$  (d)  $Nm_p + Zm_n + BE/c^2$
12. Highest number of stable nuclei belongs to  
 (a) even  $Z$ , even  $N$       (b) odd  $Z$ , even  $Z$       (c) even  $Z$ , odd  $N$       (d) odd  $Z$ , odd  $Z$   
 [Ans. 1. (b), 2. (a), 3. (a), 4. (b), 5. (d), 6. (b), 7. (b), 8. (a), 9. (c), 10. (b), 11. (c), 12. (a)]
13. Explain any two properties of nucleus.      (B.U. 2015)
14. What is binding energy? Explain the stability of nucleus and binding energy.      (B.U. 2015)
15. Explain the terms:  
 (a) binding energy of the nucleus  
 (b) packing fraction  
 (c) mass defect      (B.U. 2013)
16. Explain the reasons for the absence of electrons inside the nucleus.      (B.U. 2013)
17. Explain the proton-neutron hypothesis.      (B.U. 2013)
18. Discuss the characteristics of nuclear forces.      (B.U. 2013)
19. Describe briefly Yukawa's meson theory of nuclear forces and estimate the approximate meson mass.  
 (Awadh 85)
20. Describe Hofstadter method for measuring the nuclear radius.      (Rohilkhand 2007)
21. Mention various methods for determining the size of the nucleus and describe any one in detail.  
 (Meerut 1980, 86)
22. Explain the liquid drop model of the nucleus with suitable theory.      (B.U. 2013)
23. Derive Weizsacker semi-empirical mass formula.      (B.U. 2011)
24. On the basis of liquid drop model, give a simple derivation of Weizsacker semi-empirical mass formula giving arguments for each term. What are the important conclusions drawn from this formula?  
 (Meerut 2003)
25. Explain the salient features of nuclear shell model. What are the magic numbers?      (B.U.M.Sc. 2006)
26. Write the experimental facts in support of shell model of nuclei. Explain how the shell model can be used to predict the angular momenta and magnetic moments of odd  $A$ -nuclei in the ground state.  
 (Meerut 2009, 2002)
27. Explain the nuclear collective model.      (B.U. 2015)

## DETECTORS OF NUCLEAR RADIATIONS

**After reading this chapter, you should be able to**

- ◆ Outline the working of different types of nuclear radiation detectors
- ◆ Discuss various interaction processes of all types of radioactive radiations with the matter
- ◆ Discuss the difference between ionisation chamber, proportional counter and Geiger-Muller counter
- ◆ Explain why GM counter is quenched after every event
- ◆ Distinguish between a cloud chamber and a bubble chamber
- ◆ Explain why the biological damage by neutrons to living cells/tissues is high.

## 18.1 INTRODUCTION

Most of the nuclear reactions are accompanied by the emission of charged particles like  $\alpha$ - particles, protons, electrons and radiations like  $\gamma$ -rays. In order to understand these particles and their interaction with atomic nuclei, precise information about their mass, momentum, energy, etc., are necessary. We shall describe in this chapter some of the common techniques employed for the detection of nuclear radiations and for analysing their energies.

Several nuclear radiation detectors depend for their operation on the *ionization* that is produced in them by the passage of charged particles. This group of detectors includes ionization chambers, proportional counters, G-M counters, semiconductor radiation detectors, cloud chambers and spark chambers. In other detectors, excitation and sometimes molecular dissociation also play important roles. These phenomena, in combination with ionization, bring about the luminescence in scintillation detectors and the latent images in photographic emulsions.

## 18.2 INTERACTION BETWEEN ENERGETIC PARTICLES AND MATTER

**(a) Heavy Charged Particles.** A heavy charged particle (like a proton,  $\alpha$ -particle or fission fragment) has a fairly definite *range* in a gas, liquid, or solid. The particle loses energy primarily by the excitation and ionization of atoms in its path. The energy loss occurs in a large number of small increments. The primary particle has such a large momentum that its direction is usually not seriously changed during the slowing process. Eventually it loses all its energy and comes to rest. The distance traversed is called the *range* of the particle.

The energy loss per unit length ( $-dE/dx$ ) is called the *stopping power*. The rate  $-dE/dx$  at which a heavy particle of charge  $ze$  and speed  $v$  loses energy in an absorber of atomic number  $Z$  which contains  $N$  atoms per unit volume whose average ionization energy is  $I$  is given by

$$-\frac{dE}{dx} = \frac{z^2 e^4 N Z}{4\pi \epsilon_0^2 m_0 v^2} \left[ \ln \left( \frac{2m_0 v^2}{I} \right) - \ln \left( 1 - \frac{v^2}{c^2} \right) - \frac{v^2}{c^2} \right] \quad \dots(1)$$

$m_0$  is the electron rest mass.

**(i) Stopping power  $S(E)$ .** This is defined as the amount of energy lost per unit length of the medium by incident particle in the given medium. Thus

$$S(E) = -\frac{dE}{dx} = N_{\text{ion}} \bar{I}$$

Here,

$N_{\text{ion}}$  = number of ion pairs produced per unit length.

$\bar{I}$  = Average ionisation energy.

**(ii) Range.** The range can be calculated by integrating Eq. (1) over the energies of the particle

$$R = \int_T^0 \left( -\frac{dE}{dx} \right)^{-1} dE \quad \dots(2)$$

**(b) Electrons.** Electrons interact through coulomb scattering from atomic electrons, just like heavy charged particles. There are however, a number of important differences:

- (1) Electrons travel at relativistic speeds.
- (2) Electrons will suffer large deflections in collisions with other electrons, and therefore will follow erratic paths. The range will therefore be very different from the length of the path that the electron follows.

- (3) Very energetic electrons ( $E > 1$  MeV) lose an appreciable fraction of their energies by producing continuous X-rays (also called *Bremsstrahlung*). The cross section for this process increases with increasing  $E$ .

**(c) The Absorption of  $\gamma$ -Rays.** The interaction of  $\gamma$ -rays with matter is markedly different from that of charged particles such as  $\alpha$  or  $\beta$  particles.  $\gamma$ -rays are extremely penetrating so that they are able to pass through considerable thicknesses of matter.  $\gamma$ -rays show the exponential absorption in matter. If radiation of intensity  $I$  is incident upon an absorbing layer of thickness  $dx$ , the amount of radiation absorbed  $dI$  is proportional both to  $dx$  and  $I$ . Hence,

$$dI = -\mu I dx \text{ or } I = I_0 e^{-\mu x}.$$

Here,  $\mu$  is a constant of proportionality which is a characteristic property of the medium, known as *linear absorption coefficient*. The *mass absorption coefficient*  $\mu_m$  may be obtained by dividing  $\mu$  by the density of the medium.  $\mu_m = \mu/\rho$ . The above relation gives the intensity (number of quanta per unit area per second) of the beam of initial intensity  $I_0$ , after traversing a thickness  $x$  of the homogeneous material. At low energies (0.1 MeV to 25 MeV) there are three important processes through which  $\gamma$  photons are absorbed by matter. [Fig. 18.1].

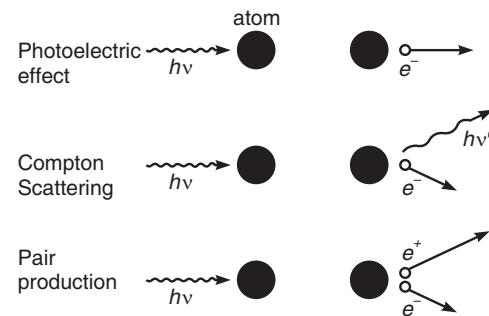


Fig. 18.1

**(i) Photoelectric effect.** In this process, the  $\gamma$ -rays knock out electrons from inside the atoms of the absorbing material. This results in the ionisation of the atoms and the emission of fluorescent radiations. Einstein's equation for those photo-electrons will be

$$\frac{1}{2}mv_k^2 = hv - W_k, \frac{1}{2}mv_L^2 = hv - W_L \text{ etc.}$$

where  $hv$  is the photon energy,  $v_k, v_L$  represent the velocities of the photoelectrons arising in the  $K, L\dots$  shells, and  $W_k, W_L\dots$  are the binding energies of  $K, L\dots$  shells.

**(ii) Pair Production.** In this process, the photon disappears and is converted to an electron-positron pair. This process can take place only when the photon energy exceeds  $2m_0c^2$ . The pair production process cannot occur in free space and usually takes place in the presence of a nuclear field. The nucleus recoils in this process conserving momentum. But the K.E., carried away by the nucleus is negligibly small due to its large mass compared with that of the electron. Photon energy, if any, in excess of  $2m_0c^2$  is shared as K.E. by the product particles.

**(iii) Compton effect.** It is an elastic scattering process in which the photon imparts energy to an electron. When a photon of energy  $hv$  strikes the perfectly free electron (at rest), the photon with diminished energy  $hv'$  is scattered at an angle  $\theta$  with the direction of incident photon and the electron recoils at an angle  $\phi$ . The energy absorbed by these Compton electrons is only a small fraction of the total energy of the incident  $\gamma$ -rays, unlike in the case of photoelectrons.

At low photon energies, the photoelectric effect is the chief mechanism of energy loss. The importance of the photoelectric effect decreases with increasing energy. In the lighter elements, Compton scattering becomes dominant at photon energies of a few tens of keV, whereas in the heavier ones this does not happen until photon energies of nearly 1 MeV are reached. Pair production becomes increasingly likely the more the photon energy exceeds the threshold of 1.02 MeV. Fig. 18.2 is a graph of the linear attenuation coefficient for photons in lead as a function of photon energy. The contributions to  $\mu$  of the photoelectric effect, Compton scattering, and pair production are shown.

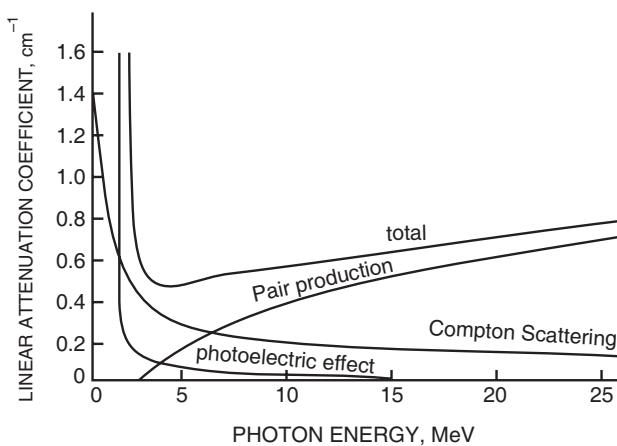


Fig. 18.2

### 18.3 IONIZATION CHAMBER

**Principle.** Charged sub-atomic particles can ionise gases. The number of ion-pairs produced gives us information not only on the nature of the incident particles, but even on their energy.

**Construction.** The ionisation chamber consists of a hollow metallic cylinder  $C$ , closed at both ends, (Fig. 18.3).

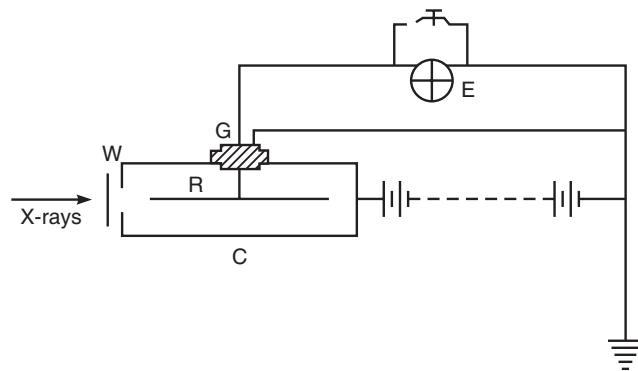


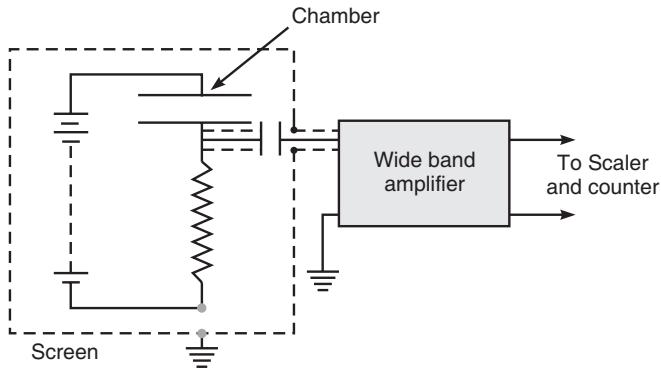
Fig. 18.3

- There is a window  $W$  at an end for the entry of ionising particles or radiation.
- A metal rod  $R$ , well insulated from the cylinder, is mounted coaxially within the cylinder.  $R$  is connected to a quadrant electrometer  $E$ . A p.d. of several hundred volts is maintained between  $C$  and  $R$ . An earthed guard ring  $G$  prevents leakage of charge from the cylinder to the rod.
- The chamber contains some gas like sulphur dioxide or methyl bromide.

**Working.** When a charged particle enters the chamber, it produces a large number of ion pairs in the enclosed gas, along its path.

- Positive ions move towards  $R$  and negative ions towards  $C$ .
- The quadrant electrometer  $E$  measures the rate of deposition of positive charges on  $R$ . The ionisation currents produced are quite small  $\approx 10^{-12} - 10^{-15}$  amperes. Special electrometers and D.C. amplifying devices are employed to measure such small currents.

If individual particles are to be counted, then the pulses of current produced are fed to a pulse amplifier, which is joined to the ionisation chamber by a coupling capacitor [Fig. 18.4].



**Fig. 18.4**

#### Uses

(1) Ionisation chambers have been used to study  $\alpha$ -particles,  $\beta$ -particles, protons, electrons and nuclei of lighter elements.

An Ionisation chamber is much less sensitive to  $\beta$ -particles (in comparison to  $\alpha$ -particles) because  $\beta$ -particles produce fewer pairs of ions in their passage through the chamber.

For detecting  $\gamma$ -rays, an ionisation chamber of thick wall made of high atomic-number material (Pt, Bi) is employed. The  $\gamma$ -rays impinging on the walls of the chamber eject high-speed electrons which produce ionisation in the gas.

(2) Ionisation chambers can also be used for measurements on X-rays and  $\gamma$ -rays.

(3) Ionisation chambers were extensively used in the early studies of cosmic rays phenomena.

(4) **Detection of neutron.** For neutron detection, the chamber is filled with boron trifluoride vapour (where the boron is enriched with  $B^{10}$ ) or the chamber walls are lined with a boron compound in the form of a paste.

**EXAMPLE 1.**  $\alpha$ -particles of energy 5 MeV pass through an ionisation chamber at the rate of 10 per second. Assuming all the energy is used in producing ion pairs, calculate the current produced. ( $35 \text{ eV}$  is required for producing an ion pair and  $e = 1.6 \times 10^{-19} \text{ C}$ ).

**SOL.** Energy of  $\alpha$ -particles =  $5 \times 10^6 \text{ eV}$ .

Energy required for producing one ion pair =  $35 \text{ eV}$

No. of ion pairs produced by one  $\alpha$ -particle

$$= \frac{5 \times 10^6}{35} = 1.429 \times 10^5$$

Since 10 particles enter the chamber in one second,

No. of ion pairs produced per second

$$= 1.429 \times 10^5 \times 10 = 1.429 \times 10^6$$

Charge on each ion =  $1.6 \times 10^{-19} \text{ C}$ .

$$\therefore \text{Current} = (1.429 \times 10^6) \times (1.6 \times 10^{-19}) \text{ C/s}$$

$$= 2.287 \times 10^{-13} \text{ A.}$$

**EXAMPLE 2.** An ionization chamber is connected to an electrometer of capacitance  $0.5 \text{ pF}$  and voltage sensitivity of 4 divisions per volt. A beam of  $\alpha$ -particles causes a deflection of 0.8 divisions. Calculate the number of ion pairs required and the energy of the  $\alpha$ -particles. Given that 1 ion pair requires energy of  $35 \text{ eV}$  and  $e = 1.6 \times 10^{-19} \text{ coulomb}$ .

**SOL.** Voltage sensitivity of electrometer = 4 divisions/volt.

$$\therefore \text{Voltage required to produce } \left. \begin{array}{l} \text{a deflection of 0.8 divisions} \\ \text{volt} \end{array} \right\} = \frac{0.8}{4} \text{ volt} = 0.2 \text{ volt}$$

$$\begin{aligned} Q &= CV = (0.5 \times 10^{-12}) \times 0.2 \quad [\text{since } C = 0.5 \text{ pF} = 0.5 \times 10^{-12} \text{ F}] \\ &= 10^{-13} \text{ C.} \end{aligned}$$

$$\therefore \text{No. of ion pairs required} = \frac{10^{-13}}{1.6 \times 10^{-19}} = 6.25 \times 10^5$$

1 ion pair requires 35 eV.

$$\begin{aligned} \therefore \text{Total energy required} &= 35 \times (6.25 \times 10^5) \text{ eV} \\ &= 21.88 \text{ MeV.} \end{aligned}$$

## 18.4 SOLID-STATE DETECTORS

**Construction.** Fig. 18.5 shows a *p-n* junction, used as a particle detector. It consists of a *p-n* junction between *p*-type and *n*-type silicon. Contact is made with the *n*-type silicon layer by a thin evaporated film of gold. In order to minimise the current flowing in the detector, when no radiation is striking it, a reverse biased diode is always used. The positive (reverse) bias applied to the gold film will push all the positive charge carriers away from the junction and produce a *depletion layer*, indicated in the figure. The depletion layer contains almost no carriers of either sign.

**Working.** When an energetic charged particle travels through the depletion layer, its interaction with the electrons in the crystal, produces electron-hole pairs. There is an electron-hole pair for every 3.5 eV (in Si) of energy lost by the charged particle. The electrons and holes are swept away by the applied electric field and registered as a voltage pulse over the resistor *R*.

### Advantage

- The number of charge carrier pairs produced in a semiconductor material is approximately 10 times as large as the number of ion pairs produced in a gas ion chamber *i.e.*, the energy extended per pair is about 3.5 eV in silicon, compared to about 30 eV for gases. The voltage pulse will therefore be about 10 times larger. Hence this detector has much better *energy resolution* than other radiation detectors.
- In solid state detectors for charged particles, silicon has been used most because of its low intrinsic conductivity. This means that the detector can be operated at room temperature without excessive leakage current.
- For gamma ray work, germanium detectors are much better than silicon because of the larger density of germanium.

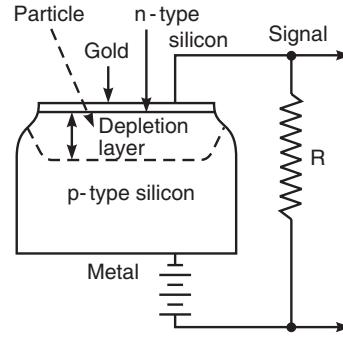


Fig. 18.5

## 18.5 PROPORTIONAL COUNTER

**Principle.** When the potential difference across the two electrodes of an ionisation chamber is increased beyond the saturation region, the ions gain sufficient energies to produce secondary ionisation due to collision with gas atoms in the chamber. This results in greater *ionisation current*. In a certain range of applied voltages (See Fig. 18.6), the pulse size is proportional to the original number of ion pairs, and the device is called a *proportional counter*.

The main regions used for measurements are:

- (1) The ionization chamber region *AB*.
- (2) The proportional counter region *CD*.
- (3) The Geiger-Muller region *EF*.

After the point *F*, the tube becomes a simple discharge tube in which the current is produced even after the ionization event has ceased.

- Since the central wire is very thin and the p.d. fairly large, the electric field  $E = dV/dr$  at a distance  $r$  from the centre is very high. If  $b$  is the radius of the cylinder and  $a$  the radius of the wire, the radial field  $E$  at a distance  $r$  from the centre is given by

$$E = \frac{V}{r \log_e (b/a)}.$$

Here  $V$  is the positive voltage of the central wire relative to the outer cylinder. Thus, in a proportional counter, the field strength near the wire is very great. Hence electrons travelling towards the wire are rapidly accelerated when near it, and produce additional electrons in that region due to the phenomena of ionization by collision. This process is called the *gas multiplication*.

**Construction.** Fig. 18.7 shows a proportional counter.

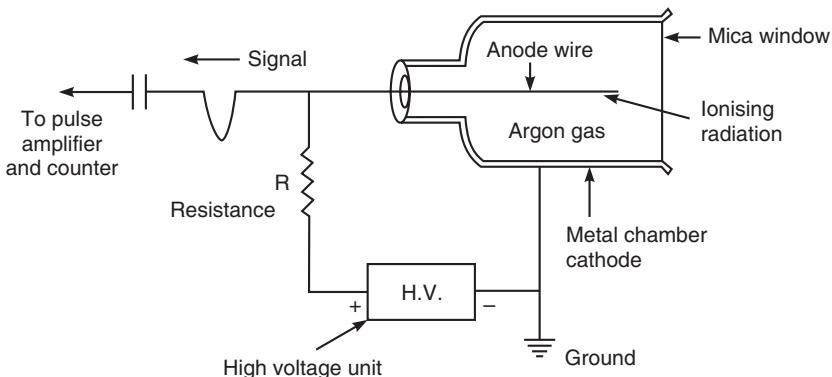


Fig. 18.7

- It is a metal chamber filled with a gas having a thin wire (diameter  $\sim 0.1$  mm) axially along the centre. This wire is made anode, by connecting it to a power supply. The metal chamber acts as the cathode.
- A voltage of a few hundred volts which does not cause a discharge is applied between the wire anode and the metal chamber cathode.

#### Definitions

**Specific ionisation.** It is defined as the number of ion pairs produced per unit distance between the electrodes.

**Gas multiplication factor.** It is defined as the total number of ion pairs produced by a single primary ion.

**Working.** Consider the entry of an ionising particle into the chamber through the thin mica “end window”. Electrons produced are attracted toward the central wire anode. The electric field in the vicinity of the fine wire is very large compared to the other regions in the chamber. So an electron near the wire acquires enough energy between two successive collisions with the gas atoms

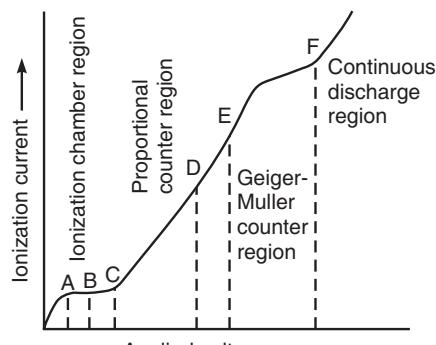


Fig. 18.6

to enable it to ionise gas atoms. This produces additional ion pairs (electrons and positive ions). This process is called *gas amplification*. The number of ions increases by a factor  $\sim 10^4$ .

- The output pulse of the current is *proportional* to the number of initial ion pairs produced by the ionising particle entering the chamber. That is, the output signal is proportional to the energy of the incident particle.
- Because of the IR drop across the resistance, the current pulse produces a voltage pulse, which is amplified and recorded.

**EXAMPLE.** It is required to operate a proportional counter with a maximum radial field of  $10^7 \text{ Vm}^{-1}$ . What is the applied voltage required if the radii of the wire and tube are 0.002 cm and 1 cm respectively?

$$\text{SOL. Radial field } E = \frac{V}{r \log_e(b/a)}$$

$$\text{Radial field at the wire surface } E = \frac{V}{a \log_e(b/a)}$$

$$\therefore 10^7 = \frac{V}{(2 \times 10^{-5}) 2.302 \log_{10}(10^{-2} / 2 \times 10^{-5})}$$

or

$$V = 1242 \text{ volts.}$$

## 18.6 GEIGER-MULLER COUNTER

**Construction.** It consists of a metal chamber  $C$  containing air or some other gas at a pressure of about 10 cm of Hg (Fig. 18.8).

- A fine tungsten wire ( $W$ ) is stretched along the axis of the tube and is insulated from it by ebonite plugs  $EE$  (Fig. 29.8).
- The wire is connected to the positive terminal of a high tension battery (about 1000 to 3000 volts) through a high resistance  $R$  (about 100 megohms) and the negative terminal is connected to the chamber  $C$ . The D.C. Voltage is kept slightly less than that which will cause a discharge between the electrodes.

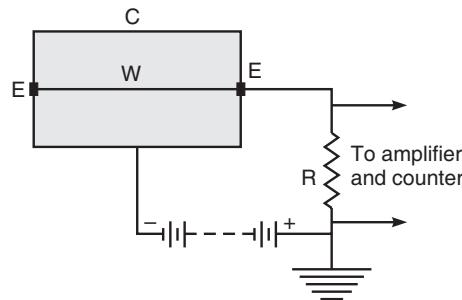


Fig. 18.8

**Working.** When an ionizing particle (say an  $\alpha$  particle) enters the counter, ionisation takes place and a few ions are produced.

- If the applied P.D. is strong enough, these ions are multiplied by further collisions. An avalanche of electrons moves towards the central wire. This is equivalent to a small current impulse which flows through the resistance  $R$ . The critical potential is lowered momentarily, causing a sudden discharge through the resistance  $R$ . The p.d. thus developed across  $R$  is amplified by vacuum tube circuits and is made to operate a mechanical counter. In this way, single particles can be registered.
- The sudden pulse of discharge sweeps away the ions from the chamber and the counter is ready to register the arrival of the next particle.

Fig. 18.9 shows the voltage characteristics of a G.M counter. This is a plot of the counting rate against the counter potential with a radioactive source placed near the counter.

- There is a threshold below which the tube does not work. This can be several hundred volts.
- As the applied potential is increased, the counting begins and rises rapidly to a flat portion of the curve called the *plateau*. This is the region of the counter operation. Here the counting rate is, more or less, independent of small changes in p.d. across the tube.
- Beyond the plateau, the applied electric field is so high that a continuous discharge takes place in the tube. The count rate increases very rapidly. It does not require any ionizing event for this to happen. So the tube must not be used in this region.

#### Efficiency of the counter

The *efficiency of the counter* is defined as the ratio of the observed counts/sec. to the number of ionizing particles entering the counter per second. Counting efficiency is defined as the ability of its counting, if at least one ion pair is produced in it.

$$\text{Counting efficiency} = \epsilon = 1 - e^{-slp}$$

- $s$  = specific ionization at one atmosphere;
- $p$  = pressure in atmospheres;
- $l$  = path length of the ionization particle in the counter.

Fig. 18.10 shows the efficiency  $\epsilon$  of a GM counter, as a function of pressure for air and hydrogen.

#### Quenching in a GM counter

The process of removing all the ions from the chamber due to continuous discharge and making it ready for fresh event is called *quenching*. This is achieved by using quenching agents like alcohol or halogens.

**Quenching Process.** When a G.M. tube operates in Geiger region, the secondary electrons increase the current pulse by further ionisation of gas molecules. The object of counter is to produce a single pulse due to entry of a single particle. The tube should not then give any succeeding spurious pulses but should recover as quickly as possible to be in the state to record the next entering particle. Therefore, it is desirable to prohibit the production of spurious pulses following the main required pulse due to a single particle entry. The process of prohibiting the undesirable secondary pulses is called the *quenching*.

**Self-quenching of GM-counter.** A typical GM-counter contains 90% Ar and 10% ethyl alcohol by weight. The ionisation potential of alcohol (11.3 eV) is lower than that of argon (15.7 volt). As a result, the inert gas ions get neutralized by transferring their charge to alcohol ions during their long path towards cathode. The alcohol ions produced, capture electrons from the cathode and are neutralized. Hence there is no multiple pulsing and the discharge is quenched soon (fraction of a milli-second) after the initial ionisation.

#### Resolving Time of a GM-Counter

Fig. 18.11 gives the three time-intervals—*dead time*, *recovery time* and *resolving time*—of operation of a GM-counter.

(i) **Dead time.** The time, during which the counter is incapable of responding to a second ionizing event, is known as *dead time* ( $\sim 200 \mu\text{sec}$ ). It is called *dead time* because during this period the counter is insensitive (dead) to further pulses. After the initial ionisation, the cloud of slow

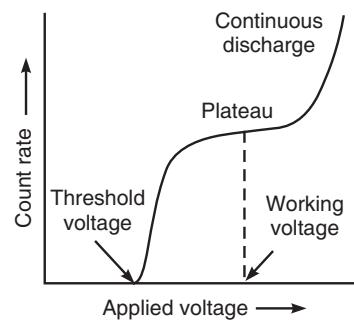


Fig. 18.9

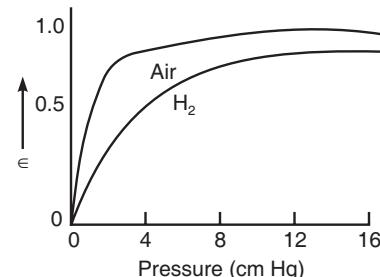


Fig. 18.10

moving massive positive ions formed over the anode wire effectively lowers its potential relative to cathode. Thus the G.M counter is dead for about 200  $\mu\text{s}$  after each ionising event.

(ii) **Recovery time.** As the positive ion-sheath moves further toward the cathode, the counter sensitivity gradually returns to its original value and the second pulse can now be recorded. This time, during which pulses of reduced size are produced, is known as *recovery time*.

(iii) **Resolving time.** The dead time plus recovery time corresponds to the time which positive ion sheath takes to reach the cathode, and is known as the *resolving time* of the counter. This is the time between just recorded pulses.

**Expression for resolving time.** Assume that a counter with a resolving time  $\tau$  responds at a rate  $n$  counts per unit time when exposed to  $N$  particles per unit time. In unit time, the total insensitive time is  $n\tau$ . The number of counts missed is  $Nn\tau$ .

The number of counts missed is the error in counting.

$$Nn\tau = N - n \quad \dots(1)$$

$$\therefore N = n/(1 - n\tau) \quad \dots(2)$$

From Eq. (2), we can find the *true* counting rate  $N$  from the *observed* counting rate  $n$ , if  $\tau$  is known.

**Determination of  $\tau$ .** By measuring separately the counting rates  $n'_1$  and  $n'_2$  with two radioactive sources 1 and 2 of nearly equal strength,  $\tau$  can be determined. Let  $n'_t$  denote the total counting rate when both sources are placed in the same position as before and  $n_b$  be the background rate (*i.e.*, with no source).

$$\therefore n_1 = n'_1 - n_b; \quad n_2 = n'_2 - n_b; \quad n_t = n'_t - n_b \quad \dots(3)$$

Here,  $n_1, n_2, n_t$  are the above counting rates corrected for background.

If  $N_1, N_2$  and  $N_t$  be the true rates of arrival of particles in the counter in the above three situations, we have, using Eq. (2),

$$N_1 = \frac{n_1}{1 - n_1\tau}; \quad N_2 = \frac{n_2}{1 - n_2\tau}; \quad N_t = \frac{n_t}{1 - n_t\tau} \quad \dots(4)$$

Also,

$$N_t = N_1 + N_2 \quad \dots(5)$$

Assuming the values of terms involving  $\tau^2$  to be negligible, we have, solving for  $\tau$

$$\tau = \frac{n_1 + n_2 - n_t}{2n_1n_2} \quad \dots(6)$$

### Advantages

(1) The counter set-up is portable (with the transistorised electronics) and serves best for mineral prospecting, apart from its several other applications in cosmic ray work.

(2) The pulse height is constant over a range of applied voltages, as in Fig. 18.9. So the power supply does not have to be precisely regulated as it does for a proportional counter. Also, the pulses are several volts in height, which makes amplifiers unnecessary.

### Disadvantages

(1) It is insensitive for a period of 200 to 400  $\mu\text{s}$  following each pulse. This prevents its use at very high counting rates.

(2) It cannot provide information about the particle or photon causing a pulse.

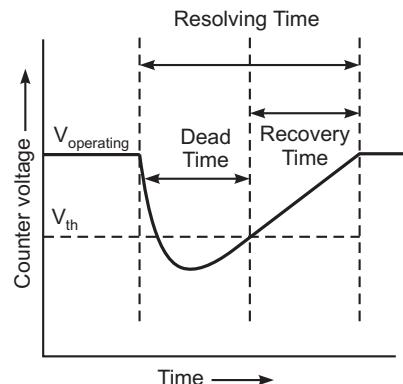


Fig. 18.11

**EXAMPLE 1.** A self-quenched G-M counter operates at 1000 volts and has a wire diameter of 0.2 mm. The radius of the cathode is 2 cm and the tube has a guaranteed lifetime of  $10^9$  counts. What is the maximum radial field and how long will the counter last if it is used on an average for 30 hours per week at 3000 counts per minute? Consider 50 weeks to a year.

**SOL.** The radial field at the central wire is

$$E_{\max} = \frac{V}{r \log_e(b/a)} = \frac{1000}{0.0001 \times 2.3026 \log_{10} \left( \frac{2 \times 10^{-2}}{10^{-4}} \right)}$$

$$= 1.89 \times 10^6 \text{ volts/metre.}$$

If the lifetime of the tube is  $N$  years, the total number of counts recorded will be

$$N \times 50 \times 30 \times 60 \times 3000 = 2.7 \times 10^8 N$$

$$\therefore 2.7 \times 10^8 \times N = 10^9$$

or  $N = 3.7$  years.

**EXAMPLE 2.** The wire in a GM counter collects  $10^{10}$  electrons per discharge. If the count rate is 1000/minute, calculate the average current in the circuit. (charge on electron is  $1.6 \times 10^{-19}$  C).

(Andhra University, 2010).

**SOL.** Counting rate = 1000 counts/min.

The wire collects  $10^{10}$  electrons per discharge.

$\therefore$  Total number of electrons collected in one min. =  $n = 1000 \times 10^{10} = 10^{13}$

$$\text{Charge/min} = n e = 10^{13} \times (1.6 \times 10^{-19}) = 1.6 \times 10^{-6} \text{ coul/min.}$$

$$\text{Charge/second} = \frac{1.6 \times 10^{-6}}{60} = 2.66 \times 10^{-8} \text{ Amp.}$$

$$\therefore \text{Average current} = \text{charge/second} = 2.66 \times 10^{-8} \text{ Amp.}$$

## 18.7 THE WILSON CLOUD CHAMBER

**Principle.** If there is a sudden expansion of saturated vapour in a chamber, supercooling of the vapour occurs. Tiny droplets will be formed by condensation over the dust particles present in the chamber. If, therefore, we have completely dust-free and saturated air, and if it is suddenly allowed to expand and thereby cool, condensation will not take place. But if ions are available in the chamber during the expansion, they serve as nuclei for condensation. Hence, if an ionising particle passes through the chamber during an expansion, ions are produced along its path and droplets condense on these ions. Hence the "track" of the particle becomes visible.

**Description.** The apparatus consists of a large cylindrical chamber  $A$ , with walls and ceiling made of glass (Fig. 18.12).

- It contains dust-free air saturated with water vapour.
- $P$  is a piston working inside the chamber. When the piston moves down rapidly, adiabatic expansion of the air inside the chamber takes place.
- The piston is connected to a large evacuated vessel  $F$  through a valve  $V$ . When the valve is opened, the air under the piston rushes into the evacuated vessel  $F$ , thereby causing the piston to drop suddenly.
- The wooden blocks  $WW$  reduce the air space inside the piston.
- Water at the bottom of the apparatus ensures saturation in the chamber.
- The expansion ratio can be adjusted by altering the height of the piston.



Wilson Cloud Chamber

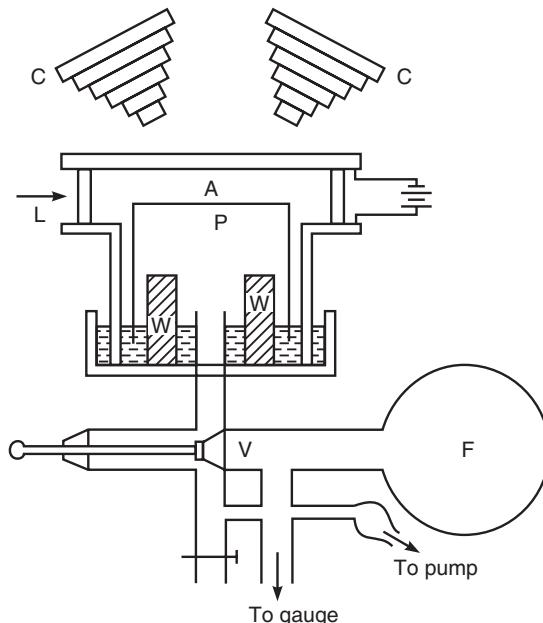


Fig. 18.12

**Working.** As soon as the gas in the expansion chamber is subjected to sudden expansion, the ionizing particles are shot into the chamber through a side window. A large number of extremely fine droplets are formed on all the ions produced by the ionizing particles. These droplets form a track of the moving ionizing particles. At this stage, the expansion chamber is profusely illuminated by a powerful beam of light  $L$ . Two cameras  $CC$  are used to photograph the tracks.

- The process of expansion, shooting of the ionising particles into the expansion chamber, illuminating the chamber and clicking the camera must all be carried out in rapid succession in order to get satisfactory results.

**Results.** The ionising agent can be easily identified from its path in the cloud chamber.

- $\alpha$ -particles, being comparatively massive, go straight and their paths are *thick, straight and sharply defined*.
- $\beta$ -particles being lighter, are easily deflected by collision and their paths are *thin and crooked*.
- The cloud chamber has led to the discovery of many elementary particles like positron, meson, etc.

**Advantages.** (1) Cloud chambers can be used to study the variation of specific ionisation along the track of a charged particle and the range of such particles.

(2) The sign of the electric charge and the momentum  $p$  of the particle can be determined if the chamber is placed in a strong magnetic field. Let a particle of mass  $m$  and charge  $q$  move with a velocity  $v$  perpendicular to the direction of the magnetic field of flux density  $B$ . The particle will be forced by the field to follow a circular path of radius  $R$ . The magnetic force  $Bqv$  is exactly balanced by the centrifugal force  $mv^2/R$ .

$$\text{Thus } Bqv = Mv^2/R \text{ or } mv = BRq.$$

The K.E. of the particle can be calculated, if the rest mass energy  $m_0c^2$  of the particle is known, by the relation,

$$K.E. = E_k = \sqrt{[p^2 c^2 + (m_0 c^2)^2]} - m_0 c^2$$

**Limitations.** (i) One is not always sure of the sense of track photographed.

(ii) The range of the particle may exceed the dimensions of the chamber so that the whole track is not photographed.

(iii) There remains a certain amount of uncertainty about the nature of the nuclei constituting the arms of the forked tracks.

### 18.8 BUBBLE CHAMBER

**Principle.** We know that normally the liquid boils with the evolution of bubbles of vapour at the boiling point. If the liquid is heated under a high pressure to a temperature well above its normal boiling point, a sudden release of pressure will leave the liquid in a *superheated state*. If an ionising particle passes through the liquid within a few milliseconds after the pressure is released, the ions left in the track of a particle act as condensation centres for the formation of vapour bubbles. The vapour bubbles grow at a rapid rate and attain a visible size in a time of the order of 10 to  $100\mu\text{s}$ . Thus in a bubble chamber, a vapour bubble forms in a superheated liquid, whereas in a cloud chamber, a liquid drop forms in a supersaturated vapour. Thus an ionising particle passing through the superheated liquid leaves in its wake a trail of bubbles which can be photographed.

**Construction.** Fig. 18.13 shows a liquid hydrogen bubble chamber, operating at a temperature of  $27\text{ K}$ .

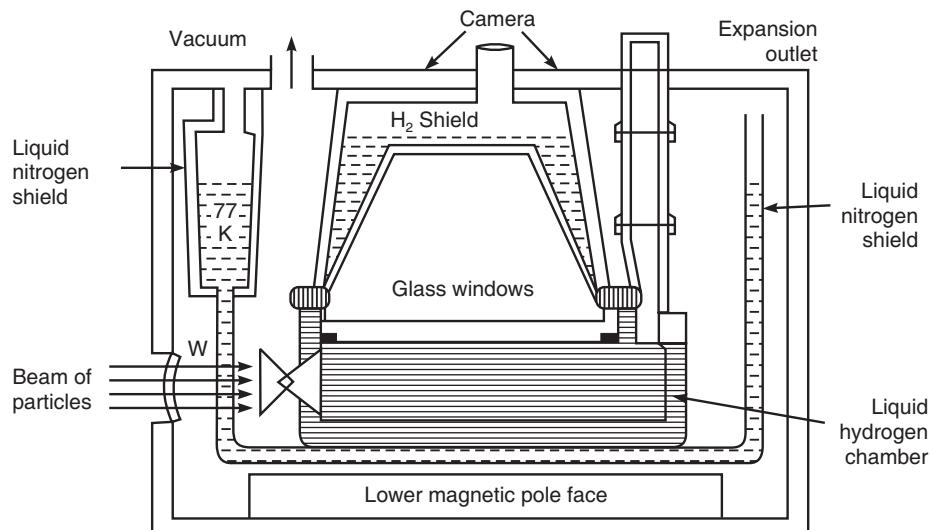


Fig. 18.13

A box of thick glass walls is filled with liquid hydrogen and connected to the expansion pressure system. To maintain the chamber at constant temperature, it is surrounded by liquid nitrogen and liquid hydrogen shields.

**Working.** High energy particles are allowed to enter the chamber from the side window  $W$ . A sudden release of pressure from the expansion valve is followed by light flash and camera takes the stereoscopic view of the chamber.

The incoming beam triggers the chamber. The charge of the tracks can be identified by the direction of their curvature in the magnetic field applied over the bubble chamber. From the curvature and length of the track, the momentum and energy of the particle can be found.

- The bubble chamber is used to *study particle interaction* and to *detect very high energy particles*.

**Advantages.** (1) The density of a liquid is very large when compared to that of a gas or even high pressure. Hence the chances of collision of a high energy particle with a molecule of the liquid are very much greater. Consequently there is a greater chance of their track being recorded. So the chances of recording events like cosmic ray phenomena are improved when compared with cloud chambers.

(2) The bubbles grow rapidly and as a result the tracks are not likely to get distorted due to convection currents in the liquid.

(3) The bubble chamber is sensitive even to particles of low ionising power.

### 18.9 NUCLEAR EMULSIONS

**Principle.** Charged particles affect a photographic plate. A heavy charged particle traversing a photographic emulsion produces a latent image of its track. The track is revealed when the plate is developed.

**Nuclear and optical emulsions.** Ordinary optical photographic emulsions are not suitable for quantitative studies with nuclear radiations. The sensitivity of such emulsions is low. Further, the tracks due to charged particles have non-clear range because the developed crystal grains are large and widely spaced.

- The composition of the emulsion was changed so as to make it more suitable for the study of various ionising particles, such as  $\alpha$ -particles, protons, mesons and even electrons.
- The nuclear emulsions differ from the optical emulsions in that they have considerably higher silver halide content and smaller grain size.
- In nuclear emulsion, the thickness is greater than that of optical emulsion (Table 18.1).

The smaller the grain size, the more sensitive is the emulsion to ionising particles. Thus different commercially available emulsions, differing chiefly in grain size, can be used to discriminate between different particles.

**Table 18.1. Nuclear and Optical Emulsions**

Property	Optical Emulsion	Nuclear Emulsion
AgBr : Gelatin (mass)	47 : 53	80 : 20
AgBr : Gelatin (volume)	15 : 85	45 : 55
Grain size (micron)	1–3.5	0.1–0.6
Thickness (,,)	2–3	25–2000
Sensitivity to light	Very high	Poor
Response to $\alpha$ particle	Dense blackening	Individual tracks
" " $\beta$ "	Moderate	Faint fog
" " $\gamma$ "	Faint	Almost none

#### Determination of the charge, mass and initial kinetic energy of the charged particle.

We can determine the charge, mass and incident K.E. of a charged particle brought to rest in a nuclear emulsion. For this purpose, the range of the particle, the grain density and its variation along the track and the small scattering angle due to collisions of the particle with nuclei in the emulsion are all accurately measured.

### Main relationships

(i) The number of grains deposited per micron  $dn/dx$  of track is closely related to the energy loss per unit length of track  $\frac{dT}{dx}$  keV/micron. These two quantities are related by the empirical relation,

$$\frac{dn}{dx} = a \left\{ 1 - e^{bz(dT/dx)^{1/2} - c^{1/2}} \right\}, \quad \dots (1)$$

Here  $a$ ,  $b$  and  $c$  are empirical constants and  $z$  is the integer characterizing the charge of ionizing particle.

(ii) If  $m$  (the mass in units of proton mass) and  $z$  of a particle are known, its kinetic energy  $T$  (MeV) is obtained from its total range  $R$  (microns) by the empirical relation,

$$T = K z^{2n} m^{1-n} R^n, \quad \dots (2)$$

Here  $K$  and  $n$  are empirical constants.

(iii) The total number of silver grains deposited ( $N$ ) and the total particle range ( $R$ ) are related by an **empirical relation**,

$$N = K' z^{2n'} m^{1-n'} R^{n'}, \quad \dots (3)$$

Here,  $K'$  and  $n'$  are empirical constants.

**Advantages.** (1) The emulsion is relatively light and cheap. Because of their lightness, they can be sent in balloons, spaceships etc., for high altitude cosmic ray experiments. The cosmic ray events once recorded can be studied by developing the exposed plates conveniently in the laboratory. Emulsions were widely employed in cosmic ray studies and led to the discovery of the  $\pi$  and  $K$  mesons.

(2) The high density of the emulsion gives it a stopping power about a thousand times that of standard air. Unstable high energy particles are brought to rest in the emulsions and their decay schemes can thus be studied.

(3) The emulsion is continuously sensitive and is consequently always available to record an event. But the cloud chamber is sensitive only for a fraction of a second after an expansion and remains ineffective for several seconds between successive expansions.

**Limitations.** (1) The main drawback of nuclear emulsions is that their sensitivity and thickness are affected by temperature, humidity, age of the emulsions before development and the conditions under which they are developed. The scanning of the plates and the analysis of the tracks obtained are also laborious, when done manually.

(2) As compared with cloud chamber, photographic emulsions have the following drawback. It is very difficult to determine the sign of the charge and the momentum of a particle from observations of the curvature of the path in a strong magnetic field. The actual path-length in the emulsion is small compared to that in a cloud chamber.

## 18.10 SCINTILLATION COUNTER

**Principle.** When high energy radiation falls on a fluorescent material, it produces flashes of light called '*scintillations*'. The energy of the particle is converted to light.

Fig. 18.14 shows the main parts of scintillation counter.

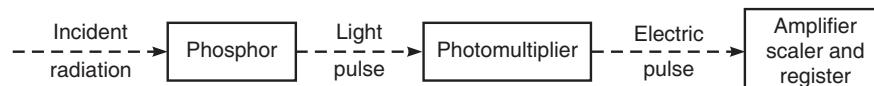


Fig. 18.14

The atoms of the phosphor are excited or ionised by the energy loss of an impinging  $\alpha$ ,  $\beta$  or  $\gamma$  ray. When the atoms return to their ground states, photons are emitted, in the blue and ultraviolet regions of the optical spectrum.

The phosphor is optically coupled to the envelope of a photomultiplier tube. The output pulse from the photomultiplier is fed to a pulse amplifier followed by a scaler circuit.

**Construction.** Fig. 18.15 shows the diagram of a scintillation counter.

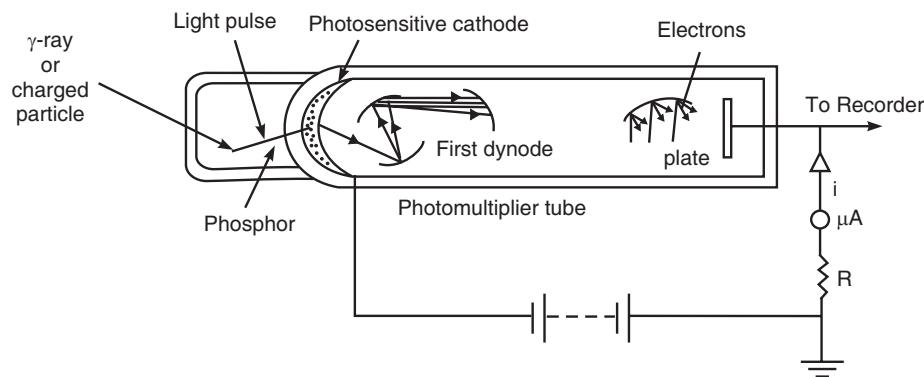


Fig. 18.15

### Working

- (i) The incident radiation is absorbed by the ‘phosphor’ (or scintillator). The atoms (or molecules) are excited.
- (ii) When the atoms (or molecules) return to their ground states, photons are emitted.
- (iii) The phosphor is optically coupled to the envelope of a photomultiplier tube. The photons strike the photocathode, causing the ejection of photo-electrons.
- (iv) As these photo-electrons leave the photocathode, they are directed by a focusing electrode to the first multiplier electrode or *dynode*. This electrode has the property of emitting three, four or five electrons for every single electron which strikes its surface. There may be from 10 to 14 such multiplier stages in a given tube. Hence from the emission of one single electron from the cathode, a burst of one million electrons may impinge on the final stage in the tube (the anode).

If  $m$  is the electron multiplication factor of a single dynode stage and there are  $n$  stages, then total amplification is  $m^n$ .

- (v) The output pulse from the photomultiplier is fed to a pulse amplifier followed by a scaler circuit.

The following scintillators are generally used for the detection of particles and photons.

*Sodium Iodide (Thallium Activated).* This is the most commonly used phosphor and is largely used in the study of  $\gamma$ -rays.

Sodium iodide crystals are usually doped with a small amount of thallium and denoted by NaI(Tl). The Tl atoms act as luminescence centres. NaI(Tl) crystals are dense enough to have good efficiency of  $\gamma$ -ray detection.

**Table 18.2. Common scintillators and their uses**

		Name	Form	Wavelength of Maximum Emission	General Remarks
1.	(a)	<b>Inorganic scintillators</b>			
	(a)	Nal (Tl)	Crystal	4100 Å	mainly used to detect $\gamma$ -rays, hygroscopic
	(b)	Cs(Tl)	Crystal	white	for $\gamma$ -rays, nonhygroscopic
	(c)	ZnS(Ag)	Powder	4500 Å	for $\alpha$ -rays
2.	(d)	LiI (Eu)	Crystal	—	used for neutron detection
		<b>Organic crystal</b>			
	(a)	Anthracene	Crystal	4470 Å	for $\beta$ detection
	(b)	Stibene	Crystal	4100 Å	for $\beta$ detection because it gives higher yield of photons [Anthracene is better]
3.		<b>Organic sulphide</b>			
	(a)	p-terphenyl in toluene	Solution	4500 Å	for $\beta$ -particle
	(b)	p-terphenyl in polystyrene	Plastic	4450 Å	for $\beta$ -particle
4.		<b>Xe-gas</b>	Gas	ultraviolet	for $\gamma$ -rays

### Advantages

- (1) The height of the pulse is proportional to the energy of the incident  $\gamma$ -radiations.
- (2) In a solid scintillator, the number of atoms per unit volume is more. Therefore a stronger pulse is observed in comparison to a gas scintillator.
- (3) The dead time of the counter is very small and hence a high counting rate is possible.
- (4) The pulse duration is extremely short and so the resolving power is high.
- (5) If a large scintillator is used, the scattered  $\gamma$ -rays are also taken into account thereby giving rise to an increased photo peak efficiency.

### Disadvantages

- (1) The light output decreases with increase in energy of radiation. It also varies with the type of radiation.
- (2) The overall efficiency is low because of the complex nature of the phenomena—namely the conversion of radiation into light and then re-conversion into electrons.

## 18.11 RADIATION HAZARDS

There are two major sources of radiation contaminating the biosphere.

- (i) People receive some natural or background radiation exposure each day from the sun, radioactive elements in the soil and rocks, household appliances (like television sets and microwave ovens), and medical and dental x-rays. These levels of natural and background radiation is normal. At sea level, the average radiation due to natural background is about 100 millirad per year.

(ii) The second source of environmental contamination is man-made. This includes nuclear reactors, materials released through accidents or through the testing of nuclear devices underground or in the atmosphere. The radioactivity spread as a result of the latter is much higher than that due to the former source.

Radioactive materials-if handled improperly-or radiation accidentally released into the environment, can be dangerous because of the harmful effects of certain types of radiation on the body. The longer a person is exposed to radiation and the closer the person is to the radiation, the greater the risk.

#### 18.11.1. Biological Effects of Nuclear Radiations

The fact that nuclear radiations such as  $\gamma$ -rays,  $\alpha$ -particles, and neutrons can cause damage to the human body is well known.

The harmful effects of nuclear radiations appear to be due to the ionization or excitation of atoms in living cells by the Compton effect, bremsstrahlung, photoelectric effect, and so on. Some of the cell constituents are altered or destroyed by ionisation and some of the products formed may act as poisons. Examples of damages are the breaking up of chromosomes, swelling of the nucleus of a cell or of the entire cell; changes in the permeability of cell membrane, and destruction of cells.

The biological effects can be divided into three groups:

- (i) Short-term recoverable effects;
- (ii) Long-term irrecoverable effects; and
- (iii) The genetic effect.

Groups (i) and (ii) are limited to the individuals who have actually received the radiation while the effects in group (iii) only appear in later generations.

When excessive doses are absorbed, the first noticeable disorder is a drop in the white blood cell count, which becomes evident in the first few hours after exposure. This is followed by a sickness pattern of diarrhoea, vomiting and fever which is now called radiation sickness. Recovery is possible from small acute doses. Large doses are lethal within a few weeks. Smaller doses produce short term effects, such as skin disorders and loss of hair, which are generally recoverable. More serious is the damage done to the bone marrow and to other cells which may not have the ability to repair the damage. This leads to leukaemia and to the production of cancerous cells and finally malignant tumours.

The effects of the third type of damage only appear in the future generations of those irradiated. Experiments on animals show that the genetic effects to be expected from breeding from radiation dosed human individuals include an increase in mental deficiencies, an increase in the number of monsters born and a general deterioration of the species in quality and population number.

The S.I. unit of radiation damage is the *gray*. 1 Gy corresponds to 1 J of energy absorbed per kilogram of target material. The gray is a large unit and the *rad*, equal to 0.01 Gy, is more widely used. *One rad is defined as the absorbed dose of any nuclear radiation that results in the absorption of 0.01 J of energy per kilogram of absorbing material.* Since the absorbed dose required to produce a certain effect may be different for various types of radiation, this difference is expressed by the relative biological effectiveness (RBE) of the particular nuclear radiation. Finally, a dose unit for biological effect is the *rem* which is the dose in rads multiplied by the RBE. The RBE of 250 keV X-rays is taken as 1, and a dosage of 1 rad of such X-rays to a person is considered 1 rem (rad equivalent man). The RBEs of other X-rays and of gamma and beta rays are close to 1. But the RBE of fast neutrons is about 10 and that of 1-MeV alpha particles is about 25; 1 rad absorbed from such alpha particles therefore means a dosage of about 25 rem.

### 18.11.2. Radiation Levels for Safety

The International Commission on Radiological Protection (ICRP) has recommended certain maximum permissible exposure (MPE) to radiation that is believed to result in no appreciable injury to a person. The ICRP requires that the maximum cumulative dose should not exceed  $5(N-18)$  REMS where  $N$  is the age of the person in years. The maximum short term dose should not exceed 3 REMS in three months a year, once a lifetime.

### 18.11.3. Radiation Protection Methods

The radiation exposure can be reduced to the safe limits by one or more of the following processes: by decreasing the time spent near a source, by allowing the source to decay for some time before approaching it, by increasing the distance from the source and by absorbing the radiation in shielding material. Other precautions include use of tongs and refraining from eating while using radioisotopes. Since persons receiving high or even lethal doses of radiation cannot sense it in any way, counters or dosimeters are used to determine if the prescribed maximum levels are exceeded. To cover individual workers in the range 10 milli REM to 1000 REM film badges are provided.

Protection can also be afforded by increasing resistance to radiation through drugs.

## 18.12 RADIATION DOSIMETERS

- A radiation dosimeter is a device used to detect the levels of exposure to ionizing radiation. Dosimeters are frequently used in environments where exposure to radiation may occur, such as nuclear power plants and medical imaging facilities.

### 18.12.1. Thermoluminescence

Some materials, upon absorption of radiation, retain part of the absorbed energy in metastable states. When this energy is subsequently released in the form of ultraviolet, visible or infrared light, the phenomenon is called *luminescence*. Phosphorescence occurs with a time delay exceeding  $10^{-8}$  s between stimulation and the emission of light. The process of phosphorescence can be accelerated with a suitable excitation. If the exciting agent is heat, the phenomenon is known as *thermoluminescence*. The material is called a thermoluminescent material, or a *TLD*.

The highly energetic secondary charged particles, usually electrons, that are produced in the primary interactions of photons with matter are mainly responsible for the photon energy deposition in matter. In a crystalline solid, these secondary charged particles release numerous low energy free electrons and holes through ionizations of atoms and ions. The free electrons and holes thus produced will either recombine or become trapped in an electron or hole trap, respectively, somewhere in the crystal.

The traps can be intrinsic or can be introduced in the crystal in the form of lattice imperfections consisting of vacancies or impurities. Two types of trap are known in general: storage traps and recombination centres.

- A storage trap merely traps free charge carriers and releases them during the subsequent heating, resulting in the thermoluminescence process.
- A charge carrier released from a storage trap may recombine with a trapped charge carrier of opposite sign in a recombination centre (luminescence centre). The recombination energy is at least partially emitted in the form of ultraviolet, visible or infrared light that can be measured with photomultiplier tubes (PMTs).

### 18.12.2. Thermoluminescent Dosimeter

**Principle.** Thermoluminescent dosimeters, or TLDs, use thermoluminescent materials to detect radiation. These materials trap electrons. To read the dosage, the TLD is heated, which causes it to generate light that can be read by electronic equipment to determine the level of dosage.

**Construction.** Fig. 18.16 shows a schematic diagram of a TLD reader.

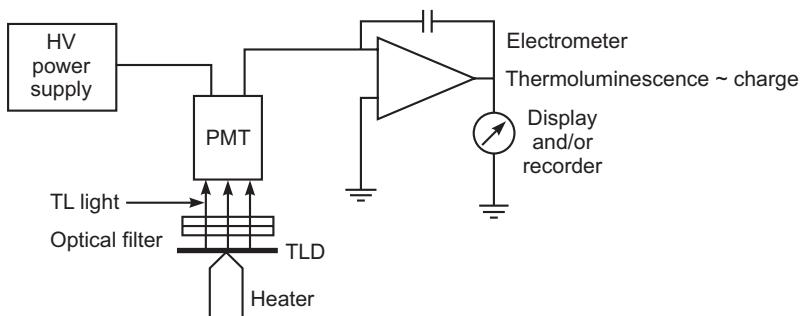


Fig. 18.16

- A **thermoluminescent dosimeter**, or **TLD**, is a type of radiation dosimeter. A **TLD** measures ionizing radiation exposure by measuring the amount of visible light emitted from a crystal in the detector when the crystal is heated. The amount of light emitted is dependent upon the radiation exposure.
- A basic **TLD** reader system consists of a planchet for placing and heating the **TLD**, a **PMT** to detect the thermoluminescence light emission and convert it into an electrical signal linearly proportional to the detected photon fluence and an electrometer for recording the **PMT** signal as a charge or current.
- The two most common types of **TLDs** are calcium fluoride and lithium fluoride, with one or more impurities to produce trap states for energetic electrons.

### Energy Level Diagram of the TL Process

A phenomenological model of the thermoluminescence mechanism is provided in terms of the band model for solids (Fig. 18.17).

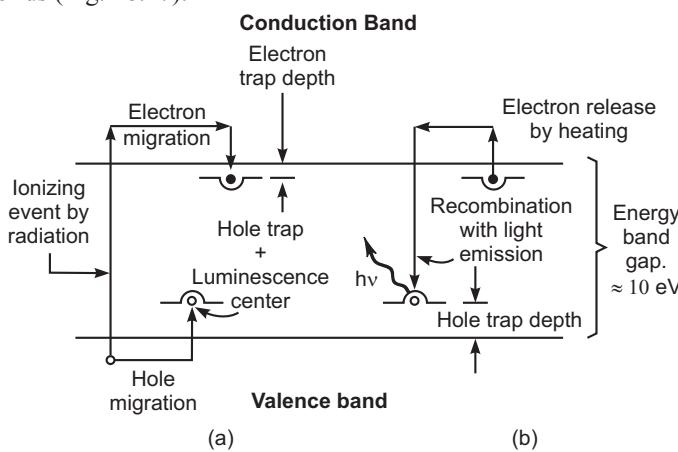


Fig. 18.17

Fig. 18.17 (a) shows *ionization by radiation, and trapping of electrons and holes*.

Fig. 18.17 (b) shows *heating to release electrons, allowing luminescence production*.

The storage traps and recombination centres, each type characterized with an activation energy (trap depth) that depends on the crystalline solid and the nature of the trap, are located in the energy gap between the valence band and the conduction band. The states just below the conduction band represent electron traps, the states just above the valence band are hole traps. The trapping levels are empty before irradiation (*i.e.*, the hole traps contain electrons and the electron traps do not).

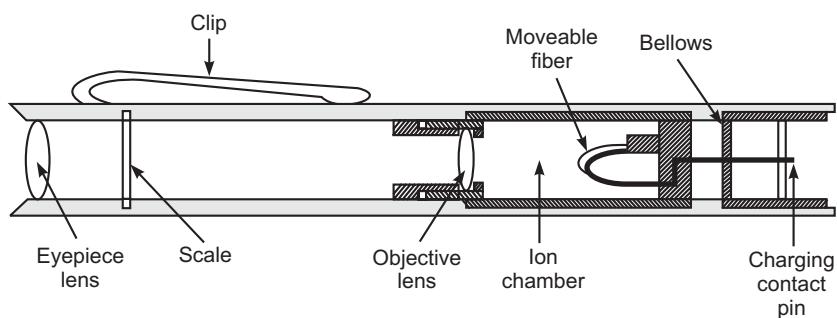
During irradiation, the secondary charged particles lift electrons into the conduction band either from the valence band (leaving a free hole in the valence band) or from an empty hole trap (filling the hole trap).

- As the radiation interacts with the crystal, it causes electrons in the crystal's atoms to jump to higher energy states, where they stay trapped due to intentionally introduced impurities (usually manganese or magnesium) in the crystal, until heated. Heating the crystal causes the electrons to drop back to their ground state, releasing a photon of energy equal to the energy difference between the trap state and the ground state.

### 18.12.3. Pocket Dosimeter

Pocket dosimeter is used to provide the wearer with an immediate reading of his or her exposure to x-rays and gamma rays. It is commonly worn in the pocket.

**Construction.** Fig. 18.18 shows pocket dosimeter.



**Fig. 18.18**

It consists of a tiny ionization chamber that contains a small quartz fiber. The chamber and fiber are contained in a metal tube with a lens that allows the fiber to be viewed.

When exposed to radiation, the quartz fiber is deflected. When viewed through the eyepiece lens, this deflection can be easily seen against a scale. The scale is graduated in units of exposure. Typical industrial radiography pocket dosimeters have a full scale reading of 200 milliroentgens.

### Digital Electronic Dosimeter

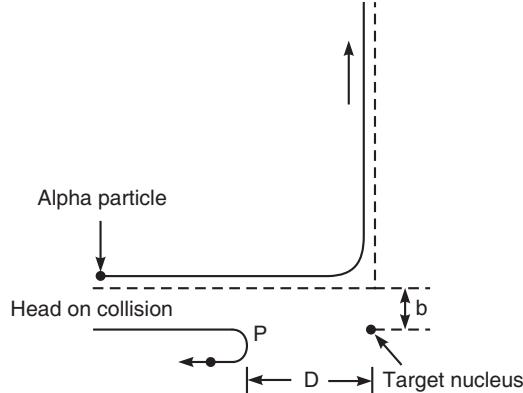
Another type of pocket dosimeter is the Digital Electronic Dosimeter. These dosimeters record dose information and dose rate. These dosimeters most often use Geiger-Müller counters. The output of the radiation detector is collected. When a predetermined exposure has been reached, the collected charge is discharged to trigger an electronic counter. The counter then displays the accumulated exposure and dose rate in digital form.

### EXERCISE

- The energy loss per unit length of the heavy charged particles is called  
 (a) photoelectric effect   (b) Compton power   (c) pair production   (d) stopping power  
 (B.U. 2015)

2. The ionization power is maximum for  
 (a) neutrons      (b) gamma rays      (c) beta particles      (d) alpha particles  
 (B.U. 2013)
3. In G.M. counter, the potential difference applied between anode and cathode is—V.  
 (a) 100      (b) 1000      (c) 10,000      (d) 500  
 (B.U. 2014)
4. The time during which pulses are recorded but are of smaller size in G.M. counter is called  
 (a) recovery time      (b) dead time      (c) resolving time      (d) none  
 (H.P.U. 2001)
5. In a scintillation counter, the scintillator generally used is  
 (a) anthracene      (b) argon      (c) camphor      (d) neon  
 (Barkatullah University, 2011)  
 [Ans. 1(d), 2(d), 3(b), 4 (a), 5(a)]
6. What is stopping power, impact parameter and range of charged particle?  
 (B.U. 2013)  
**[Hint: Impact parameter.** The impact parameter  $b$  is the minimum distance to which the  $\alpha$ -particle would approach the nucleus if there were no forces between them.]

For a head-on collision,  $b = 0$  (Fig. 18.19)



**Fig. 18.19**

7. Explain the interaction of gamma rays with matter. (B.U. 2013)
8. Explain the construction and working of ionization chamber. (B.U. 2012)
9. Explain the construction and working of solid state detectors. (B.U. 2013)
10. Explain the principle, construction and working of a proportional counter. (H.P.U. 2001)
11. Describe a GM counter. Explain its working. Mention its merit. (B.U. 2013)
12. Discuss the construction and working of a G.M. counter. How is quenching achieved in this counter? (Rohilkhand 2002)
13. Give the construction and working of Wilson cloud chamber. (B.U. 2015)
14. What is a bubble chamber? Describe the process by which ionising particles make a visible track in a bubble chamber. Give its advantages and drawbacks. (P.U. 2004)
15. Give a brief account of the nuclear emulsion technique in the study of ionising radiations.
16. Explain the principle, construction and operation of a scintillation radiation counter. (G.N.D.U. 2008)
17. A high energy-singly ionised particle leaves a track in a cloud chamber placed in a magnetic field of flux density  $2.5 \text{ Wb/m}^2$  and the radius of curvature of the track is 200 cm. Calculate the energy of the particle.  
 [Hint. Energy =  $E = pc = BRqc = 2.5 \times 2 \times 1.602 \times 10^{-19} \times (3 \times 10^8) \text{ J} = 240 \text{ MeV}$

## PARTICLE ACCELERATORS

**After reading this chapter, you should be able to**

- ◆ List the common accelerators used in the study of nuclear physics
- ◆ Discuss the constructional details and the working of the different types of particle accelerators
- ◆ Justify the motive behind constructing more and more energetic accelerators
- ◆ Explain why drift tubes are made successively longer inside a linear accelerator
- ◆ Explain the basic difference between a cyclotron and a betatron
- ◆ Explain why the cyclotron is not used for producing high energy electrons

### 19.1 VAN DE GRAAFF GENERATOR

**Principle.** In this electrostatic accelerator, electric charge is carried continuously to an insulated hollow spherical conductor ( $S$ ) and delivered to it. The potential of the hollow sphere can thus be gradually raised. When the potential has attained a sufficiently high value, it can be used to accelerate positive ions in the accelerator tube.

**Construction and working.** A well insulated endless belt  $B$  is rapidly rotated between two pulleys  $P_1$  and  $P_2$  by means of a motor (Fig. 19.1).

The upper pulley  $P_2$  is placed inside a hollow spherical conductor  $S$ . Near the lower pulley  $P_1$ , there is a metallic comb  $C$  connected to a source of steady voltage of 50 kV. Positive charge is sprayed over the moving belt by the metallic comb  $C$ . The belt moving upward carries this charge into the hollow sphere. This positive charge, on entering the sphere  $S$ , induces an equivalent negative charge on its inside and a corresponding positive charge on its outside. The negative charge is sprayed by the comb  $C'$  on to the belt, thus neutralising the positive charge on it. The net result is that the positive charges are in effect transferred from the moving belt to the outside of the sphere. The belt, returning to the bottom, is charged again and delivers up the charge, when inside the sphere. Charges are thus accumulated and distributed on the surface of the sphere. As the charge on the sphere accumulates, its potential rises to a very high value.

The charged particles which are to be accelerated are produced by a gas discharge ion source  $I$  placed inside the hollow sphere  $S$  over the accelerating tube  $G$ . The accelerating tube is evacuated and contains a number of insulated metal cylinders ( $C$ ) arranged with short gaps between them [Fig. 19.2]. The H.T. is applied to the topmost cylinder near the ion source and the bottom-most cylinder is earthed. A potential gradient exists from the top to the bottom. The ions are accelerated at the gaps between the cylinders, where intense electric field exists. The highly accelerated ions are deviated by a suitable and adjustable magnetic field  $M$  and made to strike the target ( $T$ ) at the bottom of the tube. By reversing the potential of the spray voltage, the Van de Graaff generator can also be used to accelerate electrons.

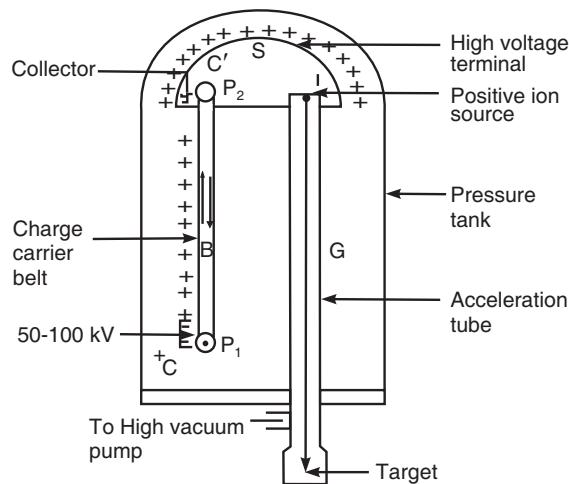


Fig. 19.1

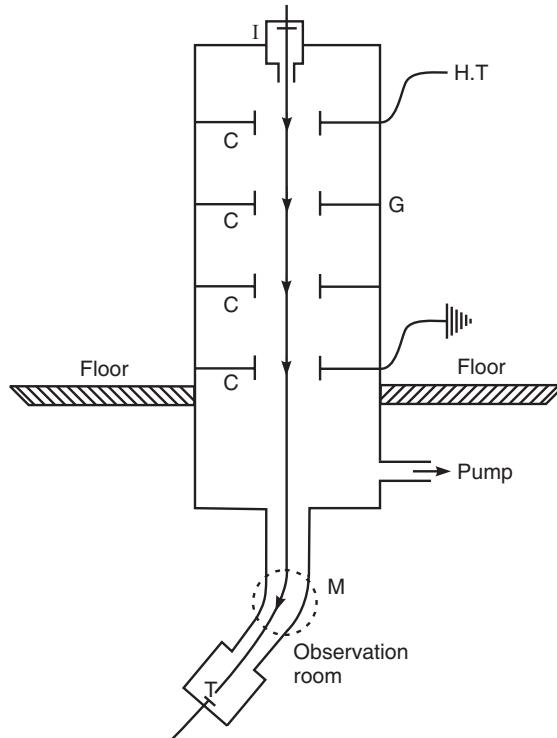


Fig. 19.2

## 19.2 THE LINEAR ACCELERATOR

**Principle.** Direct acceleration of particles by potentials above 10 million volts is a difficult problem due to insulation difficulties. For such high energies, acceleration of the particles is achieved in small successive steps.

- In linear accelerator, the P.D. between different parts of the machine and between the machine and earth, is maintained low, compared with the P.D. corresponding to the ultimate energy acquired by the particles. In this machine, high energy particles are produced without employing high P.D.'s, by using the principle of *synchronous acceleration*.

**Construction.** Fig. 19.3 shows the schematic diagram of a linear accelerator.

- It consists of a series of coaxial hollow metal cylinders or drift tubes 1, 2, 3, 4 etc. They are arranged linearly in a glass vacuum chamber. The alternate cylinders are connected together, the odd numbered cylinders being joined to one terminal and the even numbered ones to the second terminal of a H.F. oscillator.

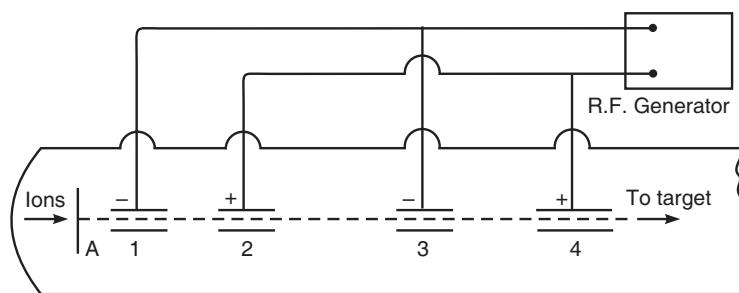


Fig. 19.3

Thus in one-half cycle, if tubes 1 and 3 are positive, 2 and 4 will be negative. After half a cycle, the polarities are reversed *i.e.*, 1 and 3 will be negative and 2 and 4 positive.

- The ions are accelerated only in the gap between the tubes where they are acted upon by the electric field present in the gaps. The ions travel with constant velocity in the *field-free space* inside the drift tubes.

**Theory and working.** Positive ions enter along the axis of the accelerator from an ion source through an aperture *A*. Suppose a positive ion leaves *A* and is accelerated during the half cycle, when the drift tube 1 is negative with respect to *A*. Let *e* be the charge and *m* the mass of the ion and *V* potential of drift tube 1 with respect to *A*. Then velocity *v*<sub>1</sub> of the ion on reaching the drift tube is given by

$$\frac{1}{2}mv_1^2 = Ve \text{ or } v_1 = \sqrt{\frac{2Ve}{m}}.$$

The length of the tube 1 is so adjusted that as the positive ions come out of it, the tube has a positive potential and the next tube (tube No. 2) has a negative potential, *i.e.*, the potentials change sign. The positive ion is again accelerated in the space between the tubes 1 and 2. On reaching the tube 2, the velocity *v*<sub>2</sub> of the positive ion is given by

$$\frac{1}{2}mv_2^2 = 2Ve \text{ or } v_2 = \sqrt{2} \sqrt{\frac{2Ve}{m}} = \sqrt{2} v_1.$$

This shows that *v*<sub>2</sub> is  $\sqrt{2}$  times *v*<sub>1</sub>. In order that this ion, on coming out of tube 2, may find tube 3 just negative and the tube 2 positive, it must take the same time to travel through the tube 2. Since

$v_2 = \sqrt{2} v_1$ , the length of tube 2 must be  $\sqrt{2}$  times the length of tube 1. For successive accelerations in successive gaps, the tubes 1, 2, 3, etc., must have lengths proportional to 1,  $\sqrt{2}$ ,  $\sqrt{3}$  etc. i.e.,  $l_1 : l_2 : l_3 : \text{etc.} = 1 : \sqrt{2} : \sqrt{3} : \text{etc.}$

**Energy of the ion.**  $n$  = the number of gaps that the ion travels in the accelerator.

$v_n$  = the final velocity acquired by the ion.

$$\left. \begin{array}{l} \text{Velocity of the ion, as it} \\ \text{emerges out of the } n^{\text{th}} \text{ tube} \end{array} \right\} = \sqrt{n} \sqrt{\frac{2Ve}{m}}$$

$$\therefore \text{K.E. acquired by the ion} = \frac{1}{2}mv_n^2 = nVe.$$

Thus the final energy of the ions depends upon (i) the total number of gaps and (ii) the energy gained in each gap.

**Limitations :** (i) The length of the accelerator becomes inconveniently large and it is difficult to maintain vacuum in a large chamber. (ii) The ion current available is in the form of short interval impulses because the ions are injected at an appropriate moment.

**EXAMPLE.** In a linear accelerator, proton accelerated thrice by a potential of 40 kV leaves a tube and enters an accelerating space of length 30 cm before entering the next tube. Calculate the frequency of the r.f. voltage and the length of the tube entered by the proton.

**SOL.** Let  $v_1$  and  $v_2$  be the velocities of the proton on entering and leaving the accelerating space. Let  $e$  and  $m$  be the mass and charge of the proton respectively. Then

$$\begin{aligned} \frac{1}{2}mv_1^2 &= 3 \times e \times 40000 \\ \therefore v_1 &= [(2 \times 3 \times 40000 \times (e/m))]^{1/2} \\ &= [2 \times 3 \times 40000 \times (9.578 \times 10^7)]^{1/2} \\ &= 4.794 \times 10^6 \text{ ms}^{-1}. \end{aligned}$$

$$\begin{aligned} \text{Similarly, } v_2 &= [2 \times 4 \times 40000 \times (9.578 \times 10^7)]^{1/2} \\ &= 5.536 \times 10^6 \text{ ms}^{-1}. \end{aligned}$$

$$\begin{aligned} \text{Mean velocity while travelling the } 0.3 \text{ m distance} \\ &= 5.165 \times 10^6 \text{ ms}^{-1}. \end{aligned}$$

The time taken to travel 0.3 m is half the period ( $T/2$ ) of the r.f. voltage.

$$\therefore \frac{T}{2} = \frac{0.3}{5.165 \times 10^6} \text{ or } T = \frac{2 \times 0.3}{5.165 \times 10^6}$$

$$\begin{aligned} \therefore \text{frequency of the } \left. \begin{array}{l} \text{r.f. voltage} \end{array} \right\} &= \frac{5.165 \times 10^6}{2 \times 0.3} \\ &= 8.608 \times 10^6 \text{ Hz} = 8.608 \text{ MHz}. \end{aligned}$$

The protons travel through the next tube for half a period with a velocity of  $5.536 \times 10^6 \text{ ms}^{-1}$ .

$\therefore$  length of the tube entered by the protons

$$L = 5.536 \times 10^6 \times \frac{1}{2 \times 8.608 \times 10^6} = 0.3216 \text{ m}$$

### 19.3 THE CYCLOTRON

**Principle.** The cyclotron is a “*magnetic resonance accelerator*”.

- The positive ions acquire energy from r.f. field when they cross the gap between two hollow metallic electrodes connected to the field. The time of transit of ions between electrodes is such that r.f. field is always in a phase to accelerate the ions when they cross the gap. Thus ions gain energy from r.f. field in steps every time they cross the gap between the two electrodes and eventually acquire energy much higher than the equivalent peak accelerating voltage.

The electric field provides the energy and magnetic field guides the motion.

- The value of the magnetic field strength depends upon the frequency of the oscillating electric field applied between the dees. It is so chosen as to give *resonance* between the arrival of the charged particle in the gap and the reversal of the voltage between the dees. This is done by adjusting the time taken by the charged particle to describe a semi-circular path equal to half the time period of oscillation of the applied high frequency electric field i.e.,

$$\frac{\pi}{\omega} = \frac{T}{2}$$

The frequency  $f = \frac{1}{T} = \frac{\omega}{2\pi} = \frac{Be}{2\pi m}$ .

This gives the *cyclotron resonance condition* for a charged particle of a given value of  $\frac{e}{m}$ .

**Construction.** The cyclotron (Fig. 19.4) consists of two hollow semicircular metal boxes,  $D_1$ ,  $D_2$  called “dees”. A source of ions is located near the mid-point of the gap between the “dees”. The

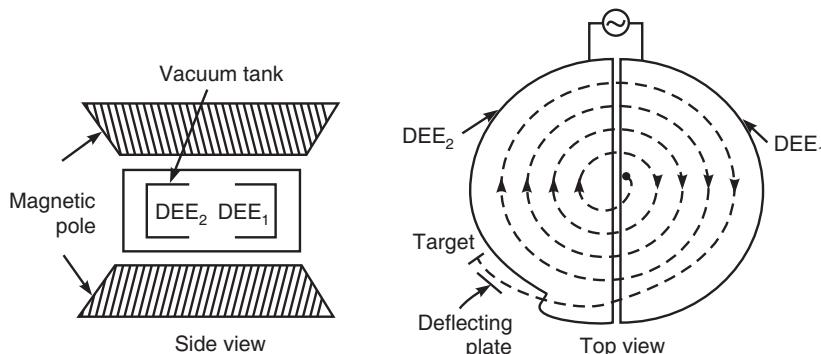


Fig. 19.4

“dees” are insulated from each other and are enclosed in another vacuum chamber. The “dees” are connected to a powerful radio-frequency oscillator. The whole apparatus is placed between the pole-pieces of a strong electromagnet. The magnetic field is perpendicular to the plane of the “dees”.

**Theory.** Suppose a positive ion leaves the ion source at the centre of the chamber at the instant when the “dees”  $D_1$  and  $D_2$  are at the maximum negative and positive A.C. potentials respectively. The positive ion will be accelerated towards the negative dee  $D_1$  before entering it.

- The ions enter the space inside the dee with a velocity  $v$  given by

$$Ve = \frac{1}{2}mv^2.$$

Here,  $V$  is the applied voltage.

$e$  and  $m$  are the charge and mass of the ion respectively.

- When the ion is inside the “dee”, it is not accelerated since this space is field free. Inside the dee, under the action of the applied magnetic field, the ions travel in a circular path of radius  $r$  given by

$$Bev = mv^2/r \quad \dots(1)$$

Here,  $B$  = the flux density of the magnetic field.

or  $r = mv/Be$

$$\text{The angular velocity of the ion in its circular path} = \omega = \frac{v}{r} = \frac{Be}{m} \quad \dots(3)$$

$$\text{The time taken by the ion to travel the semicircular path} = t = \frac{\pi r}{\omega} = \frac{\pi m}{Be} \quad \dots(4)$$

Suppose the strength of the field ( $B$ ) or the frequency of the oscillator ( $f$ ) are so adjusted that by the time the ion has described a semicircular path and just enters the space between  $D_1$  and  $D_2$ ,  $D_2$  has become negative with respect to  $D_1$ . The ion is then accelerated towards  $D_2$  and enters the space inside it with a greater velocity. Since the ion is now moving with greater velocity, it will describe a semicircle of greater radius in the second “dee”. But from the equation  $t = \pi m/Be$  it is clear that the time taken by the ion to describe a semicircle is independent of both the radius of the path ( $r$ ) and the velocity of the ion ( $v$ ). Hence the ion describes all semicircles, whatever be their radii, in exactly the same time. This process continues until the ion reaches the periphery of the dees. The ion thus spirals round in circles of increasing radius and acquires high energy. The ion will finally come out of the dees in the direction indicated, through the window.

**Energy of an ion.** Let  $r_{max}$  be the radius of the outermost orbit described by the ion and  $v_{max}$  the maximum velocity gained by the ion in its final orbit. Then the equation for the motion of the ion in a magnetic field is

$$Bev_{max} = \frac{mv_{max}^2}{r_{max}}$$

or  $v_{max} = B \frac{e}{m} r_{max}$   $\dots(5)$

$\therefore$  The energy of the ion

$$E = \frac{1}{2} mv_{max}^2 = \frac{B^2 r_{max}^2}{2} \left[ \frac{e^2}{m} \right] \quad \dots(6)$$

The condition for acceleration of the ion in the inter-dee gap is that

The time taken by the ion to  $\left. \begin{array}{l} \text{travel the semicircular path} \\ \text{travel the semicircular path} \end{array} \right\} = \text{Half the time period of oscillation of the applied high frequency voltage}$

i.e.,  $\frac{\pi m}{Be} = \frac{T}{2}$  or  $T = \frac{2\pi m}{Be}$

$\therefore$  Frequency of the oscillator

$$f = \frac{Be}{2\pi m} \quad \dots(7)$$

Hence the energy of the ion is given by

$$E = 2\pi^2 r_{max}^2 f^2 m \quad \dots(8)$$

The particles are ejected out of the cyclotron not continuously but as pulsed streams.

**Limitations of the Cyclotron.** The energies to which particles can be accelerated in a cyclotron are limited by the relativistic increase of mass with velocity. The mass of a particle, when moving with a velocity  $v$  is given by  $m = \frac{m_0}{\sqrt{1-v^2/c^2}}$  where  $m_0$  is the rest mass and  $c$  the velocity of light.

According to equation (4),

$$\left. \begin{array}{l} \text{The time taken by the ion to} \\ \text{travel the semicircular path} \end{array} \right\} = t = \frac{\pi m}{Be} = \frac{T}{2}$$

$$\therefore \text{Frequency of the ion} \} = n = \frac{1}{T} = \frac{Be}{2\pi m} \text{ or } n = \frac{Be\sqrt{1-v^2/c^2}}{2\pi m_0}.$$

Therefore, the frequency of rotation of the ion decreases with increase in velocity. The ions take longer time to describe their semicircular paths than the fixed period of the oscillating electric field. Thus, the ions lag behind the applied potential and finally they are not accelerated further. Due to this reason, the energy of the ions produced by the cyclotron is limited. This limitation can be overcome in the following two ways.

$$\text{Now, the frequency of the ion} = Be \frac{\sqrt{1-v^2/c^2}}{2\pi m_0}.$$

(i) **Field variation.** The frequency of the ion can be kept constant by increasing the magnetic field ( $B$ ) at such a rate that the product  $B\sqrt{1-v^2/c^2}$  remains constant. For this purpose, the value of the magnetic field  $B$  should increase, as velocity of the ion increases, so that the product  $B\sqrt{1-v^2/c^2}$  remains unchanged. This type of machine in which the frequency of electric field is kept constant and magnetic field is varied is called *synchrotron*.

(ii) **Frequency modulation.** In another form of apparatus, the frequency of the applied A.C. is varied so that it is always equal to the frequency of rotation of the ion. This type of machine in which magnetic field is kept constant and the frequency of the applied electric field is varied is called a *frequency modulated cyclotron* or *synchro-cyclotron*.

**EXAMPLE 1.** A cyclotron in which the flux density is 1.4 weber/m<sup>2</sup> is employed to accelerate protons. How rapidly should the electric field between the dees be reversed? Mass of the proton =  $1.67 \times 10^{-27}$  kg and charge =  $1.6 \times 10^{-19}$  C.

**SOL.** Here,  $B = 1.4$  weber/m<sup>2</sup>;  $m = 1.67 \times 10^{-27}$  kg;  $e = 1.6 \times 10^{-19}$  C.

$$\therefore t = \frac{\pi m}{Be} = \frac{\pi(1.67 \times 10^{-27})}{1.4 \times (1.6 \times 10^{-19})} = 2.342 \times 10^{-8} \text{ s.}$$

**EXAMPLE 2.** Deuterons in a cyclotron describe a circle of radius 0.32 m just before emerging from the dees. The frequency of the applied e.m.f. is 10 MHz. Find the flux density of the magnetic field and the velocity of deuterons emerging out of the cyclotron. Mass of deuterium =  $3.32 \times 10^{-27}$  kg;  $e = 1.6 \times 10^{-19}$  C.

**SOL.** We have,

$$f = \frac{Be}{2\pi m} \quad \therefore B = \frac{2\pi mf}{e}$$

Here,

$$m = 3.32 \times 10^{-27} \text{ kg}; f = 10 \text{ MHz} = 10^7 \text{ Hz}; e = 1.6 \times 10^{-19} \text{ C}$$

$$\therefore B = \frac{2\pi(3.32 \times 10^{-27})10^7}{1.6 \times 10^{-19}} = 1.303 \text{ weber/m}^2$$

We have  $\frac{mv^2}{r_{\max}} = Bev$  or  $v = \frac{Ber_{\max}}{m}$

Here,  $B = 1.303 \text{ weber/m}^2$ ;  $e = 1.6 \times 10^{-19} \text{ C}$ ;  $r_{\max} = 0.32 \text{ m}$  and  $m = 3.32 \times 10^{-27} \text{ kg}$ .

$$\begin{aligned}\therefore v &= \frac{Ber_{\max}}{m} = \frac{1.303(1.6 \times 10^{-19})0.32}{3.32 \times 10^{-27}} \\ &= 2.009 \times 10^7 \text{ ms}^{-1}.\end{aligned}$$

#### 19.4 SYNCHROCYCLOTRON

**Principle.** The resonance frequency of a cyclotron is given by

$$f_0 = \frac{qB}{2m}.$$

Here,

$m$  = mass of the particle

$q$  = Charge of the particle

$B$  = flux density of the magnetic field.

We know that the variation of mass with velocity is given by

$$m = \frac{m_0}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}}$$

If  $v = 0.8c$ , then  $v/c = 0.8$  and  $m = 1.66 m_0$ .

Thus at high velocities due to relativistic increase in mass, the orbital frequency of the particle is decreased. The particle gradually goes out of phase with the applied a.c. voltage whose frequency is constant. It is, therefore, necessary to provide corrective measures to compensate the decrease in frequency and to keep the particle in phase with the applied a.c. voltage.

Keeping the value of  $B$  constant, the oscillator frequency is decreased in step with the decrease in orbital frequency of the particle. A device which adopts this principle is called *Synchrocyclotron* or *frequency modulated cyclotron*.

In synchrocyclotron, the change in frequency should be done continuously and slowly compared with the rapid alterations of high frequency dee voltage. If this is not done, the phase stability will be disturbed. For a continuous change in frequency, a frequency modulated supply should be used.

The output of a cyclotron is continuous. The output of a synchrocyclotron is in *bursts* of a few hundred per second, each lasting for about  $100 \mu\text{s}$ .

**Construction and working.** Synchrocyclotron is a modified form of the Lawrence cyclotron. This consists of only one dee placed in a vacuum chamber between the poles of an electromagnet (Fig. 19.5). Instead of the second dee, opposite the opening of the dee, there is a metal sheet connected to the earth. The alternating P.D. is applied between the dee and the metal plate (Fig. 19.6).

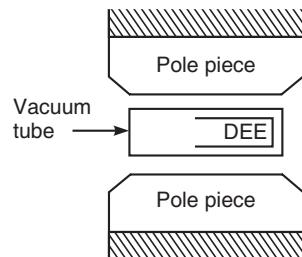


Fig. 19.5

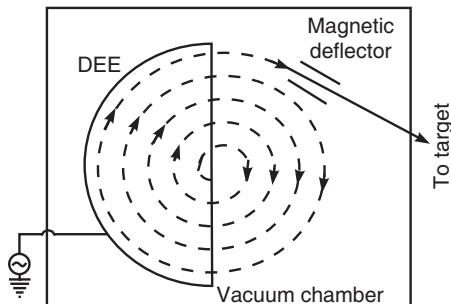


Fig. 19.6

The alternating potential applied to the “dee” is made to rise and fall periodically, instead of remaining constant. The frequency is changed at such a rate that as the ion lags a little due to the increase in mass caused by increase in velocity, the electric field frequency also automatically lags in variation. Hence the particle always enters the dee at the correct moment, when it can experience maximum acceleration.

#### Advantages of Synchro-cyclotron over cyclotron

- A single dee is used. An advantage of using one dee is that it leaves sufficient space in the vacuum chamber for the ion source and the target.
- The pole-pieces of the magnet are of suitable shape such that the field decreases outwards from the centre. This ensures good focusing of the accelerated ions.

#### Disadvantages

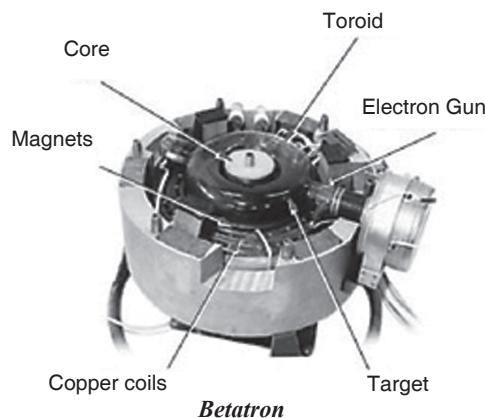
Following are the disadvantages of Synchro-cyclotron over cyclotron:

- (1) Only a small number of particles can be brought to the orbit of maximum radius and energy.
- (2) Output of the particle is in pulses.

## 19.5 THE BETATRON

Betatron is a device to accelerate electrons (beta particles) to very high energies. It was constructed in 1941 by D.W. Kerst. The action of this device depends on the principle of a transformer.

**Construction.** It consists of a doughnut-shaped vacuum chamber placed between the pole-pieces of an electromagnet. The electromagnet is energised by an alternating current. The magnet produces a strong magnetic field in the doughnut. The electrons are produced by the electron gun (*FG*) and are allowed to move in a circular orbit of constant radius in the vacuum chamber (Fig. 19.7). The magnetic field varies very slowly compared with the frequency of revolution of the electrons in the equilibrium orbit.



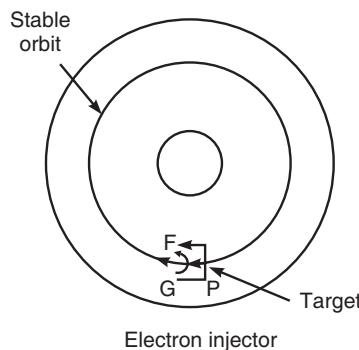


Fig. 19.7

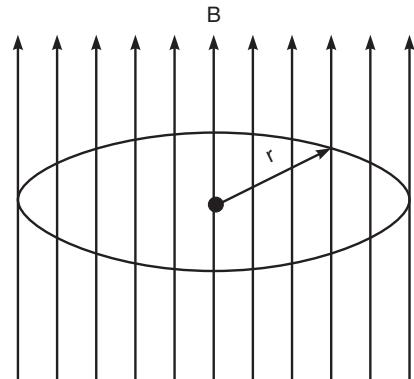


Fig. 19.8

The varying magnetic field, acting parallel to the axis of the vacuum tube, produces two effects on the electrons viz., (i) The changing flux due to the electromagnet produces the induced e.m.f. which is responsible for the acceleration of the electrons. (ii) The field of the magnet serves at the same time to bend the electrons in a circular path in the chamber and confine them to the region of the changing flux.

**Theory.** Consider the electron moving in an orbit of radius  $r$  (Fig. 19.8). Let  $\phi$  be the flux linked with the orbit. The flux increases at the rate  $d\phi/dt$  and the induced e.m.f. in the orbit is given by

$$E = -\frac{d\phi}{dt} \quad \dots(1)$$

$$\text{The work done on an electron of charge } e \text{ in one revolution} = Ee = -e \frac{d\phi}{dt} \quad \dots(2)$$

Let  $F$  be the tangential electric force acting on the orbiting electron.

For one revolution, the path length is  $2\pi r$ . Then

the work done on the electron in one revolution  $= F \times 2\pi r$

$$\begin{aligned} \therefore F \times 2\pi r &= -e \frac{d\phi}{dt} \\ \text{or } F &= -\frac{e}{2\pi r} \frac{d\phi}{dt} \end{aligned} \quad \dots(3)$$

When the velocity of the electron increases due to the above force, it will try to move into an orbit of larger radius. Because of the presence of the magnetic flux perpendicular to the plane of the electron orbit, the electron will experience a radial force inward given by

$$Bev = mv^2/r \quad \dots(4)$$

Here  $B$  is the value of the magnetic field intensity at the electron orbit of constant radius  $r$ ,  $v$  = velocity of the electron and  $m$  = mass of the electron.

From Eq. (4), the momentum of the electron  $= mv = Ber$   $\dots(5)$

From Newton's second law of motion,

$$F = \frac{d}{dt}(mv) = er \frac{dB}{dt} \quad \dots(6)$$

To maintain the constant radius of the orbit, the values of  $F$  given in equations (3) and (6) must be numerically equal.

$$\therefore \frac{e}{2\pi r} \frac{d\phi}{dt} = er \frac{dB}{dt} \text{ or } d\phi = 2\pi r^2 dB$$

$$\text{Integrating, } \int_0^\phi d\phi = \int_0^B 2\pi r^2 dB$$

$$\text{or } \phi = 2\pi r^2 B$$

...(7)

If the uniform magnetic field  $B$  acts over an area  $\pi r^2$ , the magnetic flux  $\phi' = \pi r^2 B$ . Therefore the flux through the orbit is twice the flux enclosed by the orbit, if the magnetic field were to be uniform over the area. Equation (7) represents the condition under which a betatron works and is called *betatron condition*. This distribution of magnetic flux is obtained by the special pole-pieces where the magnetic field is greater at the centre of the orbit than at its circumference.

Fig. 19.9 shows the variation of magnetic field with time. Electrons are injected into the chamber when magnetic field just begins to rise. The electrons are then accelerated by the increasing magnetic flux linked with the electron orbit. During the time the magnetic field reaches its peak value, the electrons make several thousand revolutions and get accelerated. If they are allowed to revolve any more, the decreasing magnetic field would retard the electrons. Hence, the electrons are extracted at this stage by using an auxiliary magnetic field to deflect them from their normal course. The high energy electron beam can be made to strike the target, generating X-rays. Alternately the electrons can be made to emerge out of the apparatus and used for transmutation work.

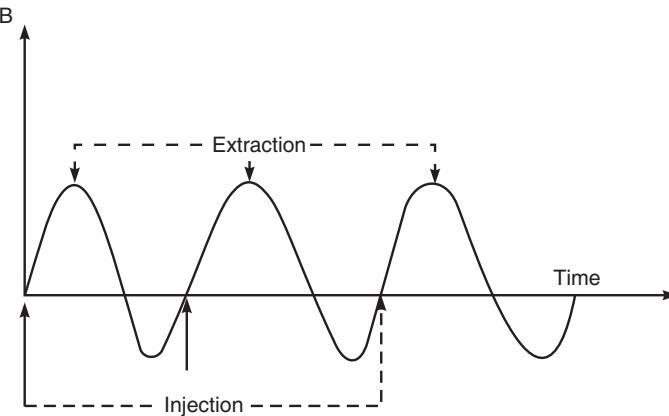


Fig. 19.9

**EXAMPLE.** In a certain betatron the maximum magnetic field at orbit was  $0.4 \text{ Wb/m}^2$ , operating at  $50 \text{ Hz}$  with a stable orbit diameter of  $1.524 \text{ m}$ . Calculate the average energy gained per revolution and the final energy of the electrons.

**SOL.** In the betatron, the electron velocities are nearly  $c$ .

$$\therefore \text{the total distance travelled in the acceleration time (i.e., one quarter cycle)} = c \times \frac{T}{4} = c \times \frac{\pi}{2\omega}$$

Total number of revolutions,

$$= N = \frac{c\pi / 2\omega}{2\pi r} = \frac{c}{4\omega r}.$$

Here, frequency  $= f = 50 \text{ Hz}$ .  $\therefore \omega = 2\pi f = 2\pi \times 50 = 100\pi$ ;  $r = 0.762 \text{ m}$ , and  $c = 3 \times 10^8 \text{ ms}^{-1}$ .

$$\therefore N = \frac{3 \times 10^8}{4(100\pi)0.762} = 3.132 \times 10^5$$

Let  $E$  be the final energy acquired by the electrons. Since the electrons must be treated relativistically,

momentum of the electron  $= mv = E/c$

But  $mv^2/r = Bev$  or  $mv = Ber$  or  $E = Berc$ .

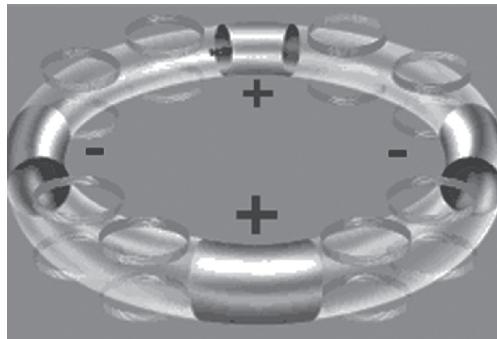
$$\therefore E = \frac{0.4(1.6 \times 10^{-19})(0.762)(3 \times 10^8)}{1.6 \times 10^{-13}} MeV \\ = 91.45 MeV.$$

$$\text{Average energy gained per revolution} = \frac{91.45 \times 10^6}{3.132 \times 10^5} = 291.9 eV.$$

## 19.6 THE SYNCHROTRONS

There are two types of synchrotrons : (a) electron synchrotron and (b) proton synchrotron.

**Electron synchrotron.** The electron synchrotron is based on the principle of the combined working of betatron and cyclotron. Electrons are injected into an orbit of fixed radius at an initial energy of about 50 to 80 keV. The main accelerating tube, the torus, is made of glass or some plastic with a circular “dee” (*D*) made of a metal. An alternating potential is applied to the “dee” as shown in Fig. 19.10. A varying magnetic field is applied perpendicular to the torus. The radius of the orbit is kept constant by increasing the magnetic field as in a betatron. The increments of energy are given, as in a cyclotron, at the beginning and ending of the *D*. The electrons, after acceleration, are made to strike the required target. Using tungsten as target, very hard X-rays of energy about 300 MeV have been produced.



Synchrotron

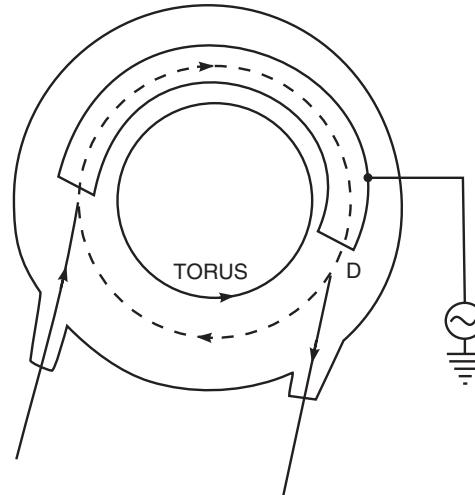


Fig. 19.10

## 19.7 THE PROTON SYNCHROTRON (BEVATRON, COSMOTRON)

The proton synchrotron has been designed to provide protons of very high energy of the order of billion electron volts.

**Construction and working.** The bevatron magnet consists of four segments connected by four equal straight sections (Fig. 19.11). The segments are quadrants of a circle of radius 30 feet and the connecting straight sections are 20 feet each. The protons are first accelerated by a linear accelerator and then injected into the synchrotron at one of the straight sections. The straight sections have an arrangement for applying the high frequency alternating potential. The magnetic field acts vertically upward. As the protons pass through one of the straight sections, they are accelerated and tend to have a bigger orbital radius. The magnetic field also increases at the same rate to constrain the proton beam to move in an orbit of constant radius. The output beam consists of a series of pulses of protons. By a suitable arrangement, the high energy particle may be made to strike a target.

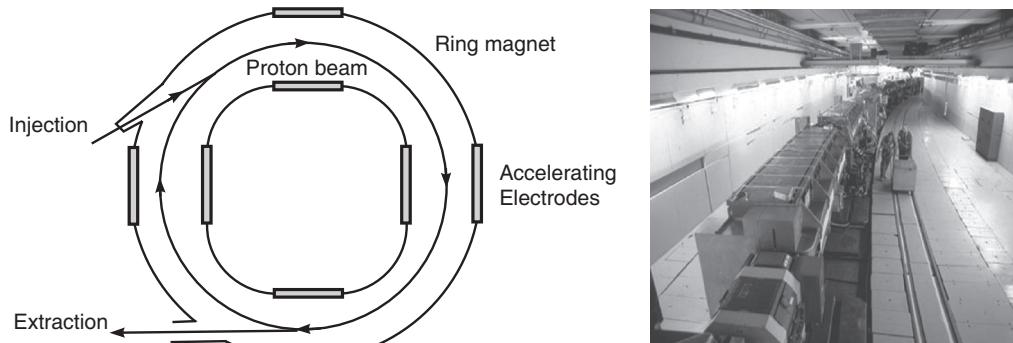


Fig. 19.11

Proton Synchrotron

**Theory.** The frequency of revolution of the ion in a circular orbit is given by  $n = Be/2\pi m$ . In this derivation, we have neglected the relativistic variation in the mass of the accelerated ion. To take the relativistic effects into consideration, let us write

$$n = \frac{1}{2\pi} \frac{Bec^2}{mc^2} = \frac{1}{2\pi} \frac{Bec^2}{(m_0 c^2 + T)} \quad \dots(1)$$

where  $m_0 c^2$  is the rest energy and  $T$  is the K.E. of the ion. Eqn. (1) gives the frequency of revolution of an ion in the absence of the straight sections. In the presence of four straight sections, each of length  $L$ , the frequency  $n'$  of circulation of protons is given by

$$n' = \frac{Bec^2}{2\pi(m_0 c^2 + T)} \frac{2\pi r_0}{(2\pi r_0 + 4L)} \quad \dots(2)$$

The proton is always guided by the magnetic field  $B$  to travel in the equilibrium orbit. Hence the following relation between  $B$  and the proton momentum  $p$  should hold at all times.

$$p = Ber_0 \quad \dots(3)$$

If  $T$  is the K.E. of a proton of momentum  $p$ , then

$$\begin{aligned} T + m_0 c^2 &= [(pc)^2 + (m_0 c^2)^2]^{1/2} = [(Becr_0)^2 + (m_0 c^2)^2]^{1/2} \\ \text{or } T^2 + 2T m_0 c^2 + m_0^2 c^4 &= B^2 e^2 c^2 r_0^2 + m_0^2 c^4 \\ \therefore B &= \frac{[T(T + 2m_0 c^2)]^{1/2}}{ecr_0} \end{aligned} \quad \dots(4)$$

Eqn. (4) describes the way in which the magnetic field  $B$  at the equilibrium orbit has to increase with the increase in the energy  $T$  of the proton as it circulates in the orbit. Substituting for  $B$  from Eqn. (4) in Eqn. (2), we get

$$n' = \frac{c[T(T + 2m_0 c^2)]^{1/2}}{(m_0 c^2 + T)(2\pi r_0 + 4L)} \quad \dots(5)$$

In the proton synchrotron, the accelerating voltage has to be in phase with the circulation frequency of the proton in the equilibrium orbit. Hence the radio-frequency of the accelerating voltage should always match  $n'$  and vary in the manner in which  $n'$  varies with the proton energy  $T$ .

**EXAMPLE.** What radius is needed in proton synchrotron to attain particle energies of 10 GeV, assuming that a guide field of 1.8 Wb/m<sup>2</sup> is available?

**SOL.** Equivalent mass of proton = 10 GeV + rest mass

$$= (10 + 0.938) \text{ GeV} = 11.75 \text{ u} = 11.75 \times (1.66 \times 10^{-27}) \text{ kg}$$

$$\text{Radius of the orbit} = \frac{mv}{Be} = \frac{(11.75 \times 1.66 \times 10^{-27})(3 \times 10^8)}{1.8 \times (1.6 \times 10^{-19})} = 20.31 \text{ m.}$$

### EXERCISE

1. A particle accelerator is used for increasing the \_\_\_\_\_ of charged particles.  
 (a) Kinetic energy      (b) Potential energy      (c) Mass      (d) e/m value  
 (B.U. 2012)
2. The accelerator that works on the principle of synchronous acceleration is  
 (a) Betatron      (b) Linear accelerator  
 (c) Vande Graff generator      (d) None of these  
 (B.U. 2013)
3. The principle used in cyclotron is  
 (a) resonance acceleration      (b) induced acceleration  
 (c) Pauli's principle      (d) Lawrence principle  
 (B.U. 2013)
4. Betatron is used to accelerate \_\_\_\_\_ particle.  
 (a) proton      (b) electron      (c)  $\alpha$  particle      (d) neutron  
 (B.U. 2006)  
 [Ans. 1. (a), 2. (b), 3. (a), 4. (b)]
5. Give the principle, construction and working of a Van-de-Graaf generator.      (H.P.U. 1992)
6. Explain the construction and working of a linear accelerator. Calculate the final energy of the ion.      (B.U. 2013)
7. Give the construction, theory and limitations of a cyclotron.      (Meerut 1983)
8. Explain the working of synchrocyclotron.      (B.U. 2006)
9. With necessary theory, explain the working of Betatron.      (B.U. 2006)
10. Describe a proton synchrotron and give its theory.      (P.U. 2007)

**After reading this chapter, you should be able to**

- ◆ Explain the types of radiation-alpha emission, beta emission, and gamma emission
- ◆ Compare and contrast briefly the properties of  $\alpha$ ,  $\beta$  and  $\gamma$ -rays
- ◆ Explain why  $\alpha$ -particles exhibit a definite range in matter
- ◆ Explain why an  $\alpha$ -particle, rather than neutrons and protons, is emitted from an unstable nucleus
- ◆ Discuss all the natural decay processes with their relevant theory
- ◆ Justify the fact that electrons are emitted from nuclei in  $\beta$ -decay, when these are not contained in the nucleus
- ◆ Discuss the laws of radioactive decay

## 20.1 INTRODUCTION

### Natural Radioactivity – definition

Nuclei with high  $A$  ( $\geq 206$ ) continuously emit positively charged  $\alpha$ -particles, negatively charged  $\beta$ -particles (high energy electrons) and high energy photons ( $\gamma$  rays). This activity is called *natural radioactivity*. The elements which exhibit this property are called *radioactive elements*. The atoms of radioactive elements emit radiations composed of three distinct kinds of rays ( $\alpha$ ,  $\beta$  and  $\gamma$ ). In the process, the elements break up, leading to an irreversible self-disintegration. The activity is *spontaneous*. Radioactivity is unaffected by any external agent like high temperature, high pressure, large electric and magnetic fields etc.

### Alpha, Beta and Gamma Rays

The existence of three distinct types of radiation is demonstrated by the following simple experiment. A small quantity of radium ( $R$ ) is placed at the bottom of a small hole drilled in a lead block [Fig. 20.1] A fairly parallel beam of radiation from  $R$  will issue through the hole. This lead block is placed inside an evacuated chamber to avoid absorption of the rays. A photographic plate ( $P$ ) is placed at a short distance above the lead block to receive the rays. A strong magnetic field is applied at right angles to the plane of the figure and directed away from the reader. After a fairly long exposure, the photographic plate is developed. Three distinct lines will be found on the photographic plate.

- The  $\alpha$ -particles will be deflected towards the left, indicating that they are *positively charged*.
- The  $\beta$ -particles will be deflected towards the right, showing that they are *negatively charged*.
- The  $\gamma$ -rays are not deflected and hit the plate  $P$  straight. This shows that the  $\gamma$ -rays are uncharged or neutral rays.

If an electrostatic field is applied, in place of the magnetic field (Fig. 20.2), the  $\beta$ -rays are deflected towards the positive plate, the  $\alpha$ -rays towards the negative and  $\gamma$ -rays do not bend at all.

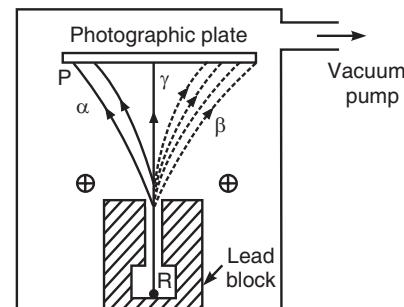


Fig. 20.1

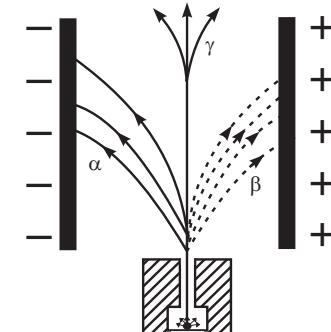


Fig. 20.2

### 20.1.1. Properties of Alpha Rays

(1) An  $\alpha$ -particle is a helium nucleus consisting of two protons and two neutrons. It carries two units of positive charge.

(2)  $\alpha$ -particles shot out from radioactive substances have high velocities, ranging from  $1.4 \times 10^7$  to  $1.7 \times 10^7$  m/s.

The velocity of emission of  $\alpha$ -particles is characteristic of the isotope from which they emanate. Rutherford's experiments gave the velocity of the alpha-particles from RaC as  $2 \times 10^7$  ms $^{-1}$ .

They move along straight lines. Their tracks can be observed in the Wilson's cloud chamber.

(3) They produce intense ionisation in the gas through which they pass. Their ionising power is 100 times greater than that of  $\beta$ -rays and 10,000 times greater than that of  $\gamma$ -rays.

(4) They affect a photographic plate but the effect is very feeble.

(5) They produce fluorescence when they fall on substances like barium platinocyanide or zinc sulphide. The fluorescence can be observed through a low power microscope.

(6) They are deflected by electric and magnetic fields. This shows that they are charged particles.

(7) They are scattered by nuclei of heavy elements like gold.

(8) They produce a heating effect. The evolution of heat is due to the stoppage of  $\alpha$ ,  $\beta$  and  $\gamma$ -rays by the radioactive substance.

(9) The value of  $E/M$  for  $\alpha$ -particles was found by Rutherford to be half of that for hydrogen ion. The charge on each  $\alpha$ -particle ( $E$ ) was found to be twice that of a hydrogen ion. Thus the mass of the  $\alpha$ -particles was shown to be four times that of hydrogen. Therefore, an  $\alpha$ -particle is the nucleus of the helium atom.

#### 20.1.2. Properties of Beta Rays

(1)  $\beta$ -particles possess negative charge and mass equal to that of an electron.

Kaufmann's experiments established that the charge to mass ratio ( $e/m$ ) of these particles is the same as that for electrons. Hence,  $\beta$ -particles are identical with electrons.

(2) All the  $\beta$ -particles emitted from a substance do not have the same velocity. The velocities of the  $\beta$ -particles emitted by a substance range continuously from  $0.3c$  to  $0.99c$ . At high velocities,  $e/m$  is found to decrease indicating an increase in mass of the particles, according to Einstein's equation,

$$m = \frac{m_0}{\sqrt{1-v^2/c^2}}.$$

(3) Their ionising power is low and hence, range is large.

(4) They affect a photographic plate.

(5) They produce fluorescence in barium platinocyanide, willemite etc.

(6) They are deflected by electric and magnetic fields. Their direction of deflection indicates that they are negatively charged particles.

(7) They penetrate through thin metal foils and their penetrating power is greater than that of  $\alpha$ -rays.

(8) The value of  $e/m$  for  $\beta$ -rays was proved to be the same as that of cathode rays. The charge on a  $\beta$ -particle was observed to be the same as the charge on an electron. Therefore,  $\beta$ -rays are identical with electrons.

#### 20.1.3. Properties of Gamma Rays

(1)  $\gamma$ -rays are electromagnetic waves of very short wavelength. The wavelengths of  $\gamma$ -rays range from  $0.005 \text{ \AA}$  to  $0.5 \text{ \AA}$ . The energy of the  $\gamma$ -rays is very large of the order of a few  $MeV$ . They are not charged particles. They travel with the velocity of light.

(2) They produce fluorescence and affect a photographic plate.

(3) They ionise the gas through which they pass but the ionisation produced is very small.

(4) They are more penetrating than even  $\beta$ -rays. They can even pass through an iron plate of 30 cm thickness.

(5) They are not affected by electric and magnetic fields.

(6) They are diffracted by crystals like X-rays.

## THE ALPHA RAYS

### 20.2 DETERMINATION OF E/M OF ALPHA PARTICLES

**Principle.** A beam of  $\alpha$ -particles is subjected to electric and magnetic fields. From the deflections produced, the value of  $E/M$  is determined.

**Action of magnetic field.** Fig. 20.3 shows the apparatus used for the magnetic deflection experiment.

- A fine wire coated with radium (*Ra C*) which emits  $\alpha$ -particles is placed in a cavity *L* of a lead block. The  $\alpha$ -particles emanating from the source are limited by a slit *S* and are incident on a photographic plate *P*. The whole arrangement is kept in an evacuated vessel *V*.
- A uniform magnetic field of flux density *B* is applied perpendicular to the plane of the paper. Under the action of the magnetic field, the  $\alpha$ -particle will describe an arc of a circle of radius *r*. On reversing the magnetic field, the path of the particle is reversed. The  $\alpha$ -particles reach the points *P*<sub>1</sub> and *P*<sub>2</sub> respectively before and after reversing the magnetic field. The velocity *v* of the  $\alpha$ -particles is more or less the same for all the particles in the case of the *Ra C* source. If *E* and *M* respectively be the charge and mass of the  $\alpha$ -particle, then

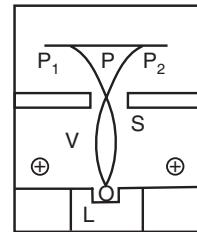


Fig. 20.3

$$BEv = \frac{Mv^2}{r}$$

or

$$\frac{E}{Mv} = \frac{1}{Br} \quad \dots(1)$$

The radius *r* of the circular path can be easily measured, since the path is defined by the three points *LSP*<sub>1</sub> or *LSP*<sub>2</sub>. Since *B* and *r* are known, *E/M* can be found if *v* is determined.

**Action of electric field.** To determine  $E/Mv^2$ , an apparatus as in Fig. 20.4 is used. In this experiment,  $\alpha$ -particles from the same source are subjected to the action of an electric field. *E*<sub>1</sub> and *E*<sub>2</sub> are two metallic plates. A strong electric field is maintained between the plates. The  $\alpha$ -particles passing through the electric field are deflected at right angles to their path. After travelling a parabolic path in the electric field, the particles are recorded on a photographic plate at *D*<sub>1</sub> and the deflection is *AD*<sub>1</sub>. When the field is reversed, the trace is obtained at *D*<sub>2</sub> and the deflection is *AD*<sub>2</sub>. The mean deflection *d* is found.

Let

*l* = the length over which the electric field acts.

*L* = distance of photographic plate from the field,

*X* = intensity of the electric field and

*d* = deflection on the photographic plate.

Then

$$d = \frac{XE}{M} \frac{l}{v^2} L$$

or

$$\frac{E}{Mv^2} = \frac{d}{XL} \quad \dots(2)$$

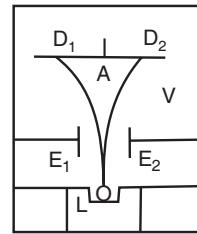


Fig. 20.4

**Determination of E/M.** Knowing  $\frac{E}{Mv}$  from Eqn. (1) and  $\frac{E}{Mv^2}$  from Eqn. (2), we can get the values of both  $E/M$  and  $v$ . The value for  $E/M$  of  $\alpha$ -particles is  $4.82 \times 10^7$  C/kg. This is very close to the value of  $E/M$  calculated for doubly ionised helium.

### 20.3 DETERMINATION OF THE CHARGE OF ALPHA PARTICLES

**Apparatus.** A known quantity of  $Ra\ C$  is placed in a shallow dish  $R$  (Fig. 20.5). The dish is covered with aluminium foil which absorbs the atoms from which  $\alpha$ -particles are emitted. These atoms cannot escape because of recoil. The  $\alpha$ -particles, after passing through a window ( $W$ ) of known area covered with a very thin aluminium foil, fall upon the collecting plate  $P$ .  $P$  is connected to an electrometer. The apparatus is kept exhausted.

**Working.** The charge received by the plate  $P$  in a given time  $t$  is found from the electrometer reading. A strong magnetic field prevents the  $\beta$ -rays emitted by the source from reaching the collecting plate. The number of  $\alpha$ -particles falling on the plate is determined by using the Geiger  $\alpha$ -particle counter. Hence the charge carried by each  $\alpha$ -particle is estimated.

$$E = 3.19 \times 10^{-19} \text{ C.}$$

#### Identification of $\alpha$ -particles

Charge on the  $\alpha$ -particle =  $E = 3.19 \times 10^{-19} \text{ C.}$

Electronic charge =  $e = 1.6 \times 10^{-19} \text{ C.}$

Hence  $E = 2e$ .

$$\text{Mass of } \alpha\text{-particles } M = \frac{E}{(E/M)} = \frac{3.19 \times 10^{-19}}{4.82 \times 10^7} = 6.62 \times 10^{-27} \text{ kg.}$$

Comparing this with the mass of the hydrogen atom

$$\frac{M}{M_H} = \frac{6.62 \times 10^{-27}}{1.67 \times 10^{-27}} \approx 4 \text{ or } M = 4M_H.$$

Thus the  $\alpha$ -particle has the mass of a helium atom and carries two positive elementary charges. This shows that the  $\alpha$ -particle is a helium atom that has lost both its electrons, i.e., a bare helium nucleus.

### 20.4 RANGE OF $\alpha$ -PARTICLES

**Definition :** The distance through which an  $\alpha$ -particle travels in a substance before coming to rest is called the 'range' of the particle in that substance.

**Explanation.**  $\alpha$ -particles are capable of ionising a gas. In the course of its passage through a gas, the  $\alpha$ -particle gets slowed down by losing energy in ionising the gas particles, until its energy falls below the ionisation potential of the gas. Then the  $\alpha$ -particle captures two electrons and becomes a neutral helium atom. The distance, the  $\alpha$ -particle travels in the gas in this manner, is called its range. Beyond this distance (range), the three properties of the  $\alpha$ -rays (ionising power, photographic action and fluorescence effect) disappear simultaneously.

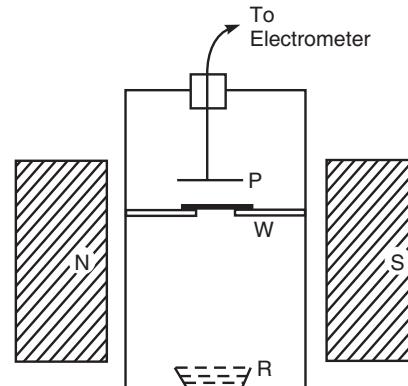


Fig. 20.5

- The range is usually expressed as so many cm in air at 0.76 m pressure and 15°C.
- The range depends on (i) the initial energy of the  $\alpha$ -particles (ii) the ionisation potential of the gas and (iii) the chances of collision between the  $\alpha$ -particle and the gas particles *i.e.*, on the nature and pressure of the gas.

#### 20.4.1. Experimental Measurement of the Range of $\alpha$ -particles

##### (i) Bragg and Kleeman method

**Principle.** Bragg and Kleeman were the first to determine the  $\alpha$ -particle range in gases by measuring the ionisation produced by  $\alpha$ -particles at different distances from the sources along their path within the medium.

**Apparatus.** Fig. 20.6 shows Bragg's apparatus for determining the range of  $\alpha$ -particles in a gas at different pressures.  $T$  is an evacuated glass tube. The radioactive substance  $R$  (in a lead block) is placed on the platform of the adjustable rod  $C$ . This rod can be raised up or down. The tube can be filled with the gas at any desired pressure which is measured by a manometer. The plate  $P$  is connected to the electrometer. The ionisation produced by the particles emitted by the source is measured by an ionisation chamber. The ionisation chamber consists of the plate  $P$  and a grid  $G$ . Between these electrodes is maintained a *P.D.* which ensures that an electrometer connected to the chamber measures the saturation ionisation current in the chamber. The second grid  $G'$  prevents the diffusion of ions from the chamber.

**Working.** The tube is filled with the required gas. The rod  $C$  is adjusted and the  $\alpha$ -particles reaching the grid  $G$  pass into the space between the grid  $G$  and plate  $P$ . The particular position of the platform is found when the ionisation current falls to zero. This is detected by the electrometer. The distance between the source and the grid  $G$  gives the range of the  $\alpha$ -particles for the particular element at a given temperature and pressure in a particular gas.

It is experimentally established that the range is inversely proportional to the density of the gas, *i.e.*, the range is inversely proportional to the pressure of the gas. *Specific ionisation* is the number of ions produced by  $\alpha$ -particles per unit length of the path. Bragg found that the curve relating the specific ionisation with the distance from the source was as shown in Fig. 20.7. It is seen that the graph shows the so called "Bragg hump" near the end of the range and also a slight "tail" or "straggling effect". The increase of ionisation near the end of the range can be understood because at this point the  $\alpha$ -particles have relatively low velocities and, therefore, have a greater probability of colliding with the gas atoms. The "tail" effect arises because all the  $\alpha$ -particles are not emitted from the source with the same initial velocity. All of them do not, therefore, reach the end of their range simultaneously.

##### (ii) Geiger and Nuttal experiment.

**Experimental arrangement.** The  $\alpha$ -particle source is in the form of a thin film on a small metal disc  $M$  (Fig. 20.8).

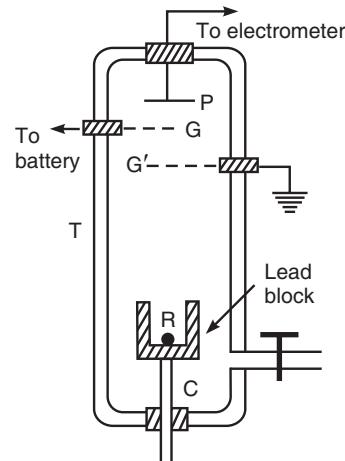


Fig. 20.6

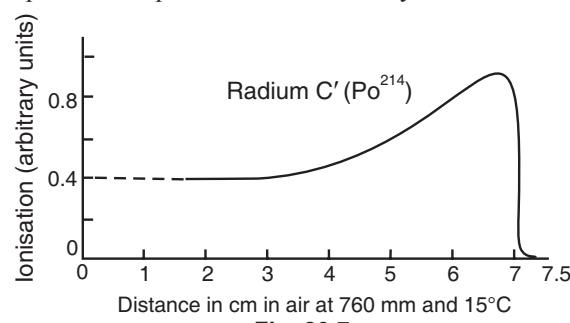


Fig. 20.7

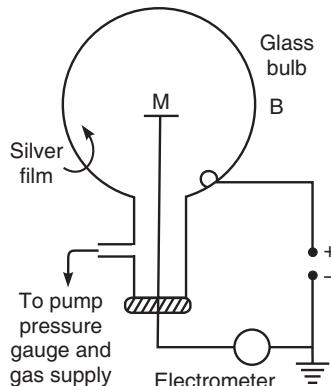


Fig. 20.8

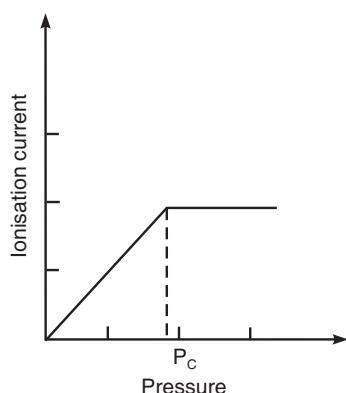


Fig. 20.9

It is mounted at the centre of a spherical glass bulb  $B$ . The bulb is coated inside with silver and a high  $P.D.$  is applied between the silver film and the disc  $M$ . The saturation ionisation current in the bulb produced by the passage of the  $\alpha$ -particles through the gas in the bulb is measured by an electrometer.

**Working.** The gas under study can be admitted at any desired pressure into the bulb. The saturation current for different gas pressures is measured. A graph is plotted between the ionisation current and the corresponding pressure  $P$  of the gas. We get curves of the type shown in Fig. 20.9.

The ionisation at first increases directly with the pressure. At a certain critical pressure  $P_c$ , the density of the gas is such that the particles are just stopped on the inner surface of the bulb, having produced the maximum ionisation possible, before reaching the bulb. If the pressure is increased beyond  $P_c$ , no rise in the current is observed, since the  $\alpha$ -particles get absorbed in a smaller path and do not reach the wall. The radius of the bulb thus gives the range of the  $\alpha$ -particles at the gas pressure  $P_c$ . As the range is inversely proportional to pressure, the range at normal pressure can be calculated.

**Geiger's law.** Geiger studied the relation between the range ( $R$ ) of an  $\alpha$ -particle and its velocity of emission ( $v$ ). He found that the range  $R$  is proportional to the cube of the velocity.  $R \propto v^3$  or  $R = av^3$ , where  $a$  is a constant.

**Range-energy relation.** The energy  $E$  of the emitted particle is directly proportional to the square of the velocity  $v$ . Hence Geiger's law can be expressed in the form  $R = bE^{3/2}$  where  $b$  is a constant.

**Geiger-Nuttal law.** The range  $R$  of an  $\alpha$ -particle and the disintegration constant  $\lambda$  of the radioactive element that emits it are related as follows:

$$\log \lambda = A + B \log R.$$

This relation is called Geiger-Nuttal law.

**Explanation.** If  $\log \lambda$  are plotted against  $\log R$  for the different  $\alpha$ -emitters in the three series, three nearly parallel straight lines are obtained, one for each series (Fig. 20.10).

In the relation  $\log \lambda = A + B \log R$ , the constant  $B$  is the same for all the series while  $A$  is different for the different series.

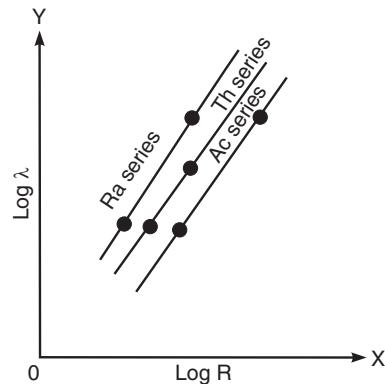


Fig. 20.10

- According to this relation, when the disintegration constant is high, the range is also high. Since the range also depends on the energy, we conclude that radioactive substances of large decay constants emit high energy  $\alpha$ -particles.

- This law is helpful in determining roughly the decay constants of radioactive substances of very short or very long lives. Experimentally measuring the ranges of  $\alpha$ -particles of such radioelements, the respective decay constants can be obtained by extrapolation from the curves representing the Geiger-Nuttal relation.

## 20.5 ALPHA PARTICLE DISINTEGRATION ENERGY

When an  $\alpha$ -particle is emitted from a nucleus, the nucleus recoils in order to conserve momentum. Let  $m$  and  $v$  be the mass and velocity of the  $\alpha$ -particle. Let  $M$  and  $V$  be the mass and velocity of the daughter nucleus. According to the law of conservation of momentum,

$$mv = MV.$$

The sum of the kinetic energies of the  $\alpha$ -particle and the product nucleus is called alpha disintegration energy ( $E$ ).

$$\begin{aligned} E &= \frac{1}{2}mv^2 + \frac{1}{2}MV^2 = \frac{1}{2}mv^2 + \frac{1}{2}M\left(\frac{mv}{M}\right)^2 \\ \therefore E &= \frac{1}{2}mv^2\left(1 + \frac{m}{M}\right). \end{aligned}$$

The  $\alpha$  disintegration energy can thus be easily obtained by multiplying the K.E. of the  $\alpha$ -particles by the factor  $(1 + m/M)$ .

## 20.6 ALPHA PARTICLE SPECTRA

It was thought earlier that the nucleus of a radioactive substance emitted alpha particles of some fixed energy. By measuring the energies of  $\alpha$ -particles accurately using a magnetic spectrograph, it was shown that several radioactive nuclei emitted  $\alpha$ -particles of different energies falling within a narrow range.

**Example.** Consider a  $ThC$  nucleus which decays into a  $ThC''$  ( $_{81}Tl^{208}$ ) nucleus after emitting an  $\alpha$ -particle (Fig. 20.11).

- $ThC$  emits five groups of  $\alpha$ -particles with energies 5.709, 5.728, 5.873, 6.161 and 6.201 MeV.
- The emission of groups of  $\alpha$ -particles with different energies suggests that nuclei may exist in a number of discrete, excited energy states above the ground state. The parent nucleus decays from its ground state to anyone of the several excited states or to the ground state of the daughter nucleus.
- The daughter nucleus  $ThC''$  formed as a result of disintegration of  $ThC$  exists in several excited energy states  $a, b, c, d, e$ .
- $e$  is the ground state and  $a$  is the highest excited state of the daughter nucleus.
- Each excited state has got some energy which it gives up in the form of  $\gamma$  radiation during its transition to the ground state.
- When a parent nucleus goes from its ground state to the ground state of the daughter nucleus, it emits during such a transition, an  $\alpha$ -particle of maximum energy ( $\alpha_1$  in the energy diagram).
- Any transition from the ground state of the parent nucleus to any one of the excited states of the daughter nucleus will give rise to  $\alpha$ -particle with smaller energy ( $\alpha_2, \alpha_3, \alpha_4$  and  $\alpha_5$ ).

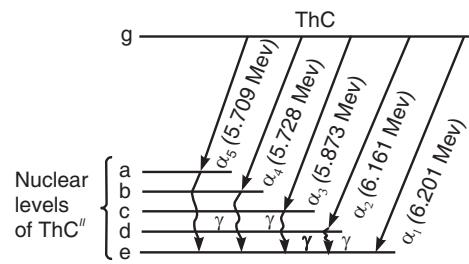


Fig. 20.11

- The transition from any excited state to the ground state involves the emission of  $\gamma$ -rays. The energy of the  $\gamma$ -ray photons have been measured and found to be equal to the energy difference between one of the excited energy states and the ground state.
- Thus, the  $\alpha$ -spectrum provides a conclusive evidence of the existence of *discrete energy levels in a nucleus*.

## 20.7 THEORY OF ALPHA DECAY

Only heavy nuclei with  $A > 200$  undergo  $\alpha$ -decay. The  $\alpha$ -particles emitted from nuclei have a discrete energy spectrum and consist of several groups. Usually the most intensive is the group with  $\alpha$ -particles of highest energy. The existence of several groups of  $\alpha$ -particles is called the *fine structure of the  $\alpha$ -spectrum*.

Before emission, the  $\alpha$ -particle can be considered to be inside the nucleus. Coulomb's law is applicable when the  $\alpha$ -particle is outside the nucleus. When the  $\alpha$ -particle is inside the nucleus or very close to it, Coulomb's law does not hold good. For an  $\alpha$ -particle of charge  $2e$  and a nucleus of charge  $(Z-2)e$  separated by a distance  $r$ , the potential energy is given by

$$V_r = \frac{2e(Z-2)e}{4\pi\epsilon_0 r} = \frac{2(Z-2)e^2}{4\pi\epsilon_0 r}.$$

Here  $(Z - 2)$  is the atomic number of the daughter nucleus. An attractive potential, called the potential well, represents the position (Fig. 20.12). When  $r < r_0$ , the  $\alpha$ -particle is within the potential well and it is bound by the nucleus. Here  $r_0$  is equal to the sum of the radii of the nucleus and the  $\alpha$ -particle. According to wave mechanics, an  $\alpha$ -particle can have different energy levels  $E_0, E_1, E_2, \dots$ , etc., within the potential well. Only an  $\alpha$ -particle at a level  $E_3$  or above can come out of the potential well. In the case of radioactive elements, the height of the potential barrier is about 9 MeV. But uranium emits  $\alpha$ -particles of energy 4 MeV. Classical mechanics cannot explain how a particle having an energy of 4 MeV can come out of a well having a potential barrier of 9 MeV.

The escape of an  $\alpha$ -particle from a radioactive nucleus can be explained on the basis of wave mechanics and Schrodinger equation. According to it, it is possible for an  $\alpha$ -particle to *leak* through the potential barrier, even though its K.E. is less than the potential energy of the height of potential barrier. This probability of leaking of an  $\alpha$ -particle through the barrier is called *tunnel effect*. The  $\alpha$ -particle within the nucleus must present itself again and again at the barrier surface until conditions are ripe for penetration or leakage.

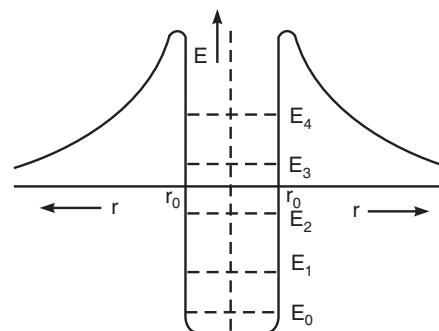


Fig. 20.12

## 20.8 GAMOW'S THEORY OF ALPHA DECAY

Classical physics fails to explain  $\alpha$ -decay. Quantum mechanics provides a successful explanation of the problem of  $\alpha$ -decay. The basic notions of this theory are :

- An alpha particle may exist as an entity within a heavy nucleus.
- Alpha particle is in constant motion and bounces back and forth from the barrier walls. In each collision with 'wall' there is a definite probability that the particle will leak through the potential barrier. Let  $v$  be the frequency with which the  $\alpha$ -particle collides with the walls in order to escape from the nucleus and  $P$  the probability of transmission in each collision. Then the decay probability per unit time (*i.e.*, disintegration constant  $\lambda$ ) is given by

$$\lambda = vP \quad \dots(1)$$

Suppose that at any moment only one  $\alpha$ -particle exists as such in a nucleus and that it moves back and forth along a nuclear diameter.

Then

$$v = \frac{v}{2R} \quad \dots(2)$$

where  $v$  is the alpha particle velocity and  $R$  is the nuclear radius.

Since  $V > T$  [where  $V$  = height of the potential barrier and  $T$  = K.E. of  $\alpha$ -particle], according to classical physics  $P = 0$  [Fig. 20.13]. But in quantum mechanics a moving particle is regarded as a wave, and the result is a small but definite value for  $P$ . The probability  $P$  can be calculated quantum mechanically using WKB perturbation theory. According to this theory,

$$\log_e P = \frac{-2}{\hbar} \int_R^{R_1} \sqrt{2m(V(x)-T)} dx \quad \dots(3)$$

where

$m$  = mass of the  $\alpha$ -particle.

$$V(x) = \frac{2Ze^2}{4\pi\epsilon_0 x}$$

is the electrostatic P.E. of an alpha particle at a distance  $x$  from the centre of a nucleus of charge  $Ze$  ( $Ze$  is the charge of the daughter nucleus i.e., nuclear charge minus the alpha particle charge of  $2e$ )

$R$  = Nuclear radius

$T$  = K.E. of  $\alpha$ -particle such that  $T < V(x)$ .

The region from  $x = R$  to  $x = R_1$  is called the thickness of the barrier.

We therefore have:

$$\log_e P = \frac{-2}{\hbar} \int_R^{R_1} \sqrt{2m \left( \frac{2Ze^2}{4\pi\epsilon_0 x} - T \right)} dx \quad \dots(4)$$

When

$$x = R_1, T = V. \therefore T = \frac{2Ze^2}{4\pi\epsilon_0 R_1}$$

∴

$$\log_e P = \frac{-2}{\hbar} (2mT)^{1/2} \int_R^{R_1} \left( \frac{R_1}{x} - 1 \right)^{1/2} dx$$

$$\begin{aligned} \log_e P = & \frac{-2}{\hbar} (2mT)^{1/2} R_1 \left[ \cos^{-1} \left( \frac{R}{R_1} \right)^{1/2} \right. \\ & \left. - \left( \frac{R}{R_1} \right)^{1/2} \left( 1 - \frac{R}{R_1} \right)^{1/2} \right] \end{aligned} \quad \dots(5)$$

The width of the potential barrier is very large compared with the nuclear radius i.e.,  $R_1 \gg R$ .

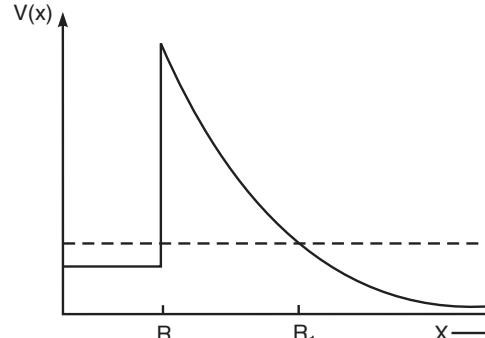


Fig. 20.13

Therefore

$$\cos^{-1}\left(\frac{R}{R_l}\right)^{1/2} \approx \frac{\pi}{2} - \left(\frac{R}{R_l}\right)^{1/2} \text{ and } \left(1 - \frac{R}{R_l}\right)^{1/2} \approx 1$$

Hence  $\log_e P = \frac{-2}{\hbar}(2mT)^{1/2} R_l \left( \frac{\pi}{2} - 2\left(\frac{R}{R_l}\right)^{1/2} \right)$

Substituting  $R_l = 2Ze^2 / 4\pi \epsilon_0 T$ , we have

$$\log_e P = \frac{4e}{\hbar} \left( \frac{m}{\pi \epsilon_0} \right)^{1/2} Z^{1/2} R^{1/2} - \frac{e^2}{\hbar \epsilon_0} \left( \frac{m}{2} \right)^{1/2} Z T^{-1/2} \quad \dots(6)$$

or  $\log_e P = 2.97 Z^{1/2} R^{1/2} - 3.95 Z T^{-1/2}$

where  $R$  is in fermi, and  $T$  is in MeV.

Since

$$\lambda = v P, \log_e \lambda = \log_e v + \log_e P = \log_e \left( \frac{v}{2R} \right) + \log_e P$$

$\therefore \log_e \lambda = \log_e \left( \frac{v}{2R} \right) + 2.97 Z^{1/2} R^{1/2} - 3.95 Z T^{-1/2} \quad \dots(7)$

Changing the base of log (from  $e$  to 10),

$$\log_{10} \lambda = \log_{10} \left( \frac{v}{2R} \right) + 0.4343(2.97 Z^{1/2} R^{1/2} - 3.95 Z T^{-1/2})$$

or  $\log_{10} \lambda = \log_{10} \left( \frac{v}{2R} \right) + 1.29 Z^{1/2} R^{1/2} - 1.72 Z T^{-1/2} \quad \dots(8)$

The changes in atomic number and nuclear radius are negligible when compared to the changes in energy. The first term is almost same for heavier nuclei. So Eq. (8) reduces to

$$\log_{10} \lambda = c + d T^{-1/2} \quad \dots(9)$$

where "c" and "d" are constants.

Eq. (9) shows that the emitters having lesser decay constants emit  $\alpha$ -particles of greater energy ( $T$ ) which is the Geiger and Nuttal law.

#### Experimental verification of theory of $\alpha$ -decay

For a number of alpha-active nuclei, a plot of  $\log_{10} \lambda$  vs  $ZT^{-1/2}$  is shown in Fig. 20.14. The straight line has slope  $-1.72$  as predicted by the theory. The intercept on the  $Y$ -axis gives the value of  $\log_{10} \left( \frac{v}{2R} \right) + 1.29 Z^{1/2} R^{1/2}$ .

This can be used to determine the value of nuclear radius  $R$ . The nuclear radius calculated in this way comes out to be of the same order as obtained from scattering experiments.

Further this theory makes understandable the enormous variation in half-life with disintegration energy. Nuclei emitting most energetic  $\alpha$ -particles are short lived and those emitting least energetic  $\alpha$ -particles are long lived. This correlation between the half-life time (or disintegration constant  $\lambda$ ) and the energy of the  $\alpha$ -particle is predicted by Eq. (8).

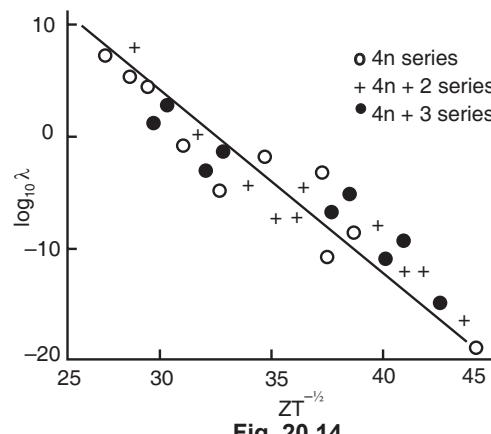


Fig. 20.14

## BETA RAYS

### 20.9 DETERMINATION OF E/M OF BETA PARTICLES : KAUFMANN'S EXPERIMENT

**Apparatus.** Fig. 20.15 shows the apparatus used by Kaufmann.

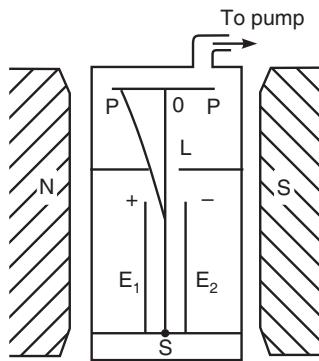


Fig. 20.15

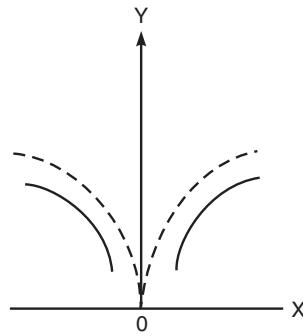


Fig. 20.16

In a highly evacuated vessel, a speck of radium bromide ( $S$ ) is placed. The charged particles from the source  $S$  are limited by a narrow slit  $L$  in a lead plate and are recorded on a photographic plate  $P$ . They are acted upon by a uniform electric field  $X$  parallel to the slit in the course of their travel from  $S$  to  $L$ . The electric field is produced by applying a high P.D. between two metal plates  $E_1$  and  $E_2$ . The whole arrangement is placed between the pole-pieces of an electromagnet, so that the particles are acted upon by a uniform magnetic field  $B$  parallel to  $X$ . Hence the magnetic field produces a deflection in a direction perpendicular to that due to electric field. The comparatively heavy  $\alpha$ -particles from the source produce a central spot ( $O$ ) on  $P$ , since they are almost undeviated by the fields.

**Theory.** The deflection due to the electric field  $= x = \frac{Xell}{mv^2}$

Here,

$X$  = intensity of the electric field,

$l$  = the length of the electrodes  $E_1$  or  $E_2$  over which the field acts  
and  $L$  = the distance of the photographic plate from the field.

Due to the magnetic field, the particles travel in arcs of radius  $r$  given by

$$Bev = mv^2/r.$$

The deflection  $y$  due to the magnetic field is proportional to  $Be/mv$  and can be calculated from the geometry of the apparatus.

The source emits  $\beta$ -particles with a range of values of velocity  $v$ . Hence, the particles fall on the photographic plate along an arm of a parabola, each point on it corresponding to particles of a certain velocity (as in the case of Thomson's parabola method of positive ray analysis). When the electric or magnetic field is reversed, the other arm of the parabola is obtained. The parabolic traces obtained are shown in Fig. 20.16 by the continuous lines. The values of  $x$  and  $y$  are found for different points on the curve and hence  $e/m$  is calculated.

**Increase of Beta particle mass with velocity.** The traces actually obtained are found not to coincide with the parabolic curves that would be expected if  $e/m$  is the same for all velocities (shown by dotted lines in the figure). This shows that  $e/m$  is not constant but varies. For slow particles,  $e/m$  is found to be equal to that of an electron. The value of  $e/m$  decreases, as the velocity

increases. This can be explained by the relativistic law of variation of mass with velocity  $m = m_0 / \sqrt{1 - v^2 / c^2}$ . With the experimental value of  $e/m$  for the different values of  $v$ , Kaufmann calculated the value of  $e/m_0$ . The value of  $e/m_0$  was found to be a constant for all values of  $v$ .

## 20.10 BETA RAY SPECTRUM

**Theory.** The energies of  $\beta$ -particles from radio-active elements are determined by measuring the radii of curvature of their paths in a magnetic field of known flux density  $B$ . The circular path traversed by the  $\beta$ -particles of velocity  $v$  is governed by the relation

$$Bev = \frac{mv^2}{r}.$$

∴

$$v = Br(e/m).$$

From the geometry of the arrangement, the radius of the circular path  $r$  can be found.

The value of  $e/m$  can be assumed. Hence, the velocity  $v$  can be calculated.

For particles moving with very high velocities, the kinetic energy of the particle,

$$E_k = mc^2 - m_0c^2 = m_0c^2 \left[ \frac{1}{(1 - v^2/c^2)^{1/2}} - 1 \right]$$

### Magnetic spectrograph

**Experimental study.** The energy spectrum of  $\beta$ -particles is studied with the help of a magnetic spectrograph (Fig. 20.17).

The source is the radioactive substance  $S$  under study, coated on a fine wire.  $A$  and  $B$  are diaphragms used for limiting the beam of  $\beta$ -particles.  $G$  is a Geiger counter.

$\beta$ -rays from the source at  $S$  are bent around by the magnetic field and focused on the aperture  $O$ .

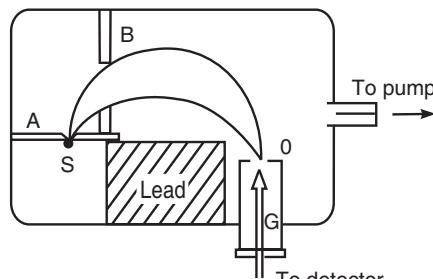


Fig. 20.17

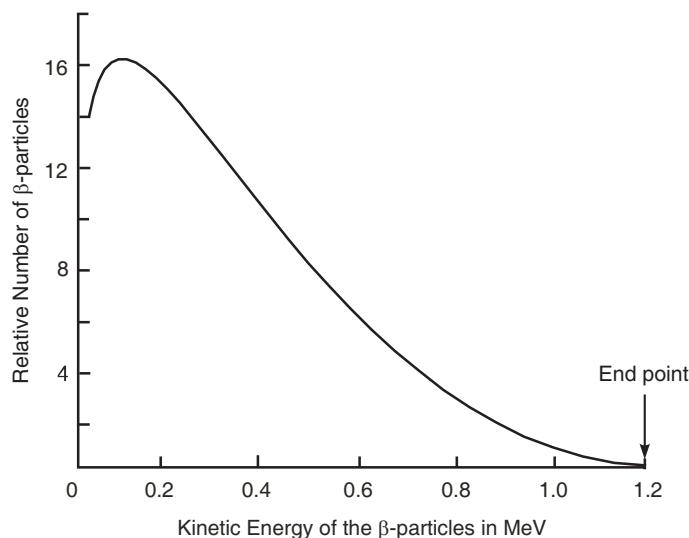


Fig. 20.18

**Working.** The number of  $\beta$ -particles entering the aperture  $O$  is counted at a given value of the magnetic field of intensity  $B$ . The intensity of the magnetic field is then changed to a new value and the number of  $\beta$ -particles entering the aperture  $O$  is again counted. In this manner, the velocity distribution of the  $\beta$ -particles is determined.

Fig. 20.18 shows the continuous  $\beta$ -ray spectrum of radium  $E$ .

- The curve first rises to a maximum and then decreases to zero at a well defined velocity, which corresponds to the maximum velocity of the  $\beta$ -particles.
- Below the maximum energy ( $E_{\max}$ ), the  $\beta$ -ray particles give a continuous spectrum.

**End point energy.** The  $\beta$ -spectrum is continuous, having energies ranging from zero to a certain well defined limit called *End point energy*, which is characteristic of the  $\beta$ -emitter. The energy distribution has a well defined maximum. The height and position of the maximum also depends upon the nature of the radioactive substance emitting the  $\beta$ -particles.

Thus *end point energy* is the maximum energy with which a  $\beta$ -particle is emitted from a radioactive nuclide.

**Results.** The results of these measurements show that there are apparently two distinct types of  $\beta$ -ray spectrum, one a *sharp line spectrum* and the other a *continuous spectrum*.

- The lines correspond to  $\beta$ -particles emitted with discrete energies.

The sharp line spectra are due to electrons that have been ejected from the  $K$ ,  $L$ ,  $M$  and  $N$  shells of the atom by the process of *internal conversion*.

The continuous spectrum of  $\beta$ -particle energies found for RaE is shown in Fig. 20.18. In this case, no line spectrum is found. The upper limit or maximum energy is at 1.17 MeV.

● When a line spectrum is also present, the lines appear as distinct peaks superimposed upon the continuous distribution curve, as in Fig. 20.19. Fig. 20.19 shows Beta-particle spectrum of  $Au^{198}$ .

- The continuous spectrum is due to a continuous spread of energy among the emitted  $\beta$ -particles having a fixed maximum value of the energy. The upper energy limit of the continuous  $\beta$ -particle spectrum is characteristic of the radioactive atom concerned.
- The continuous  $\beta$ -ray spectrum is that produced by the electrons that have been ejected from the nuclei of radioactive atoms.

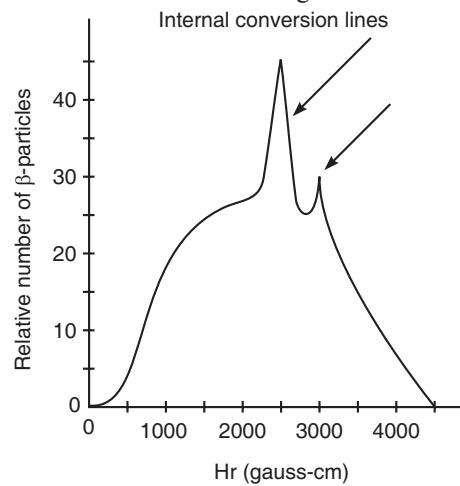


Fig. 20.19

#### 20.10.1. Origin of the Continuous Spectrum

The explanation of the existence of the continuous energy spectrum for the  $\beta$ -particles presents some difficulties.

(1) **Law of conservation of energy.** When a nucleus emits a  $\beta$ -particle, a neutron in the nucleus changes to a proton. Hence the atomic number increases by unity and the mass number remains the same.

Let  $M_1$  = mass of the *neutral parent atom* of atomic number  $Z$ ,

$M_2$  = mass of *neutral daughter atom* of atomic number  $(Z + 1)$ ,

$m$  = mass of the  $\beta$ -particle and  $e$  = Charge on the  $\beta$ -particle.

Then, according to the principle of mass energy,

$$\left. \begin{aligned} & \text{Rest mass of} \\ & \text{the parent nucleus} \end{aligned} \right\} = \text{Rest mass of daughter nucleus} + \text{Rest mass of electron} \\ + \text{Energy of the electron.}$$

$$\text{or } (M_1 - Zm) c^2 = [M_2 - (Z + 1)m] c^2 + mc^2 + Q \\ \therefore \text{Energy of the electron} = Q = (M_1 - M_2) c^2.$$

Hence all the  $\beta$ -particles from a given radioactive substance must be emitted with the same K.E. But actual measurements show that only a few  $\beta$ -particles are emitted with this maximum value of energy. The majority of  $\beta$ -particles are emitted with smaller energies. What happens to the remaining energy?

(2) **Law of conservation of angular momentum.** Another difficulty comes in the conservation of angular momentum. Every nucleus has an angular momentum (nuclear spin) which is an odd multiple of  $\frac{1}{2}\hbar$  for nuclei of odd mass number and an even multiple of  $\frac{1}{2}\hbar$  for nuclei of even mass number. The electron has an angular momentum  $\frac{1}{2}\hbar$ . In  $\beta$ -decay, mass number remains unchanged.

How is it possible for a nucleus of even mass number and therefore an integral spin to give rise to a daughter nucleus of the same mass number and also an integral spin and yet emit an electron of spin  $\frac{1}{2}\hbar$ ? The same is the difficulty for a nucleus of odd mass number.

(3) There is also an apparent failure to conserve linear momentum in  $\beta$ -decay.

#### **Pauli's Neutrino Hypothesis**

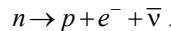
In 1930, Pauli proposed that if an uncharged particle of zero mass and spin  $\frac{1}{2}$  is emitted in

$\beta$ -decay together with the electron, the energy, angular momentum and linear momentum discrepancies discussed above would be removed. The particle was named *neutrino*. It was supposed that neutrino carries off an energy equal to the difference between  $Q$  and the actual electron K.E. Subsequently it was found that there are two kinds of neutrino involved in  $\beta$ -decay, the neutrino itself (symbol  $\nu$ ) and the anti-neutrino (symbol  $\bar{\nu}$ ). The reason neutrinos were not experimentally detected until recently is that their interaction with matter is extremely feeble. Lacking charge and mass, and not electromagnetic in nature, the neutrino can pass unimpeded through vast amounts of matter. A neutrino would have to pass through over 100 light-years of solid iron on the average before interacting.

#### **20.10.2. The Neutrino Theory of Beta Decay**

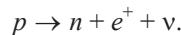
In 1934, Fermi developed a theory to explain the continuous  $\beta$ -ray spectrum. This theory is called neutrino theory of  $\beta$ -decay. According to this theory, a  $\beta$ -particle and a neutrino are created in the nucleus and both are emitted simultaneously. The *total energy* of these two particles is a constant which is equal to the end-point energy observed in the  $\beta$ -ray spectrum. This maximum energy is shared by the  $\beta$ -particle, the neutrino and also by the recoiling nucleus. The electron will carry the maximum energy when the energy of the neutrino is zero. In all other cases, electron will carry an energy less than the maximum. The sum of the energies carried by the electron and the neutrino will always be the same. This energy may be shared by the two particles in any proportion. Hence it explains the continuous  $\beta$ -ray spectrum.

When the nucleon shifts from the neutron quantum state to the proton quantum state, electron and antineutrino are emitted. This process is represented by

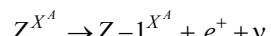
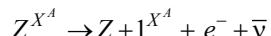


In ordinary beta decay, it is an antineutrino that is emitted.

Positron emission corresponds to the conversion of a nuclear proton into a neutron, a positron, and a neutrino.



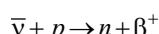
Positron emission leads to a daughter nucleus of lower atomic number  $Z$  while leaving the mass number  $A$  unchanged. Thus negative and positive beta decays may be represented as



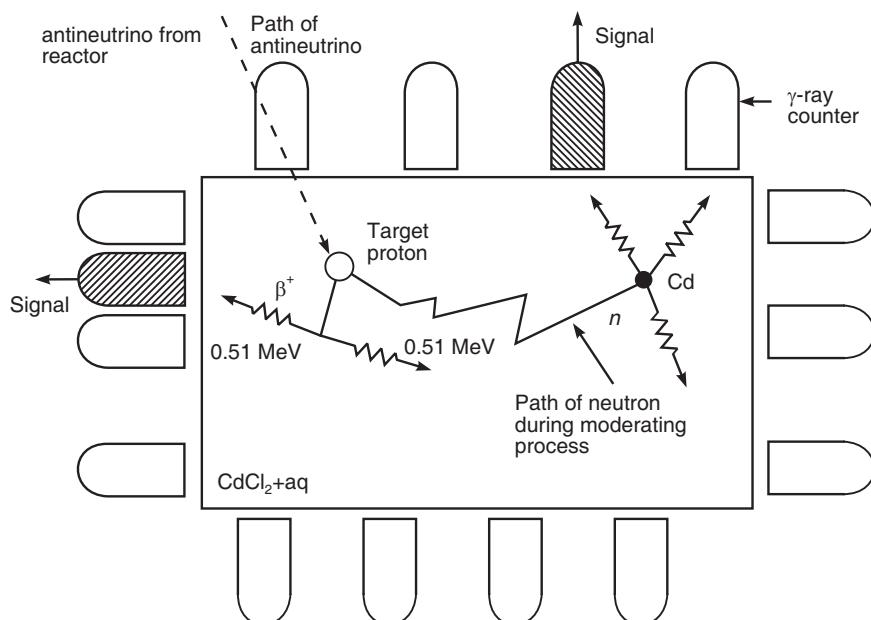
The electron, neutrino and product nucleus share among them the energy, angular momentum and linear momentum available from the nuclear transitions. Thus the neutrino theory of  $\beta$ -decay successfully explains the continuous energy spectrum of  $\beta$ -rays.

### 20.10.3. Detection of Neutrino

Reines and Cowan made use of the abundant supply of antineutrinos emitted from a nuclear reactor. Energetic antineutrinos can be detected by the reaction

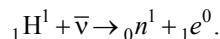


The experimental arrangement is shown in Fig. 20.20.

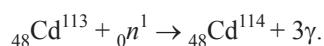


**Fig. 20.20**

An intense beam of antineutrinos, proceeding from a nuclear reactor is allowed to traverse a large plastic tank filled with an aqueous solution of cadmium chloride,  $CdCl_2$ . The tank is surrounded by many photomultiplier tubes. Suppose an antineutrino is absorbed by a proton and this system decays by  $\beta^+$  emission to become a neutron. Thus



The positron combines with an electron to create two  $\gamma$  photons, which are detected by the photomultiplier tubes. The neutron, moderated, *i.e.*, slowed down by collisions with protons, is finally captured by a cadmium nucleus when three  $\gamma$  photons are emitted. Thus,



The almost simultaneous detection of these  $\gamma$  photons by the tubes established, beyond all reasonable doubt, the existence of antineutrinos.

**EXAMPLE :** Which of the following isobars would you expect to be  $\beta^-$  active and how would it decay? Why?  $_{28}^{64}\text{Ni} = 63.9280\text{u}$ ,  $_{29}^{64}\text{Cu} = 63.9298\text{u}$ .

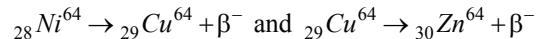
**SOL.** We know that for a  $\beta^-$  decay to be possible

$$_Z^A M^A > {}_{Z+1}^{A+1} M^A + {}_{-1}^0 e^0 \quad (\text{Nuclear masses}).$$

By adding Z electron masses to both sides, the nuclear masses are changed into isotopic masses. The same condition now appears as

$$_Z^A M^A > {}_{Z+1}^{A+1} M^A \quad (\text{Atomic masses}).$$

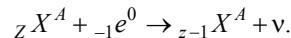
The given isobars are  $\beta^-$  emitters, i.e.,



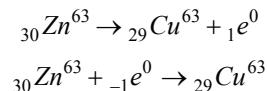
if  ${}_Z^A M^A > {}_{Z+1}^{A+1} M^A$ . Now mass of  $\text{Cu}^{64}$  is greater than that of  $\text{Ni}^{64}$ . Hence  $\text{Ni}^{64}$  is not a  $\beta^-$  emitter.

Mass of  $\text{Cu}^{64}$  is greater than that of  ${}_{30}^{64}\text{Zn}$  (63.92915 u). Hence  $\text{Cu}^{64}$  is a  $\beta^-$  emitter.

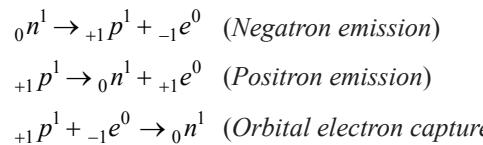
**K-electron capture.** In an excited nucleus, there is a possibility of a proton turning into a neutron by absorbing an electron, instead of emitting a positron. In this case, an orbital electron of the parent nuclide, usually an electron in the K-shell which is closest to the nucleus, interacts with a proton in the nucleus, converting it into a neutron i.e., the K-electron falls into the nucleus. The general basic equation is



The daughter nuclide is seen to be the same as that which would have been produced if positrons had been emitted. There is no question of any electron coming out from this process. The vacancy in the K-shell of the daughter nuclide resulting from K-capture is filled by the rearrangement of the Z-1 electrons in it. This leads to the emission of the characteristic X-rays of the daughter element. We also note that in  $\beta$  activity and in K-capture processes, the parent and the daughter nuclides form a pair of isobars. Some examples of the lighter radioactive nuclides which undergo electron capture are :  $\text{Be}^7$ ,  $\text{Al}^{24}$ ,  $\text{Al}^{26}$ ,  $\text{Ar}^{37}$ ,  $\text{K}^{40}$ ,  $\text{Ca}^{41}$ ,  $\text{V}^{48}$ ,  $\text{V}^{49}$ ,  $\text{Cr}^{51}$ ,  $\text{Fe}^{55}$  and several others. In many cases,  $\gamma$ -radiation is also emitted.  $\gamma$ -radiation is emitted after electron capture if the nucleus is left in an excited state. In a few cases, both positron emission and electron capture occur. Example :  ${}_{30}^{63}\text{Zn} \rightarrow {}_{29}^{63}\text{Cu} + {}_{-1}^0 e^0$



The three modes of beta radioactivity can be shown by the following reactions :



## 20.11 NON-CONSERVATION OF PARITY IN BETA DECAY

**Non-conservation of parity by neutrino.** Prior to 1957, parity was assumed to remain conserved in all nuclear interactions, implying that the mirror image of an object or process is not distinguishable from the object or the process itself. In 1957, Lee and Yang hinted at the non-conservation of parity in  $\beta$ -decay.

Consider the mirror image of a neutrino. Let the neutrino travel toward a mirror with momentum  $\vec{p}_v$ . Its spin  $\vec{s}_v$  is directed opposite to  $\vec{p}_v$  (Fig. 31.27). Clearly, the mirror image of this neutrino is

a different particle — the anti-neutrino — because in the mirror image, the vectors  $\vec{s}_v$  and  $\vec{p}_v$  are in the same direction. Thus, the motion of the neutrino as given by the mirror image is not a possible motion for it; it refers to a different particle viz., an antineutrino. This is known as *non-conservation of parity* by neutrino.

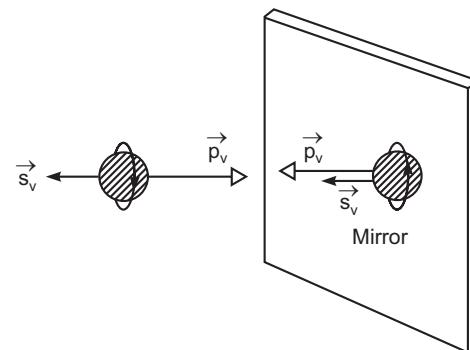


Fig. 20.21

### Experimental verification of non-conservation of parity in $\beta$ -decay

The violation of parity in the process of  $\beta$ -decay was experimentally demonstrated by *Wu et al.* in their demonstration of the asymmetric  $\beta$ -particle distribution from polarized cobalt nuclei.

**Experimental arrangement.** Fig. 20.22 shows the experimental set up used by *Wu et al.*, in their demonstration of the asymmetric  $\beta$ -particle distribution from polarized cobalt nuclei.

- The source consists of a thin layer of  $\text{Co}^{60}$  grown as a crystalline layer on top of good single crystal of cerium-magnesium-nitrate. Cobalt-60 emits an electron of energy 0.312 Mev in over 99% of its decays and two other electrons in the remaining decays. Several  $\gamma$ -rays are also emitted. Polarised samples of  $\text{Co}^{60}$  nuclei are prepared by applying the magnetic fields at low temperatures. The sample is cooled to 4.2 K by liquid helium and further cooled to below 0.01 K by adiabatic demagnetization.

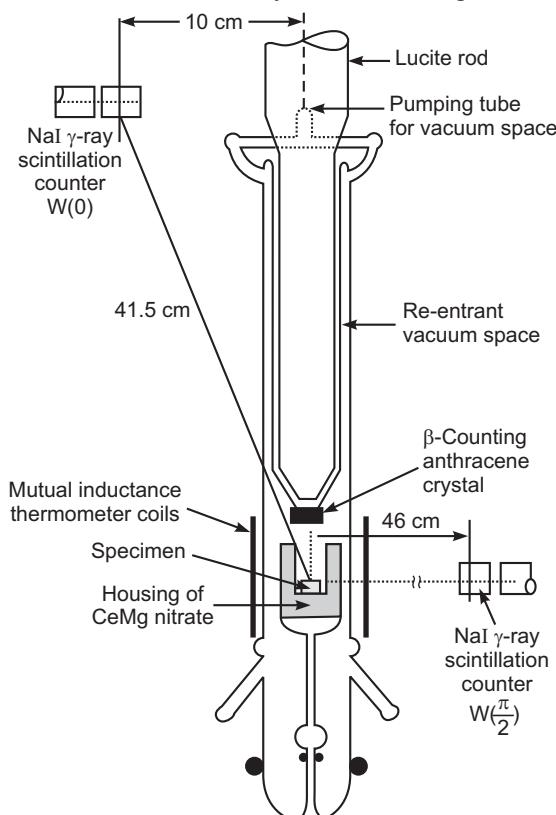


Fig. 20.22

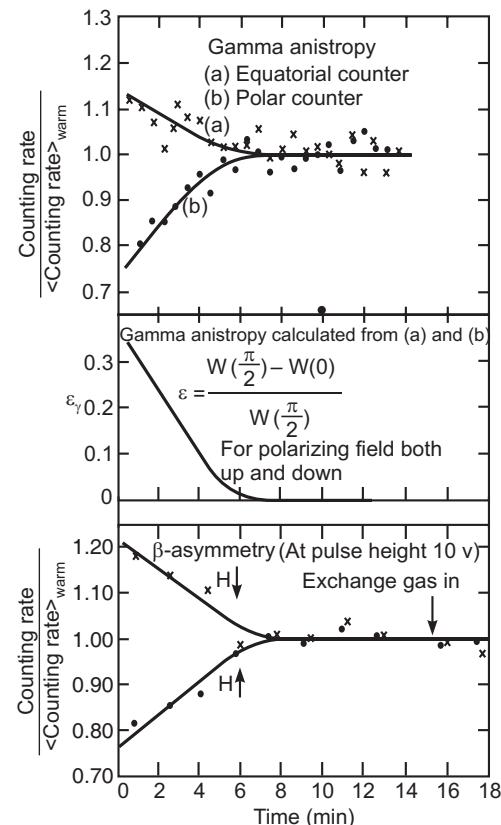


Fig. 20.23

- The beta-particles emitted by the polarized  $^{60}\text{Co}$  nuclei are detected by an anthracene crystal mounted 2 cm above the source. The space between them is a vacuum. The light flashes in the scintillator, produced by the beta-rays, are transmitted through a lucite light pipe to a photomultiplier tube mounted outside the cryostat.
- To measure the extent of the polarization of the  $\text{Co}^{60}$  nuclei, two NaI  $\gamma$ -ray scintillation counters are used, one in the equatorial plane and one near the polar position. The observed  $\gamma$ -ray anisotropy provided a measure of the polarization and hence of the temperature of the sample.

**Experiment.** After the material in the cryostat is cooled to about 0.01 K, the magnet used in the cooling process is turned off. A vertical solenoid is raised around the lower part of the cryostat within 20 sec after the demagnetization so that the warming starts. The counting is started.

**The results of the experiment.** Fig. 20.23 shows the results of the experiment.

(i) **Gamma-ray anisotropy.** The *time scale is actually a temperature scale measuring time from the instant when cooling is stopped*. The curve labelled "gamma-anisotropy" is a measure of the extent of polarization. After about 8 min the nuclei have become sufficiently warm so that their thermal motion causes the alignment to become random, and the  $\gamma$ -anisotropy disappears.

(ii) **Beta asymmetry.** The curve labelled " $\beta$  asymmetry" is the significant one. It indicates the number of electrons emerging in the direction of the magnetic field (the direction in which the nuclei are aligned), and the number emerging in the opposite direction. It is evident from the figure that more electrons are emitted in the direction opposite that of the magnetic field, *i.e.*, in the direction opposite to that in which the nuclei are aligned. The electrons are emitted in a preferred direction, the principle of right-left symmetry is violated, and parity is not conserved. The size of the effect is remarkable. The electron intensity in one direction along the axis of rotation was found to be 40% greater than in the opposite direction. There was no doubt about the result.

## GAMMA RAY SPECTRA

### 20.12 DETERMINATION OF THE WAVELENGTH OF GAMMA RAYS—

#### Du Mond Curved Crystal Spectrometer

**Principle.** The wavelength of  $\gamma$ -rays is determined by using the Bragg equation,  
 $2d \sin \theta = n\lambda$ .  $\gamma$ -ray energy is calculated using the relation,

$$E = \frac{ch}{\lambda}.$$

Here,  $E$  is the gamma ray energy.

**Construction.** Fig. 20.24 shows the apparatus. The quartz diffracting crystal  $C$  is bent and clamped in a steel frame so that the diffracting planes meet, when extended, in a line  $B$  normal to the plane of the figure. The radius of curvature of the crystal is then equal to the diameter of the focusing circle  $F$ . The  $\gamma$ -ray source is kept on the focusing circle at point  $S$ . A detector like a scintillation counter is placed behind the crystal.

If the Bragg-condition ( $2d \sin \theta = n\lambda$ ) is satisfied, there will be a diffracted beam which will enter the detector (scintillator) as if it comes from a virtual source  $V$ . A lead collimator, which is cut with various channels at different angles is used to prevent the intense direct transmitted radiations

from reaching the scintillator. But the lead collimator permits the diffracted rays at angle  $2\theta$  w.r.t. the direct beam to reach the scintillator.

**Working.** The counting rate is measured as a function of the source position (angle  $\theta$ ).  $S$  is moved on a circle and the output of the detector is plotted against  $\theta$ . At the diffraction maxima, peaks are found. From the value of  $\theta$  corresponding to the peaks, the value of  $\lambda$  is calculated from Bragg's formula,  $2d \sin \theta = n\lambda$ .

**Advantages.** (1) For low energy  $\gamma$ -rays, the method is very precise and measures with an accuracy of 1 part in  $10^4$ .

(2) This method avoids the necessity of collimating the incident  $\gamma$ -rays and makes full use of the aperture of the crystal.

**Disadvantages** (1) The measurements become more difficult and less precise as the energy of the  $\gamma$ -rays increases and the wavelength decreases. The method is not very precise for  $\gamma$ -ray energies above about 1 MeV.

(2) This method requires highly active sources which are often not available.

(3) The intensity of  $\gamma$ -rays diffracted by the lattice planes of a crystal increases with the thickness of the crystal. There is an optimum thickness beyond which the absorption of  $\gamma$ -rays becomes significant. It is also difficult to bend large quartz crystals which are thicker.

### 20.13 ORIGIN OF $\gamma$ -RAYS

The large values of energy associated with  $\gamma$ -rays show that they must be of nuclear origin.  $\gamma$ -ray spectrum consists of sharp lines and this indicates the existence of a number of energy levels in the nucleus. Stable nuclides are usually in the state of least energy or ground state, but they can be excited by particle or photon bombardment. Hence nuclei can exist in states of definite energies, just as atoms can. An excited nucleus is denoted by an asterisk (\*) after its usual symbol. Thus  ${}_{38}^{87*}\text{Sr}$  refers to  ${}_{38}^{87}\text{Sr}$  in an excited state. One way an excited nucleus can return to the ground state is by the emission of  $\gamma$ -rays.  $\gamma$ -ray decay is represented schematically by

$$({}_{Z}X^A)^* \rightarrow {}_Z X^A + \gamma.$$

The star (\*) indicates an excited nucleus, and both the daughter and the parent have the same structure of nuclear particles. If  $E^*$  is the energy associated with the excited state and  $E$  is the energy of the ground state, then the  $\gamma$ -rays have an energy

$$h\nu = E^* - E.$$

Here  $\nu$  is the frequency of the emitted  $\gamma$ -ray.

A simple example of the relationship between energy levels and decay schemes is shown in Fig. 20.25 which pictures the  $\beta$ -decay of  ${}_{12}^{27}\text{Mg}$  to  ${}_{13}^{27}\text{Al}$ . The half-life of the decay is 9.5 minutes, and it may

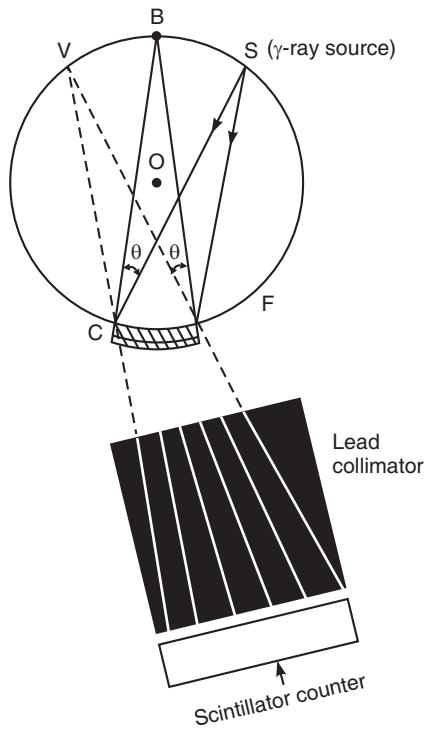


Fig. 20.24

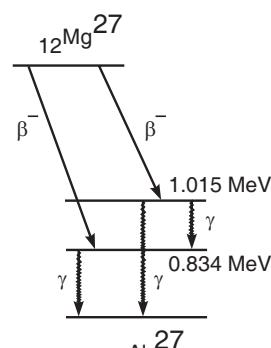


Fig. 20.25

take place to either of the two excited states of  $_{13}Al^{27}$ . The resulting  $_{13}Al^{27*}$  nucleus then undergoes one or two gamma decays to reach the ground state.

Most excited nuclei have very short half-lives against  $\gamma$ -decay, but a few remain excited for as long as several hours. A long lived excited nucleus is called an *isomer* of the same nucleus in its ground state. The excited nucleus  $_{38}Sr^{87*}$  has a half-life of 2.8 hours and is an isomer of  $_{38}Sr^{87}$ .

## 20.14 NUCLEAR ISOMERISM

There are nuclei which have the same atomic and mass numbers (same  $Z$  and same  $A$ ) but differ from one another in their nuclear energy states and exhibit differences in their internal structure. These are called *nuclear isomers*.

The existence of nuclear isomers is called *nuclear isomerism*. The excited nucleus  $_{38}Sr^{87*}$  is an isomer of  $_{38}Sr^{87}$ . The difference between the nuclear isomers is attributed to a difference of nuclear energy states. One isomer represents the nucleus in its ground state, whereas the other is the same nucleus in an excited state of higher energy. The phenomenon of nuclear isomerism was discovered by O. Hahn in 1921. He found that  $UX_2$  and  $UZ$  both have the same atomic number and the same mass number but have different half-lives and emit different radiations.  $UX_2$  has 0.394 MeV more energy in its nucleus than  $UZ$ . Both these nuclei are formed out of  $UX_1$  by  $\beta$ -decay.  $UX_2$  has half-life of 1.17 minutes and  $UZ$  has a half-life of 6.7 hours. The higher energy isomer  $UX_2$  may directly decay to  $UII$  by  $\beta$  emission with a half-life of 1.17 minutes, or it may first come to the lower energy isomer by emitting a  $\gamma$ -ray of energy 0.394 MeV and then decay to  $UII$  by  $\beta$ -emission with a half-life of 6.7 hours.

Nuclear isomerism has also been detected in artificial radioactive substances. Many *isomeric pairs* have been produced by bombarding radionuclides with neutrons.

## 20.15 INTERNAL CONVERSION

When a nucleus passes from a higher excited state to the ground state, the difference in energy of the two states is emitted as a  $\gamma$ -ray. As an alternative to  $\gamma$ -decay, an excited nucleus, in some cases, may return to its ground state by giving up its excitation energy to one of the orbital electrons around it. The emitted electron has a K.E. equal to the lost nuclear excitation energy minus the binding energy of the electron in the atom. *i.e.*,

$$\text{K.E. of the ejected electron} = E_e - W.$$

Here,  $E_e$  = the available excitation energy and

$W$  = binding energy of the ejected electron in its shell of origin.

This process is called *internal conversion*. The emitted electron is called a *conversion electron*. Thus internal conversion and emission of a  $\gamma$ -ray from the nucleus are two alternate ways of accomplishing the same nuclear transition. The internal conversion is not a *two step* process in which a  $\gamma$ -ray photon is first emitted and then it knocks out an orbital electron. It is in better accord with experiment to regard internal conversion as representing a direct transfer of excitation energy from a nucleus to an orbital electron.

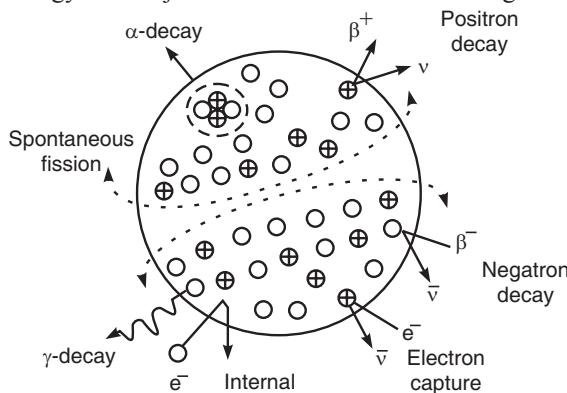


Fig. 20.26

Hence, internal conversion is a *single step* process in which the excited nucleus interacts directly with the orbital electron. The energy of the ejected electron ( $\beta$ -particle) has discrete values. Therefore, the corresponding  $\beta$ -particle energy spectrum is a line spectrum having discrete energies.

Fig. 20.26 illustrates the various kinds of disintegration processes that radioactive nuclei may undergo. The nucleus is represented as an assembly of protons and neutrons. A proton is indicated by a cross, and a neutron by an open circle.

## 20.16 MOSSBAUER EFFECT

**The Mossbauer Effect.** The emission of gamma rays is generally accompanied by the emission of an  $\alpha$  or  $\beta$ -particle. If after the emission of an  $\alpha$  or  $\beta$ -particle the product nucleus is left in an excited state, it reaches the ground state by releasing or emitting photons called  $\gamma$ -rays. When a nucleus emits a photon it recoils in the opposite direction. This reduces the energy of the  $\gamma$ -rays from its usual transition energy  $E_0$  to  $E_0 - R$ , where  $R$  is the recoil energy.

In certain crystals, the atoms are tightly bound to the lattice. When such an atom emits a  $\gamma$ -ray, the entire crystal and not merely the emitting atom recoils as a whole. In this case the recoil energy is not transferred to *one* emitting atom, but to a *group* of atoms. The whole mass of crystal — not just the recoiling particle — takes part in the recoil. This implies that the recoiling mass is extremely large and the *recoil velocity is almost zero*. Now the  $\gamma$ -ray photon has the transition energy  $E_0$ . The emission of  $\gamma$ -rays without recoil, and therefore without reduction of the  $\gamma$ -ray energy, is called the *Mössbauer effect*.

**Experimental arrangement.** Fig. 20.27 (a) shows the experimental arrangement.

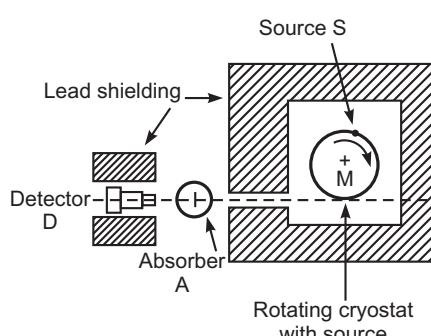


Fig. 20.27 (a)

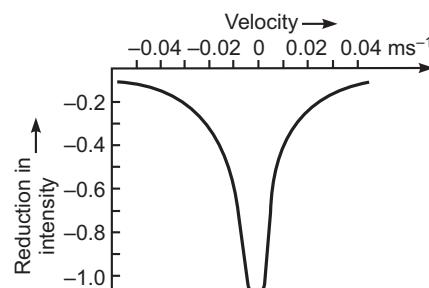


Fig. 20.27 (b)

- $S$  is the source at the rim of rotating cryostat. The source  $\text{Os}^{191}$  decays by  $\beta^-$  emission to an excited state of stable  $\text{Ir}^{191}$  with energy 0.129 MeV.  $M$  is the region in which the source is seen from the scintillation detector.
- $A$  is the absorber in a cryostat. The absorber is  $\text{Ir}^{191}$  (a foil of natural crystalline iridium).
- $D$  is a scintillation detector. Lead blocks are used for shielding purposes.

**Procedure.** The recoil-free  $\gamma$ -rays from the source  $S$  are collimated through a channel in a  $Pb$ -block that encloses the source. The radiation falls on the absorber  $A$ . Both the source and the absorber are placed in two cryostats maintained at a temperature of 88 K. The cooling reduces the thermal broadening to an enormous extent. The  $\gamma$ -photons transmitted by  $A$  are detected with the scintillation counter  $D$ .

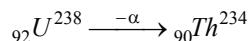
The cryostat containing  $S$  is mounted on a lathe carriage and is set into rotation. The detector is turned on only when the source is in line with the absorber  $A$ . The Doppler shift in the frequency of  $\gamma$ -photon offsets the resonance absorption by the stationary absorber  $A$ .

**Resonance Curve.** Fig. 20.27 (b) shows the variation of  $\gamma$ -intensity with relative velocity of source and absorber. The faster the motion of the source, greater is the mismatch between the emission and absorption frequencies and so less is the absorption in the foil or more is the number of  $\gamma$ -photons reaching the counter C. The resonance absorption of  $\gamma$ -rays is so sharp that even a relative speed of  $0.02 \text{ ms}^{-1}$  reduces the absorption by more than 50% at 88 K. The curve in Fig. 20.27(b) corresponds to the theoretical resonant nuclear absorption for 129 kV  $\gamma$ -photons from  $^{191}\text{Ir}$ , based on a natural width of  $5 \times 10^{-6} \text{ eV}$  for its excited state. The experimental points of Mössbauer fall close to this curve.

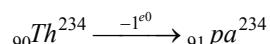
## FUNDAMENTAL LAWS OF RADIOACTIVITY

### 20.17 SODDY FAJAN'S DISPLACEMENT LAW

- (1) In all known radioactive transformations either an  $\alpha$  or a  $\beta$ -particle (i.e., never both or more than one of each kind) is emitted by the atom.
- (2) When a radioactive atom emits an  $\alpha$  particle, a new atom is formed whose mass number is less by four units and atomic number less by two units than those of the parent atom.



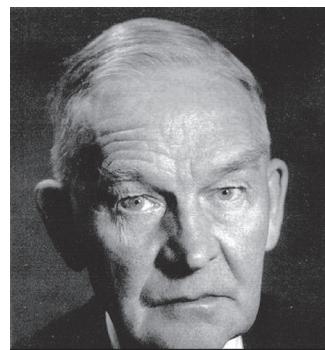
- (3) When a radioactive atom emits a  $\beta$ -particle, the new atom formed has the same mass number but the atomic number is increased by one unit.



#### Natural Radioactive Series

With the help of the displacement law, one can determine easily the mass and atomic numbers of the different elements in the successive radioactive changes, if the mass and atomic numbers of the parent are known. The different atoms produced at the different stages of disintegration will naturally form a *series*.

The first member is called the *parent*, the intermediate members are called *daughters* and the final stable member is called the *end product*. (1) Uranium series (2) Actinium series (3) Thorium series and (4) Neptunium series are four such series known at present. Each of the series consists of a succession of daughter elements which are all derived from a single parent nuclide. An explanation can be given as to why there are exactly four series. We know that alpha decay reduces the mass number of a nucleus by 4. Thus the nuclides, whose mass numbers are all given by  $A = 4n$ , where  $n$  is an integer, can decay into one another in descending order of mass number. All radioactive nuclides whose mass numbers obey the equation  $A = 4n$  are said to be members of the  $4n$  series. Radioactive nuclides whose mass numbers obey the equation  $A = 4n + 1$  belong to the  $(4n + 1)$  series. Similarly radioactive nuclides whose mass numbers satisfy the equation  $A = 4n + 2$  and  $A = 4n + 3$  belong to the  $(4n + 2)$  series and  $(4n + 3)$  series respectively.



Frederick Soddy

Table 20.1. Four radioactive series

Mass numbers	Series	Parent	Half-life, years	Stable end product
$4n$	Thorium	$_{90}Th^{232}$	$1.39 \times 10^{10}$	$_{82}Pb^{208}$
$4n + 1$	Neptunium	$_{93}Np^{237}$	$2.25 \times 10^6$	$_{83}Bi^{209}$
$4n + 2$	Uranium	$_{92}U^{238}$	$4.51 \times 10^9$	$_{82}Pb^{206}$
$4n + 3$	Actinium	$_{92}U^{235}$	$7.07 \times 10^8$	$_{82}Pb^{207}$

Table 20.1 is a list of the names of four important radioactive series, their parent nuclides and the half-lives of these parents, and the stable daughters which are end products of the series.

## 20.18 LAW OF RADIOACTIVE DISINTEGRATION

Let  $N$  be the number of atoms present in a particular radioelement at a given instant  $t$ . Then the rate of decrease  $-dN/dt$  is proportional to  $N$ .

$$\therefore -\frac{dN}{dt} = \lambda N \quad \dots(1)$$

Here  $\lambda$  is a constant known as the *disintegration constant* or *decay constant* of the radioactive element. It is defined as the ratio of the amount of the substance which disintegrates in a unit time to the amount of substance present.

Eq. (1) can be written as  $\frac{dN}{N} = -\lambda dt$ .

$$\text{Integrating, } \log_e N = -\lambda t + C \quad \dots(2)$$

Let the number of radioactive atoms initially present be  $N_0$ .

Then, when  $t = 0$ ,  $N = N_0$ .  $\therefore \log_e N_0 = C$ .

Substituting for  $C$  in (2), we get,  $\log N = -\lambda t + \log N_0$

$$\text{or } \log_e \frac{N}{N_0} = -\lambda t$$

$$\text{or } N = N_0 e^{-\lambda t} \quad \dots(3)$$

This equation shows that the number of atoms of a given radioactive substance decreases exponentially with time (Fig. 20.28).

**Definition of half-life period.** The half-life period of a radioactive substance is defined as the time required for one-half of the radioactive substance to disintegrate.

### Value of half-life period

We know the relation  $N = N_0 e^{-\lambda t}$ .

If  $T_{1/2}$  be the half-life period, then at

$$t = T_{1/2}, N = N_0/2.$$

$$\therefore \frac{N_0}{2} = N_0 e^{-\lambda T_{1/2}} \text{ or } e^{\lambda T_{1/2}} = 2$$

$$\text{or } \lambda T_{1/2} = \log_e 2 \text{ or } T_{1/2} = \frac{\log_e 2}{\lambda}$$

$$\therefore T_{1/2} = \frac{\log_e 2}{\lambda} = \frac{0.6931}{\lambda}$$

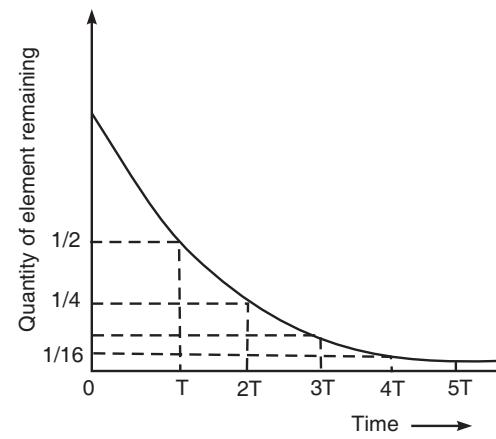


Fig. 20.28

**EXAMPLE 1.** The half-value period of radium is 1590 years. In how many years will one gram of pure element (a) lose one centigram, and (b) be reduced to one centigram?

**SOL.** Here, half-life period of radium =  $T_{1/2} = 1590$  years.

$$\therefore \text{Radioactive constant } = \lambda = \frac{0.6931}{T_{1/2}} = \frac{0.6931}{1590}.$$

(a) Let  $t$  be the time in which one gram of radium loses one centigram (0.01g).

$$\therefore \text{Radium left behind} = 1 - 0.01 = 0.99 \text{ gram.}$$

$$\text{Now, } N = N_0 e^{-\lambda t} \text{ or } \log_e N = \log_e N_0 - \lambda t$$

$$\text{or } \lambda t = \log_e \left( \frac{N_0}{N} \right)$$

$$\therefore t = \frac{1}{\lambda} \log_e \left( \frac{N_0}{N} \right) = \frac{1590}{0.6931} \log_e \left( \frac{1}{0.99} \right)$$

$$= 23.25 \text{ years.}$$

(b) Here,  $N = 0.01$  gram;  $N_0 = 1$  gram;  $t = ?$

$$t = \frac{1}{\lambda} \log_e \left( \frac{N_0}{N} \right) = \frac{1590}{0.6931} \log_e \left( \frac{1}{0.01} \right) = 10560 \text{ years.}$$

**EXAMPLE 2.** 1 gram of radium is reduced by 2.1 mg in 5 years by  $\alpha$ -decay. Calculate the half-life period of radium. (B.U. 2011)

**SOL.** Initial mass of radium = 1 gram

Mass of radium left behind after 5 years =  $1 - 0.0021 = 0.9979$  g.

$$\text{We have, } N = N_0 e^{-\lambda t}; \text{ Here, } \frac{N}{N_0} = \frac{0.9979}{1}; t = 5 \text{ years.}$$

$$\therefore 0.9979 = e^{-5\lambda} \text{ or } e^{5\lambda} = \frac{1}{0.9979}$$

$$\text{or } 5\lambda = \log_e \left( \frac{1}{0.9979} \right)$$

$$\therefore \lambda = \frac{1}{5} \log_e \left( \frac{1}{0.9979} \right) = \frac{2.3026}{5} \log_{10} \left( \frac{1}{0.9979} \right) \\ = 41.45 \times 10^{-5} \text{ per year.}$$

$$T_{1/2} = \frac{0.6931}{\lambda} = \frac{0.6931}{41.45 \times 10^{-5}} = 1672 \text{ years.}$$

**EXAMPLE 3.** Calculate the time required for 10% of a sample of thorium to disintegrate. Assume the half-life of thorium to be  $1.4 \times 10^{10}$  years. (B.U. 2012)

**SOL.** We have,  $N = N_0 e^{-\lambda t}$ . Here  $N = 0.9 N_0$ .

$$\lambda = \frac{0.6931}{T_{1/2}} = \frac{0.6931}{1.4 \times 10^{10}} a^{-1}.$$

$$\therefore 0.9 N_0 = N_0 e^{-\lambda t} \text{ or } \lambda t = \log_e \left( \frac{1}{0.9} \right) \text{ or } t = \frac{1}{\lambda} \log_e \left( \frac{1}{0.9} \right)$$

$$\begin{aligned} t &= \frac{1.4 \times 10^{10}}{0.6931} \times \log_e 1.111 = \frac{1.4 \times 10^{10}}{0.6931} \times 2.302 \times \log_{10} 1.111 \\ &= 2.1 \times 10^9 \text{ years.} \end{aligned}$$

### 20.18.1. The Mean Life

It is not possible to predict which atom of a radioactive substance will disintegrate at any instant. The atom which disintegrates first has zero life and that disintegrates last has infinite life. Thus the life of every atom is different and the actual lives of the various atoms range from zero to infinity.

$$\therefore \text{The mean-life of a radioactive element} = \frac{\text{Sum of the lives of all the atoms}}{\text{Total number of atoms}}.$$

**Definition.** *The mean-life of a radioelement is defined as the ratio of the total life time of all the radioactive atoms to the total number of such atoms in it.*

**Value of mean-life.** Let  $N_0$  be the total number of radioactive atoms in the beginning. Let  $N$  be the number of atoms of that element after time  $t$ . Then  $N = N_0 e^{-\lambda t}$ . Let  $dN$  be the number of atoms disintegrating between time  $t$  and  $t + \delta t$ . These  $dN$  atoms have had a life between  $t$  and  $(t + \delta t)$ . Since  $\delta t$  is very small, each of these atoms had a life of  $t$ .

$$\therefore \text{total life of } dN \text{ atoms} = (dN)t.$$

The possible life of any of the total number  $N_0$  radioactive atoms varies from 0 to  $\infty$ .

$$\therefore \text{Total life time of all } N_0 \text{ atoms} = \int_0^\infty t dN$$

$$\text{Now, mean life} = \bar{T} = \frac{\text{total life - time}}{\text{total number of atoms}}$$

$$= \frac{\int_0^\infty t dN}{N_0}$$

Now,

$$N = N_0 e^{-\lambda t}$$

$$\therefore \frac{dN}{dt} = -\lambda N_0 e^{-\lambda t}$$

or

$$\begin{aligned} dN &= -\lambda N_0 e^{-\lambda t} dt \\ dN &= \lambda N_0 e^{-\lambda t} dt \end{aligned}$$

(Leaving the negative sign which merely indicates the decrease in the number of atoms with time).

$$\text{Hence, } \bar{T} = \frac{\int_0^\infty t \lambda N_0 e^{-\lambda t} dt}{N_0} = \lambda \int_0^\infty t e^{-\lambda t} dt$$

Integrating by parts,

$$\bar{T} = \lambda \left[ \frac{te^{-\lambda t}}{-\lambda} - \int \frac{e^{-\lambda t} dt}{-\lambda} \right]_0^\infty$$

$$\begin{aligned}
 &= \lambda \left[ \frac{te^{-\lambda t}}{-\lambda} - \frac{e^{-\lambda t}}{-\lambda^2} \right]_0^\infty = \lambda \left( \frac{1}{\lambda^2} \right) = \frac{1}{\lambda} \\
 \therefore \bar{T} &= \frac{1}{\lambda}.
 \end{aligned}$$

Thus the mean life ( $\bar{T}$ ) of a radioactive substance is the reciprocal of the decay constant ( $\lambda$ ).

**EXAMPLE 1.** 1 gram of a radioactive substance disintegrates at the rate of  $3.7 \times 10^{10}$  disintegrations per second. The atomic weight of the substance is 226. Calculate its mean life.

**SOL.** Number of atoms disintegrated in one second =  $3.7 \times 10^{10}$

$$\begin{aligned}
 \text{The mass of the substance disintegrated in one second} &= \frac{(3.7 \times 10^{10}) \times 226}{6.02 \times 10^{26}} \\
 &= 1.389 \times 10^{-14} \text{ kg}
 \end{aligned}$$

Here,  $N = 1 \text{ g} = 10^{-3} \text{ kg}$ ;  $-dN/dt = 1.389 \times 10^{-14} \text{ kg}$

We have,  $-dN/dt = \lambda N$

$$\text{or } \lambda = \frac{-dN/dt}{N} = \frac{1.389 \times 10^{-14}}{10^{-3}} = 1.389 \times 10^{-11} \text{ s}^{-1}$$

$$\therefore \text{Mean life } \bar{T} = \frac{1}{\lambda} = \frac{1}{1.389 \times 10^{-11}} \text{ s} = 2282 \text{ years}$$

## 20.19 UNITS OF RADIOACTIVITY

Generally we express the activity of a radioactive substance in terms of *curie* and its submultiples *millicurie* ( $mCi$ ) and *micro curie* ( $\mu Ci$ ). The curie is defined as the quantity of a radioactive substance which gives  $3.70 \times 10^{10}$  disintegrations / second. This is the number of disintegrations per second per gram of radium. Another *unit rutherford* ( $Rd$ ) is also used. It is defined as the quantity of a radioactive substance which gives  $10^6$  disintegrations / second. It is to be noted that for a substance with very short half-life period, very little of the substance is required for 1 curie of activity. On the other hand, for the substance with very long half-life period, a very large quantity of the substance is required for 1 curie of activity.

**EXAMPLE.** Calculate the weight in kg of one curie of Ra B ( $Pb^{214}$ ) from the half-life of 26.8 minutes.

**SOL.** 1 curie =  $3.7 \times 10^{10}$  disintegrations / sec.

Here,  $-dN/dt = \lambda N = 3.7 \times 10^{10}$

$$\text{or } N = \frac{3.7 \times 10^{10}}{\lambda} = \frac{3.7 \times 10^{10} \times T_{1/2}}{0.693}$$

$$\therefore N = \frac{(3.7 \times 10^{10})(26.8 \times 60)}{0.693} = 8.585 \times 10^{13} \text{ atoms.}$$

From Avogadro's hypothesis,  $6.023 \times 10^{26}$  atoms are contained in 214 kg of RaB.

$$\begin{aligned}
 8.585 \times 10^{13} \text{ atoms are contained in } &\frac{214 \times 8.585 \times 10^{13}}{6.023 \times 10^{26}} \text{ kg.} \\
 &= 3.1 \times 10^{-11} \text{ kg.}
 \end{aligned}$$

**Activity.** The *activity* of a sample of any radioactive nuclide is the rate at which the nuclei of its constituent atoms decay. If  $N$  is the number of nuclei present in the sample at a certain time, its activity  $R$  is given by  $R = -dN/dt$ . The S.I. unit of activity is named after Henri Becquerel.

$$1 \text{ becquerel} = 1 \text{ Bq} = 1 \text{ event/s.}$$

$$1 \text{ MBq} = 10^6 \text{ Bq} \text{ and } 1 \text{ GBq} = 10^9 \text{ Bq.}$$

**EXAMPLE.** Find the activity of 1 mg ( $10^{-6} \text{ kg}$ ) of radon.

**SOL.** The decay constant of radon is

$$\lambda = \frac{0.693}{3.8d} = \frac{0.693}{3.8 \times 86400 \text{ s}} = 2.1 \times 10^{-6} \text{ s}^{-1}.$$

$$\text{The No. of atoms in } 10^{-6} \text{ kg of Rn}^{222} = N = \frac{(6.023 \times 10^{26}) \times 10^{-6}}{222} = 2.7 \times 10^{18}$$

$$\begin{aligned} \therefore R &= \lambda N = (2.1 \times 10^{-6}) \times (2.7 \times 10^{18}) \\ &= 5.7 \times 10^{12} \text{ events / s} \\ &= 5.7 \times 10^3 \text{ GBq} = 153 \text{ Ci} \end{aligned}$$

In this case, the “events” are alpha decays.

## 20.20 LAW OF SUCCESSIVE DISINTEGRATION

Chain disintegration means a process in which a radioactive substance disintegrates to form a new substance, which disintegrates to form another new substance and so on. Consider a substance  $A$  which decays to form a substance  $B$ . Let the substance  $B$  decay to form a substance  $C$  and so on.

$$A \rightarrow B \rightarrow C \rightarrow \dots X \text{ (stable).}$$

In a radioactive series, any two adjacent elements are considered as *parent* and *daughter*. Evidently in a series, the parent of the following element will be daughter of the preceding one.

At time  $t = 0$ , the number of initial atoms in  $A = N_0$

and the number of initial atoms in  $B = 0$

At time  $t$ , let the number of atoms in  $A = N_1$

and the number of atoms in  $B = N_2$ .

Let  $\lambda_1$  and  $\lambda_2$  be the decay constants of  $A$  and  $B$  respectively. Every time an atom of  $A$  disappears, an atom of  $B$  is produced.

$\therefore$  Rate of formation of daughter  $B = \lambda_1 N_1$ .

The rate at which  $B$  decays  $= \lambda_2 N_2$ .

$\therefore$  The net increase in the number of  $B$  atoms

$$\frac{dN_2}{dt} = \lambda_1 N_1 - \lambda_2 N_2$$

$$\text{Hence } \frac{dN_2}{dt} = \lambda_1 N_0 e^{-\lambda_1 t} - \lambda_2 N_2 \text{ (since } N_1 = N_0 e^{-\lambda_1 t}).$$

$$\therefore \frac{dN_2}{dt} + \lambda_2 N_2 = \lambda_1 N_0 e^{-\lambda_1 t}$$

Multiplying both sides by the integrating factor  $e^{\lambda_2 t}$ ,

$$\frac{dN_2}{dt} e^{\lambda_2 t} + \lambda_2 N_2 e^{\lambda_2 t} = \lambda_1 N_0 e^{(\lambda_2 - \lambda_1)t}$$

$$\text{or } \frac{d}{dt}[N_2 e^{\lambda_2 t}] = \lambda_1 N_0 e^{(\lambda_2 - \lambda_1)t}$$

$$\text{Integrating, } N_2 e^{\lambda_2 t} = \frac{\lambda_1}{\lambda_2 - \lambda_1} N_0 e^{(\lambda_2 - \lambda_1)t} + C \quad \dots(1)$$

The constant of integration  $C$  can be determined from the initial conditions.

$$\text{i.e., when } t = 0, N_2 = 0; \therefore C = \frac{-\lambda_1 N_0}{\lambda_2 - \lambda_1}$$

Substituting the value of  $C$  in (1),

$$N_2 e^{\lambda_2 t} = \frac{\lambda_1 N_0}{\lambda_2 - \lambda_1} [e^{(\lambda_2 - \lambda_1)t} - 1]$$

$$\text{or } N_2 = \frac{\lambda_1 N_0}{\lambda_2 - \lambda_1} [e^{-\lambda_1 t} - e^{-\lambda_2 t}] \quad \dots(2)$$

#### Radioactive equilibrium :

**Case (i). Secular or permanent equilibrium.** Suppose  $T_1 \gg T_2$  with  $T_1 \approx \infty$  and  $T_2 = 0$ . (i.e., the half-life of  $A$  is very much longer than that of  $B$ ).

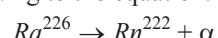
Then,  $\lambda_1 \ll \lambda_2$  and  $\lambda_1 \approx 0$ . In equation (2),  $e^{-\lambda_1 t} \rightarrow 1$

$$\therefore N_2 = \frac{N_0 \lambda_1}{\lambda_2} (1 - e^{-\lambda_2 t}) \quad \dots(3)$$

After an appreciably long time,  $e^{-\lambda_2 t}$  becomes negligible and

$$\begin{aligned} N_1 &\approx N_0 \\ \therefore N_1 \lambda_1 &= N_2 \lambda_2 \end{aligned} \quad \dots(4)$$

This relation shows that at equilibrium, the rate of decay of any radioactive product is just equal to its rate of production from the previous member of the chain. Now the daughter is said to be in *secular or permanent equilibrium* with the parent. An example of this is provided by the decay of radium into radon according to the equation.



The half-life period of  $Ra$  (1590 years) is very large compared with that of  $Rn$  (3.8 days). Fig. 20.29 shows the experimental decay and recovery curves. After a time, long enough compared with its mean life,  $Rn$  is in permanent equilibrium with  $Ra$ . Its amount becomes constant as shown by the sum of the ordinates of the two curves which becomes constant beyond a certain value of time indicated by the point of intersection.

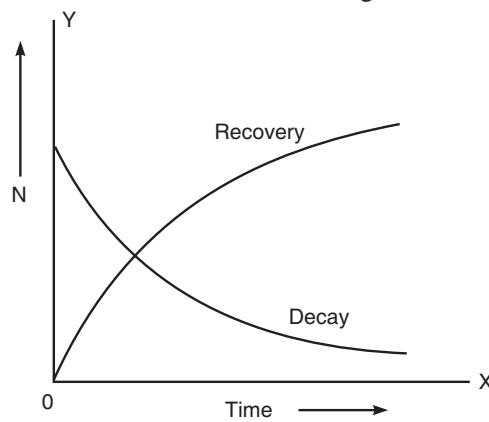


Fig. 20.29

**Case (ii). Transient equilibrium.** Suppose  $\lambda_1 \ll \lambda_2$  but  $\lambda_1 \neq 0$  i.e., the decay constant  $\lambda_1$  which is very small compared with  $\lambda_2$ , is not small enough to be ignored. In this case, after a sufficiently long time,  $e^{-\lambda_2 t}$  becomes negligible compared with  $e^{-\lambda_1 t}$ . Hence

$$N_2 = \frac{N_0 \lambda_1}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} = \frac{\lambda_1 N_1}{\lambda_2 - \lambda_1}$$

or

$$\frac{N_2}{N_1} = \frac{\lambda_1}{\lambda_2 - \lambda_1} \quad \dots(5)$$

In this case, both  $A$  and  $B$  decay while the ratio  $N_2 / N_1$  remains constant.

## 20.21 RADIOACTIVE DATING : THE AGE OF THE EARTH

The age of the earth is estimated from the relative abundance of the two isotopes of uranium,  $U^{238}$  and  $U^{235}$ . The half-periods of  $U^{238}$  and  $U^{235}$  are  $4.5 \times 10^9$  years and  $7 \times 10^8$  years respectively. Assume that at the beginning when the earth was formed the proportions of the two isotopes were equal. The present relative abundance of  $U^{238}$  to  $U^{235}$  in natural uranium is 99.3% to 0.7%.

$$\therefore \frac{N_1}{N_2} = \frac{99.3}{0.7} = \frac{N_0 e^{-\lambda_1 t}}{N_0 e^{-\lambda_2 t}} = e^{(\lambda_2 - \lambda_1)t}$$

where

$$\lambda_1 = \frac{0.6931}{4.5 \times 10^9} \text{ and } \lambda_2 = \frac{0.6931}{7 \times 10^8}$$

$$\therefore \log_e \left( \frac{99.3}{0.7} \right) = (\lambda_2 - \lambda_1)t$$

$$\begin{aligned} \therefore t &= \frac{1}{\lambda_2 - \lambda_1} \log_e \left[ \frac{99.3}{0.7} \right] \\ &= \frac{1}{\left( \frac{0.6931}{7 \times 10^8} \right) - \left( \frac{0.6931}{4.5 \times 10^9} \right)} \log_e \left( \frac{99.3}{0.7} \right) \\ &= 5.93 \times 10^9 \text{ years.} \end{aligned}$$

This value agrees nearly with that given by astronomical evidence for the age of the universe.

**Radio-carbon Dating.** One of the interesting properties of radionuclides is that they have constant half-life irrespective of the state of chemical combination of the atoms in which they reside. This property finds applications in dating or age estimation.  ${}^6C^{14}$ , a carbon isotope is used in archaeological dating. This isotope has a half-life of 5730 years. The interaction of cosmic rays with the atmospheric gas atoms produces the radioactive isotope  $C^{14}$ . This is assimilated by living organism along with the more abundant  $C^{12}$ . The production of  ${}^6C^{14}$  and its subsequent decay are in equilibrium in the earth's atmosphere. Hence  ${}^6C^{14}$  is present in equilibrium concentration in all living plants. However, after the death of the creature or plant, the intake of  ${}^6C^{14}$  stops. With the death of organism,  $C^{14}$  begins to decay. Hence the number of  $C^{14}$  atoms to the number of stable  $C^{12}$  atoms in the object decreases with time after death. Thus the time elapsed since death can then be estimated by measuring this ratio. Using this procedure, ages of wooden slabs from tombs, charcoal used for old paintings etc., have been studied.

- The same technique has been used in estimating the age of the earth from measurement on the relative amounts of  $U^{238}$  and  $Pb^{206}$  in geological specimens. In 1913 Joly and Rutherford suggested that if igneous rock, formed as a result of a prehistoric volcanic eruption, contained a small amount of uranium it would steadily decay, leaving less uranium and depositing more stable  $Pb-206$ . By measuring the ratio of uranium to lead in rock samples, a rather exact time can be determined for the origin of the geological deposits. Uranium dating measures times of the order of millions of years.

### EXERCISE

1. The maximum velocity of  $\alpha$ -particle is \_\_\_\_\_ m/sec.  
 (a)  $1.6 \times 10^5$       (b)  $1.7 \times 10^4$       (c)  $1.7 \times 10^6$       (d)  $1.7 \times 10^7$   
 (B.U. 2014)
2.  $\alpha$ -particle carries \_\_\_\_\_ of positive charge.  
 (a) one unit      (b) two units      (c) three units      (d) 8 units  
 (B.U. 2015)
3. Beta particles are nothing but  
 (a) Electrons      (b) Protons      (c) Neutrons      (d) Cosmic rays  
 (B.U. 2015)
4. Geiger Nuttal relation is  
 (a)  $\lambda = A + B \log R_\alpha$       (b)  $\log \lambda = A + B \log R_\alpha$   
 (c)  $\lambda = \log R_\alpha$       (d)  $\log \lambda = \log R_\alpha$   
 (B.U. 2013)
5. The particle emitted in  $\beta$  decay together with electron  
 (a) Photon      (b) Neutrino      (c) Antineutrino      (d) Meson  
 (B.U. 2006)
6. The Half-life period ( $T_{1/2}$ ) of a radioactive element is  
 (a)  $0.693 \lambda$       (b)  $0.693 \lambda^2$       (c)  $\frac{0.693}{\lambda}$       (d) None of these  
 (B.U. 2012)  
 [Ans.  $T_{1/2} = \frac{0.693}{\lambda} = \frac{0.693}{0.00231} = 300$  days ]
7. The disintegration constant of a radioactive element is 0.00231 per day. The half-life period is  
 (a) 300 days      (b) 400 days      (c) 500 days      (d) 3 days  
 (B.U. 2012)
8. The reciprocal of the radioactive disintegration constant is called  
 (a) half life period      (b) average life of atom      (c) mean life period      (d) binding energy  
 (B.U. 2011)  
 [Ans. 1. (d), 2 (b), 3. (a), 4. (b), 5. (b), 6. (c), 8 (c)]
9. List out the properties of  $\alpha$ -rays and  $\beta$  rays.      (B.U. 2013)
10. Describe the method that is used to determine the e/m of alpha particles.      (B.U. 2011)
11. What is meant by range of  $\alpha$ -particles? Explain how the range can be measured experimentally.  
 (P.U. 2007)
12. Explain about  $\alpha$  particle disintegration energy.      (B.U. M.Sc. 2015)
13. Polonium-212 emits  $\alpha$ -particles whose K.E. is 10.54 MeV. Determine the  $\alpha$  disintegration energy.  
 (B.U. 2006)

**Sol.** When polonium – 212 emits an  $\alpha$ -particle, the mass of the daughter nucleus is 208 units.

Here,  $M = 208$  units,  $m = 4$  units

$$\begin{aligned} \text{K.E. of the } \alpha\text{-particles} &= \frac{1}{2} mv^2 = 10.54 \text{ MeV. E = ?} \\ \therefore E &= \frac{1}{2} mv^2 \left(1 + \frac{m}{M}\right) = 10.54 \left(1 + \frac{4}{208}\right) = 10.74 \text{ MeV} \end{aligned}$$

14. Write a short note on Geiger-Nuttal law.      (Calicut U. 2002)
15. Outline Gamow's theory of alpha decay and explain how it leads to Geiger-Nuttal law      (B.U. 2006)
16. Describe an experiment for measuring the variation in the value of  $e/m$  with velocity in the case of  $\beta$ -particles.

17. Explain the existence of continuous spectrum of  $\beta$  particles. Describe neutrino theory of  $\beta$  decay.  
(B.U. 2006)
18. Explain the experimental verification of the existence of neutrino.  
(B.U. 2006)
19. Define Parity. Explain the violation of parity conservation during  $\beta$ -decay process. How has this been experimentally verified?  
(Meerut 1988)
20. Describe an experiment to determine the wavelength of  $\gamma$ -rays.  
(B.U. 2012)
21. Write short notes on (i) Nuclear isomerism (ii) Internal conversion.
22. What is Mossbauer effect? Describe experimental set up to study this effect.  
(Mumbai University, October 2010)
23. State and explain the laws of radioactive disintegration process.  
(B.U. 2014)
24. Write the fundamental laws of radioactivity.  
(B.U. 2011)
25. Explain "half-period" of a radioactive element and derive an expression for it.
26. Explain the terms, decay constant, half-life and average life as applied to a radioactive substance. Find the relation between them.
27. Give the relevant theory of successive disintegration of radioactive elements. What do you mean by radioactive equilibrium? Obtain the conditions for secular and transient equilibrium.  
(Osmania Univ. 91, 90)
28. A carbon specimen found in a cave contained  $1/8$  as much  $C^{14}$  as an equal amount of carbon in living matter. Calculate the approximate age of the specimen. Half-life period of  $C^{14}$  is 5568 years.  
(Ans. 16710 years)

# ARTIFICIAL TRANSMUTATION OF ELEMENTS

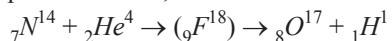
**After reading this chapter, you should be able to**

- ◆ Explain the terms: Q-value of a nuclear reaction, endoergic reactions and exoergic reactions
- ◆ Explain why the centre of mass frame is preferred in describing the kinematics of a nuclear reaction
- ◆ Explain stripping and pick-up reactions
- ◆ Explain the concept of radiotherapy
- ◆ Discuss production, properties and detection of neutrons and their interaction with matter in bulk
- ◆ Distinguish between the slow and fast neutrons.

## 21.1 THE DISCOVERY OF ARTIFICIAL TRANSMUTATION

**Rutherford's experiment.** Artificial transmutation is the conversion of one element into another provoked by artificial means. The apparatus with which Rutherford demonstrated the first artificial transmutation is shown in Fig. 21.1. It consists of a chamber  $C$  one side of which has an opening covered by a thin sheet of silver foil  $F$ . A zinc sulphide screen  $S$  is placed close to  $F$  and the scintillations on it are observed by a microscope  $M$ . The chamber can be filled with various gases through the side tubes  $TT$ . The source of  $\alpha$ -particles ( $RaC'$ ) was placed at  $R$  on a small disc. The distance of  $R$  from  $F$  can be varied by means of a long rod attached to the disc at one end and manipulated from outside the chamber at the other. The range of the  $\alpha$ -particle used was known to be about 7 cm in air, corresponding to an energy of about 8 MeV. The effects produced by these  $\alpha$ -particles on the various gases can be studied by observing the scintillations on  $S$ .

When the chamber was filled with oxygen or  $CO_2$  no scintillations were observed on the screen when the distance between  $R$  and  $F$  was greater than the range of the  $\alpha$ -particles (7 cm). But when the chamber was filled with nitrogen, scintillations were observed on the screen even when the distance between  $R$  and  $F$  was as high as 40 cm. The  $\alpha$ -particles themselves could not penetrate such a distance. Thus, obviously, scintillations were not caused by the  $\alpha$ -particles but by particles emitted by the nitrogen atoms. Magnetic deflection experiments indicated that these particles were hydrogen nuclei or protons. On the basis of the Bohr's theory of the compound nucleus, the reaction can be written as



The reaction may be pictorially represented as in Fig. 21.2. The new element formed was the rare isotope of oxygen of mass number 17. Thus, the first transmutation of nitrogen into oxygen was achieved and established.

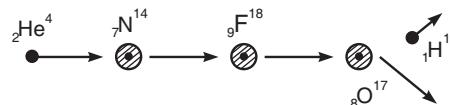


Fig. 21.2

### Bohr's Theory of Nuclear Disintegration

Bohr assumed that the nucleus is somewhat like a liquid drop. When a projectile strikes the nucleus, the projectile is captured by the nucleus. The nucleus formed after the projectile is captured, is called a *compound nucleus*. The energy of the projectile is quickly distributed among the nucleons inside the compound nucleus. Therefore, the compound nucleus resembles a heated-up liquid drop and may be considered to exist in a variety of quasi-stationary excitation states. The compound nucleus persists in its excited state until a particular nucleon momentarily happens to have a large fraction of the excitation energy to escape from the compound nucleus, thus, leading to a disintegration. This process is similar to the slow evaporation of particles from the surface of a liquid drop.

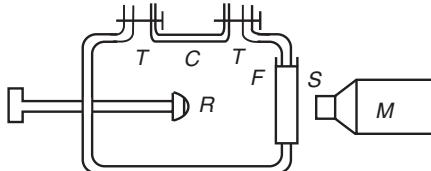


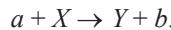
Fig. 21.1



Ernest Rutherford (1871–1937)

In 1919 at Manchester, he discovered that the nuclei of certain light elements, such as nitrogen, could be “disintegrated” by the impact of energetic alpha particles coming from some radioactive source, and that during this process fast protons were emitted.

A nuclear reaction may be represented by



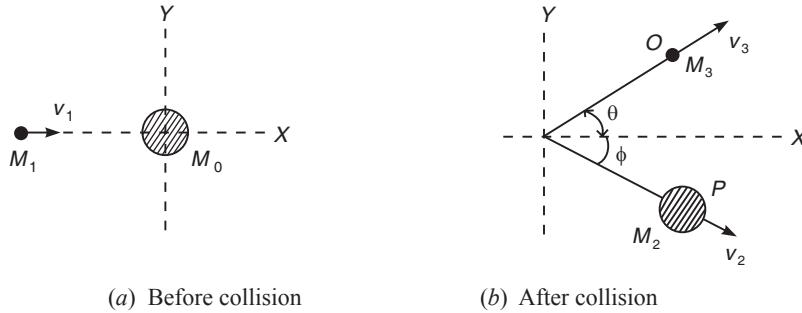
This formula signifies that the particle  $a$  interacts with the nucleus  $X$  to yield the nucleus  $Y$  and particle  $b$ . The above equation is usually abbreviated and is written as

$$X(a, b) Y.$$

For example,  ${}_6^C{}^{12}(d, n) {}_7^N{}^{13}$  stands for a reaction between an incident deuteron ( $d = {}_1^H{}^2$ ) and a  ${}_6^C{}^{12}$  nucleus to produce a  ${}_7^N{}^{13}$  nucleus with the emission of a neutron.

## 21.2 THE Q-VALUE EQUATION FOR A NUCLEAR REACTION (NUCLEAR REACTION KINEMATICS)

Let a particle of mass  $M_1$  moving with velocity  $v_1$  collide with a target nucleus of mass  $M_0$  at rest (Fig. 21.3). After collision, the particle  $O$  of mass  $M_3$  is emitted with velocity  $v_3$  at an angle  $\theta$  and the recoiling nucleus  $P$  of mass  $M_2$  is emitted with a velocity  $v_2$  at an angle  $\phi$ . The conservation of linear momentum in the plane of paper yields the equations



**Fig. 21.3**

$$M_1 v_1 = M_3 v_3 \cos \theta + M_2 v_2 \cos \phi \quad \dots(1)$$

$$0 = M_3 v_3 \sin \theta - M_2 v_2 \sin \phi \quad \dots(2)$$

$$\therefore M_2 v_2 \cos \phi = M_1 v_1 - M_3 v_3 \cos \theta \quad \dots(3)$$

$$M_2 v_2 \sin \phi = M_3 v_3 \sin \theta \quad \dots(4)$$

Squaring and adding, we get

$$M_2^2 v_2^2 = M_1^2 v_1^2 + M_3^2 v_3^2 - 2 M_1 M_3 v_1 v_3 \cos \theta \quad \dots(5)$$

Using the kinetic energy relations,

$$E_{k_1} = \frac{1}{2} M_1 v_1^2, E_{k_2} = \frac{1}{2} M_2 v_2^2 \text{ and } E_{k_3} = \frac{1}{2} M_3 v_3^2$$

Substituting these values in Eq. (5), we have

$$2 M_2 E_{k_2} = 2 M_1 E_{k_1} + 2 M_3 E_{k_3} - 2 (M_1 M_3 E_{k_1} E_{k_3})^{1/2} 2 \cos \theta$$

$$\text{or } E_{k_2} = \frac{M_1 E_{k_1}}{M_2} + \frac{M_3 E_{k_3}}{M_2} - \frac{2}{M_2} (M_1 M_3 E_{k_1} E_{k_3})^{1/2} \cos \theta \quad \dots(6)$$

Let

$$Q = (E_{k_2} + E_{k_3}) - E_{k_1} \quad \dots(7)$$

Substituting the value of  $E_{k_2}$  from Eq. (6) in Eq. (7), we have

$$Q = \frac{M_1 E_{k_1}}{M_2} + \frac{M_3}{M_2} E_{k_3} - \frac{2}{M_2} (M_1 M_3 E_{k_1} E_{k_3})^{1/2} \cos \theta + E_{k_3} - E_{k_1}$$

or

$$Q = E_{k_3} \left( 1 + \frac{M_3}{M_2} \right) - E_{k_1} \left( 1 - \frac{M_1}{M_2} \right) - \frac{2}{M_2} (M_1 M_3 E_{k_1} E_{k_3})^{1/2} \cos \theta \quad \dots(8)$$

If

$\theta = 90^\circ$ ,  $\cos 90^\circ = 0$ . Hence, Eq. (8) reduces to

$$Q = E_{k_3} \left( 1 + \frac{M_3}{M_2} \right) - E_{k_1} \left( 1 - \frac{M_1}{M_2} \right) \quad \dots(9)$$

**Threshold energy.** The threshold energy ( $E_{th}$ ) is defined as the minimum kinetic energy of the incident particle which will initiate an endoergic reaction.

We can calculate the threshold energy by using the centre of mass coordinate system. For any collision or reaction in centre of mass coordinate system, the linear momentum is always zero before and after the reaction. Let  $E'_{k_1}$  be the kinetic energy of the incident particle in centre of mass coordinate system. An endoergic reaction is possible if

$$E'_{k_1} \geq |Q| \quad \dots(10)$$

Let  $M_{red}$  be the reduced mass of incident particle and the target nucleus. Then,

$$\begin{aligned} M_{red} &= \frac{M_1 M_0}{(M_0 + M_1)} \\ \therefore E'_{k_1} &= \frac{1}{2} M_{red} v_1^2 = \frac{1}{2} \frac{M_1 M_0}{(M_0 + M_1)} v_1^2 \end{aligned} \quad \dots(11)$$

From Eqs. (10 and 11),

$$\begin{aligned} \frac{1}{2} \frac{M_1 M_0}{(M_0 + M_1)} v_1^2 &\geq |Q| \text{ or } \frac{1}{2} M_1 v_1^2 \geq \left( \frac{M_0 + M_1}{M_0} \right) |Q| \\ \therefore \text{Threshold energy } E_{th} &= \left( 1 + \frac{M_1}{M_0} \right) |Q| \end{aligned} \quad \dots(12)$$

### 21.2.1. Energy Balance in Nuclear Reactions and the Q-value

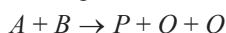
In all nuclear reactions, the total sum of mass and energy is conserved. Thus taking the equation  $A + B \rightarrow P + O$ , the target is supposed to be at rest and let its mass be  $m_1$ . The projectile has a mass  $m_2$  and K.E. =  $E_2$ .

The product nucleus has a mass  $m_3$  and K.E.  $E_3$ . The outgoing particle has a mass  $m_4$  and K.E.  $E_4$ . The equation representing the conservation of total energy is written as

$$m_1 c^2 + m_2 c^2 + E_2 = m_3 c^2 + E_3 + m_4 c^2 + E_4$$

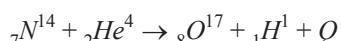
Let  $Q = E_3 + E_4 - E_2 = (m_1 + m_2 - m_3 - m_4)c^2$ .

The quantity  $Q$  is called the energy balance or the  $Q$ -value of the reaction. If  $Q$  is positive, the reaction is *exothermic* (or *exoergic*). Now the K.E. of products of transmutation is greater than the K.E. of the reactants and energy is released in the process. If  $Q$  is negative, the reaction is *endothemic* (or *endoergic*). The equation for a nuclear reaction is written as



where  $Q$  may be positive or negative.

**EXAMPLE 1.** Compute the  $Q$  value for the reaction



where  $Q$  denotes the energy absorbed or evolved during the nuclear reaction.

**SOL.** The atomic masses of the particles are :  $N^{14} = 14.003074$ ,  $He^4 = 4.002604$ ,  $O^{17} = 16.99913$  and  $H^1 = 1.007825u$ .

Substituting these values in the equation,

$$\begin{aligned} 14.003074 + 4.002604 &= 16.99913 + 1.007825 + Q \\ \therefore Q &= 18.005678 - 18.006955 = -0.001277u \end{aligned}$$

According to Einstein's mass-energy relation,  $1u = 931.3$  MeV

$\therefore$  The calculated value of  $Q = -0.001277 \times 931.3 = -1.189$  MeV.

This is in close agreement with the experimental value (*i.e.*,  $-1.26$  MeV). Incidentally, Einstein's mass-energy relation gets verified.

**EXAMPLE 2.** Determine the product nuclei and  $Q$  values in the following reactions :  $Al^{27}$  ( $d, \alpha$ ) and  $Mg^{25}$  ( $\alpha, d$ ). Masses of  $Al^{27}$ ,  $Mg^{25}$ ,  $\alpha$  and  $d$  are  $26.9901$ ,  $24.9936$ ,  $4.0039$  and  $2.0147$  amu respectively. Comment on your results.

**SOL.** (i) The nuclear reaction is

$$\begin{aligned} {}_{13}Al^{27} + {}_1H^2 &\rightarrow {}_{12}Mg^{25} + {}_2He^4 + Q \\ \therefore Q &= 26.9901 + 2.0147 - 24.9936 - 4.0039 = 0.0073 \text{ amu.} \end{aligned}$$

*i.e.*,  $Q = 0.0073 \times 931.3$  MeV =  $6.797$  MeV.

$$\begin{aligned} (ii) \quad {}_{12}Mg^{25} + {}_2He^4 &\rightarrow {}_{13}Al^{27} + {}_1H^2 + Q \\ \therefore Q &= 24.9936 + 4.0039 - 26.9901 - 2.0147 = -0.0073 \text{ amu.} \\ \text{i.e., } Q &= -0.0073 \times 931.3 \text{ MeV} = -6.797 \text{ MeV.} \end{aligned}$$

### 21.2.2. Threshold Energy of an Endoergic Reaction

In an endoergic reaction  $-Q$  is the energy which is needed to excite the reaction. This energy is supplied by K.E. of the incoming particle. But all the K.E. of the projectile particle is not available for the nuclear reaction because part of the energy is used to give K.E. to the compound nucleus. Hence, for a reaction in which  $-Q$  energy is to be absorbed, the incident particle should supply some energy in addition to  $-Q$ . Thus the minimum K.E. which the projectile should possess so that the nuclear reactions may take place is called the *threshold energy*.

Let  $m_i$  and  $v_i$  be the mass and initial velocity of the projectile. Let  $m_c$  and  $v_c$  be the mass and velocity of the compound nucleus. Applying conservation of momentum principle,

$$m_i v_i = m_c v_c \text{ or } v_c = m_i v_i / m_c.$$

$$\text{K.E. of compound nucleus} = \frac{1}{2} m_c v_c^2 = \frac{1}{2} m_c \left( \frac{m_i}{m_c} \right)^2 v_i^2$$

$$\text{Energy available for the reaction} = \frac{1}{2} m_i v_i^2 - \frac{1}{2} m_c v_c^2 = \frac{1}{2} m_i v_i^2 \left( 1 - \frac{m_i}{m_c} \right)$$

But  $m_c = m_i + m_t$  where  $m_t$  = mass of target nucleus.

$$\therefore \frac{1}{2} m_i v_i^2 - \frac{1}{2} m_c v_c^2 = \frac{1}{2} m_i v_i^2 \left( \frac{m_t}{m_i + m_t} \right)$$

$$\text{or } -Q = \frac{1}{2} m_i v_i^2 \left( \frac{m_t}{m_i + m_t} \right)$$

$$\therefore \text{The threshold energy} = E_{th} = \frac{1}{2} m_i v_i^2 = -Q \left( \frac{m_i + m_t}{m_t} \right)$$

**EXAMPLE.** The  $Q$  value of the  $\text{Na}^{23}(n, \alpha)\text{F}^{20}$  reaction is  $-5.4 \text{ MeV}$ . Determine the threshold energy of the neutrons for this reaction.

**SOL.** Here,

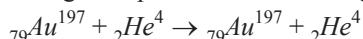
$$\begin{aligned} Q &= -5.4 \text{ MeV}; m_i = 1.008665u, m_t = 22.9898 u. \\ E_{th} &= -Q \left( \frac{m_i + m_t}{m_t} \right) = -(-5.4) \left( \frac{1.008665 + 22.9898}{22.9898} \right) \\ &= 5.635 \text{ MeV}. \end{aligned}$$

### 21.2.3. Types of Nuclear Reactions

Following are some of the main types of nuclear reactions.

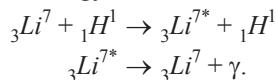
(i) **Elastic scattering.** In this case, the incident particle strikes the target nucleus and leaves without loss of energy, but its direction may change.

**EXAMPLE :** Scattering of  $\alpha$ -particles from a thin gold foil.

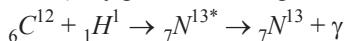


The target nucleus remains unaffected.

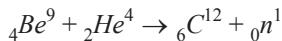
(ii) **Inelastic scattering.** In this case, the incident particle loses a part of its energy in exciting the target nucleus to a higher allowed energy level. The excited nucleus later decays to the ground state, radiating the excess energy in the form of a  $\gamma$ -ray photon. *Example :*



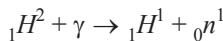
(iii) **Radiative Capture.** Here the incident particle is captured by the target nucleus and a new nucleus is formed. The new nucleus, in general, has a considerable excess of energy and decays with the emission of one or more  $\gamma$ -ray photons. *Example :*



(iv) **Disintegration.** Here the incident particle is absorbed by the target nucleus and the ejected particle is a different one. The composition of the resultant nucleus is also different from the parent nucleus. An example is the disintegration of beryllium by  $\alpha$ -particle producing neutrons.



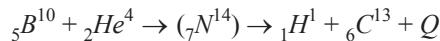
(v) **Photodisintegration.** When target materials are bombarded with radiations, the resulting compound nuclei are usually formed in excited states. These nuclei generally get rid of the excess excitation energy through neutron emission. For example,



This requires a photon of energy  $2.225 \text{ MeV}$ .

### 21.2.4. Conservation Laws

The conservation laws are stated and illustrated by reference to some specific nuclear reaction, say,



(i) **Conservation of charge :** Total charge is conserved in every type of *nuclear reaction*. In  $\text{B}^{10}(\alpha, p)\text{C}^{13}$  there are seven protons initially, also seven in the compound nucleus and in the products of the reaction.

(ii) **Conservation of nucleons :** The total number of nucleons entering and leaving the reaction is constant. In  $\text{B}^{10}(\alpha, p)\text{C}^{13}$ , we find 14 nucleons at each stage of the reaction.

(iii) **Conservation of mass-energy :** In nuclear reactions neither kinetic energy nor rest mass is conserved by itself. But their total is always conserved.

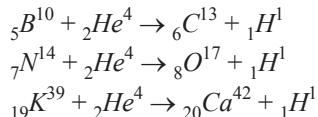
**(iv) Conservation of parity :** The net parity before the reaction must equal the net parity after the reaction.

**(v) Linear momentum, angular momentum, spin and isotopic spin** are the other physical quantities which are also conserved in a nuclear reaction.

**(vi) Quantities not conserved :** The most prominent physical characteristics which are not conserved in nuclear reactions are the magnetic dipole moments and the electric quadrupole moments of the reacting nuclei. These moments depend upon the internal distribution of mass, charge and current within the nuclei involved and are not subject to conservation laws.

### 21.3 NUCLEAR TRANSMUTATIONS

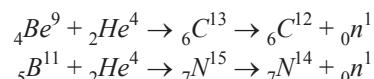
**Transmutations by  $\alpha$ -particles.** (i) ( $\alpha$ , p) **reaction.** The elements are bombarded by  $\alpha$ -particles and protons are ejected. Rutherford and Chadwick succeeded in disintegrating a number of elements from boron to potassium by  $\alpha$ -particle bombardment. The equations are :



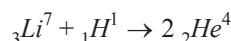
In these reactions, the atomic number of the product nucleus increases by 1 and mass number by 3.

(ii) ( $\alpha$ , n) **reaction.** When certain nuclei are bombarded by  $\alpha$ -particles, neutrons are ejected. In this type of reaction, the atomic number increases by 2 and mass number by 3. It was this type of reaction that led to the discovery of neutrons.

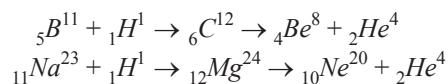
**EXAMPLE.**



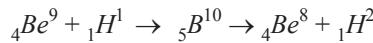
**Transmutation by protons.** (i) (p,  $\alpha$ ) **reactions.** Cockcroft and Walton artificially accelerated the protons with their tension multiplier and used them as projectiles in nuclear reactions. When lithium was bombarded by protons having energies from 100 to 700 keV, lithium broke up into two  $\alpha$ -particles after proton capture.



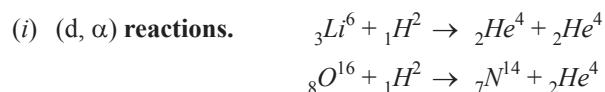
There are many other disintegrations produced by the bombardment with protons and they are given below :

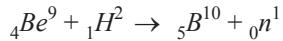
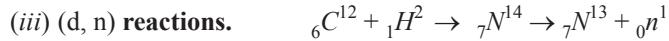


(ii) (p, d) **reaction.** In some cases instead of  $\alpha$ -particles, deuterons are also emitted.

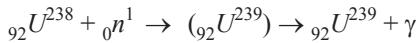
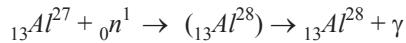
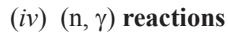
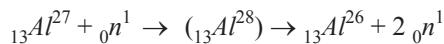
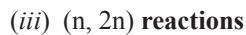
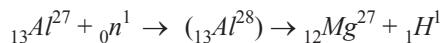
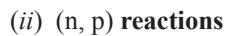
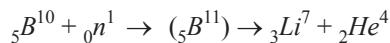
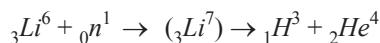
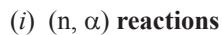


**Transmutation by deuterons.** In many nuclear reactions, high energy deuterons are used as bombarding particles. The deuterons are accelerated in a cyclotron and they possess very high energy. Some of the nuclear reactions are :





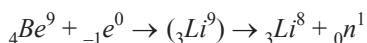
**Transmutation by neutrons.** Neutrons are very effective in producing transmutations. They are not affected by electrostatic fields because they do not possess any charge and are not repelled by the nuclei. Due to these reasons, neutrons have been found to be more penetrating into the nuclei as compared to  $\alpha$ -particles, protons and deuterons. The disintegrations produced by neutrons leads to the emission of various particles like  $\alpha$ -particles, protons, neutrons, photons and deuterons.



These reactions are called radiative captures. Such reactions usually produce artificially radioactive isotopes.

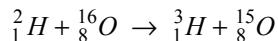
**Disintegration by electrons.** Electrons of very high energy obtained from betatrons cause the disintegration of a number of elements.

**EXAMPLE.** When beryllium target is bombarded with high energy electrons, neutron is ejected.



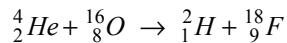
**Pick up reactions.** In the pick up reactions, the target loses nucleons to the projectile.

**Example.** In the reaction



${}_8^{16}O$  has lost a neutron to  ${}_1^2H$ .

**Stripping reactions.** The stripping reactions are inverse of the pick up reactions. In these reactions, the target gains nucleons from the projectile. For example, in the reaction

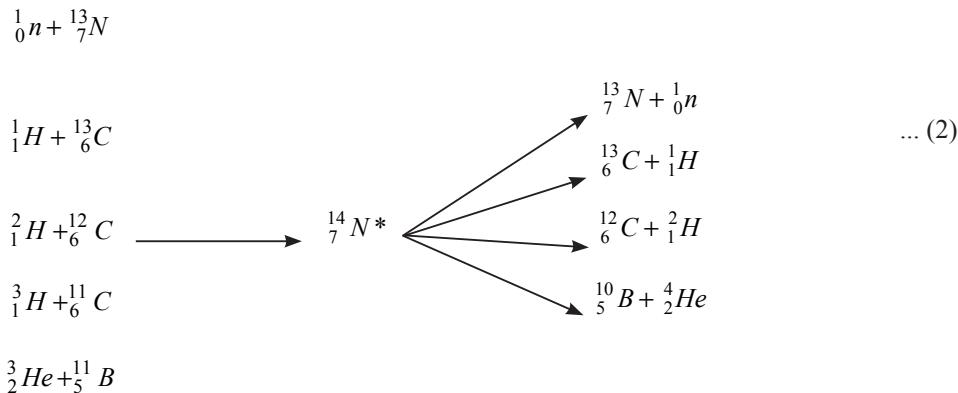


the target has gained one proton and one neutron from the  $\alpha$ -particle.

**Compound nuclear reactions.** In the compound nuclear reactions, all the nucleons take part. In the first stage, projectile and target combine to form a compound nucleus.



In the second stage, the formed compound nucleus, which is unstable, splits into components  $W$  and  $V$  after a short time. A typical life time of the compound nucleus is of the order of  $10^{-16}$  sec., which is about a million times greater than the time taken by the projectile to cross the target. Hence, the formation of a compound nucleus and its decay are two independent processes. Because of it, a particular compound nucleus can be formed in a number of ways and it can also decay in more than one way. For example,  ${}_{\frac{1}{7}}^{14} N^*$  can be formed in five different ways. It can decay in four different ways, as shown below:



- In the *initial channel*, we have  $X$  and  $Y$  nuclei.  
The *final channel* has  $W$  and  $V$  nuclei.

## 21.4 THE SCATTERING CROSS-SECTION

The properties of scattering interactions are usually expressed in terms of the *Scattering cross-section*. To define this term, we consider the following scattering experiment : A beam of particles (called *beam particles*) is directed on a scatterer consisting of *target particles*. As a result of *collisions* between beam and target particles, there are particles which emerge from the reactions (called *reaction products*) and these are detected in *particle detectors*. To describe this quantitatively, we suppose that the scatterer contains  $N_t$  target particles. The target is uniformly illuminated by a flux  $F_s$  (expressed as the number of beam particles per unit area per unit time arriving at the target) of beam particles. We suppose also that the scatterer is sufficiently small that the beam is negligibly attenuated in passing through it. Then, if there are  $\delta N_s$  scattering interactions per unit time which lead to detected particles, we define the scattering cross section  $\delta\sigma$  as  $\delta\sigma = \delta N_s / N_t F_s$ . In the limit that the detectors subtend very small solid angles, as seen from the target, we define the *differential scattering cross section*  $d\sigma$ . When a single detector subtending a solid angle  $\delta\Omega$  is used to define  $\delta\sigma$ , we may define the cross section per unit solid angle as

$$\frac{d\sigma}{d\Omega} = \lim_{\delta\Omega \rightarrow 0} \frac{\delta\sigma}{\delta\Omega}$$

The total scattering cross section  $\sigma$  is obtained by summing  $\delta\sigma$  over all scattering events:  $\sigma = \Sigma \delta\sigma$ .

It is customary practice to associate a cross-section with each particular type of nuclear interaction. A scattering cross-section is used when dealing with nuclear scattering processes;

an absorption cross-section for nuclear absorption processes; and a fission cross-section when studying nuclear collisions, leading to the fission of the target nucleus.

#### Determination of cross-section.

Consider a slab of some material whose area is  $A$  and thickness is  $dx$  (Fig. 21.4). Its volume is  $A dx$ . If the target material contains  $n$  nuclei per unit volume, the total No. of nuclei in the slab =  $n A dx$ . Each nucleus has a cross-section of  $\sigma$  for some particular interaction, so that the aggregate cross-section of all the nuclei in the slab is  $n A \sigma dx$ . Let  $N$  be the No. of incident particles in a bombarding beam. Let  $dN$  be the No. of particles that interact with nuclei in the slab. Then,

$$\frac{dN}{N} = \frac{\text{Aggregate cross section}}{\text{target area}} = \frac{n A \sigma dx}{A} = n \sigma dx.$$

Now let us consider the same beam of particles incident on a slab of thickness  $x$ . Let  $N_0$  be the number of incident particles. If each particle can interact only once,  $dN$  particles may be thought of as being removed from the beam in passing through the first  $dx$  of the slab. Hence

$$\frac{-dN}{N} = n \sigma dx \quad \text{or} \quad \int_{N_0}^N \frac{dN}{N} = -n \sigma \int_0^x dx$$

$$\begin{aligned} \text{or} \quad & \log_e N - \log_e N_0 = -n \sigma x \\ \text{or} \quad & N = N_0 e^{-n \sigma x} \end{aligned}$$

The number of surviving particles  $N$  decreases exponentially with increasing slab thickness  $x$ . Here,  $\sigma$  is the cross-section per nucleus or the *microscopic cross-section*.  $n$  is the number of nuclei per unit volume. The product  $n\sigma$  is called the *macroscopic cross-section and is denoted by  $\Sigma$* .  $\Sigma$  has the dimensions of  $L^{-1}$ .

Thus

$$N = N_0 e^{-\Sigma x}$$

The unit used for  $\sigma$  is an area of  $10^{-28} \text{ m}^2$ . This unit is called the *barn*.

If the incident flux (beam intensity) is  $I_0$  and the flux after penetrating a distance  $x$  is  $I$ , then  $I = I_0 e^{-\Sigma x}$  or

$$\log_e I = \log_e I_0 - \Sigma x$$

$\Sigma$  can be found by measuring the beam intensity with and without targets. A plot of  $\log_e I$  against  $x$  for several foil thicknesses gives  $\Sigma$  graphically and hence  $\sigma$ .

The cross-sections for most nuclear reactions depend upon the energy of the incident particle.

- Fig. 21.5 shows how the neutron capture cross-section of  $^{48}\text{Cd}^{113}$  varies with neutron energy.
- In this reaction, the absorption of a neutron is followed by the emission of a gamma ray :  $\text{Cd}^{113}(n, \gamma)\text{Cd}^{114}$ .
- The narrow peak at 0.176 eV is a resonance effect

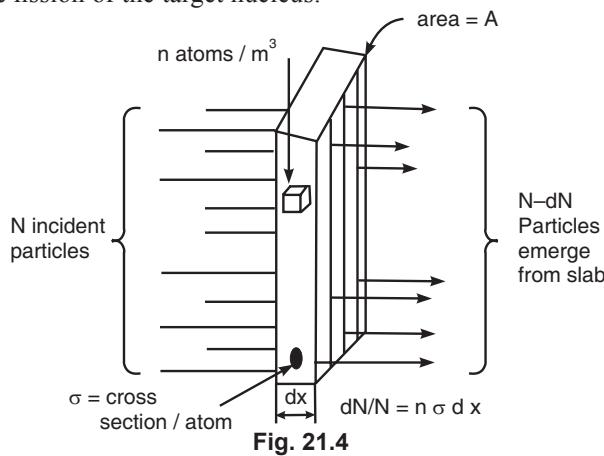


Fig. 21.4

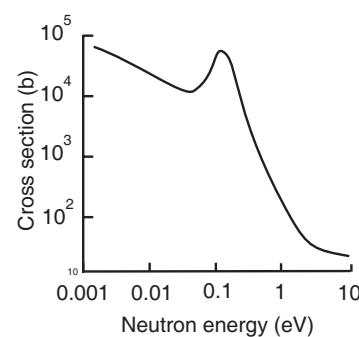


Fig. 21.5

associated with an excited state in the  $Cd^{114}$  nucleus. The capture cross-sections of  $Cd^{113}$  for slow neutrons are so great that cadmium is widely used in control rods for nuclear reactors.

**Macroscopic cross-section and mean free path.** The mean free path ( $\lambda$ ) of a particle in a material is the average distance it can travel in the material before interacting there.

$$\lambda = \frac{\int_0^\infty xe^{-\Sigma x} dx}{\int_0^\infty e^{-\Sigma x} dx} = \frac{1}{\Sigma}$$

#### 21.4.1. Neutron Thermalization

Consider a particle of mass  $m_1$  and initial velocity  $v_1$  that collides head-on with another particle of mass  $m_2$  at rest (Fig. 21.6). After collision they move with velocities  $v_1'$  and  $v_2'$ . The collision is elastic.

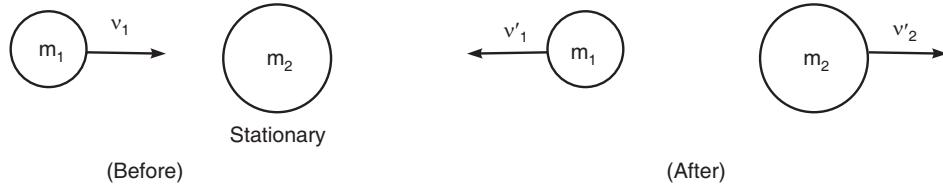


Fig. 21.6

By the law of conservation of linear momentum,

$$m_1 v_1 = m_2 v_2' - m_1 v_1'$$

$$\text{or } m_1 (v_1 + v_1') = m_2 v_2' \quad \dots(1)$$

Since the collision is elastic, kinetic energy is also conserved.

$$\frac{1}{2} m_1 v_1^2 = \frac{1}{2} m_1 v_1'^2 + \frac{1}{2} m_2 v_2'^2$$

$$\text{or } m_1 (v_1^2 - v_1'^2) = m_2 v_2'^2 \quad \dots(2)$$

Dividing Eq. (2) by Eq. (1), we obtain

$$v_1 - v_1' = v_2'$$

$$\text{or } v_1' = v_1 - v_2' \quad \dots(3)$$

Substituting this value of  $v_1'$  in Eq. (1), we get

$$v_2' = \frac{2m_1}{(m_1 + m_2)} v_1 \quad \dots(4)$$

Let  $K_1$  be the initial kinetic energy of the particle of mass  $m_1$  and  $K_2'$  the kinetic energy lost by it in the collision. Then,

$$\frac{K_2'}{K_1} = \frac{\frac{1}{2} m_2 v_2'^2}{\frac{1}{2} m_1 v_1^2} = \frac{4m_1 m_2}{(m_1 + m_2)^2} = \frac{4(m_2 / m_1)}{[1 + (m_2 / m_1)]^2}$$

If  $m_1 = m_2$ , then  $K'_2 = K_1$  and the energy transfer is a maximum. It means that *all* the energy of  $m_1$  is given to  $m_2$ . The neutron may transfer its entire kinetic energy to a proton on collision with it, since the mass  $m_p \approx m_n$ .

The above considerations show that in order to “slow down” the fast neutrons in a nuclear reactor, they must be made to collide with the stationary targets (nuclei) of nearly the same mass as of the neutrons themselves ( $m_1 = m_2$ ). That is why paraffin (which is rich in hydrogen whose nucleus ‘proton’ has nearly the same mass as a neutron) is a very good moderator.

## 21.5 CENTRE-OF-MASS COORDINATE SYSTEM

When the law of conservation of energy and the law of conservation of momentum were written to solve the nuclear collision discussed in Section 21.4.1, it was implicit that the velocities in the various expressions were the values in the *laboratory coordinate system*. In this *L-system*, it was assumed that the target nucleus was at rest before the collision and that only the bombarding particle was in motion. The treatment of nuclear collisions is easier if the centre of mass of the particles is taken as the reference system. In the *centre-of-mass* or *C-system*, the centre of mass of the incident particle and the target nucleus is considered to be at rest and both particles are approaching it.

The relationship between the *L* and *C*-systems is shown in Fig. 21.7. In the *L-system*, the observer views the collision occurring as shown in Fig. 21.7 (a). A particle of mass  $m_1$  and velocity  $v$  is incident upon a stationary particle of mass  $m_2$ . In the *C-system*, the observer views the collision occurring as shown in Fig. 21.7 (b). To an observer located at the centre of mass, the particles have equal and opposite momenta. The velocity  $V$  of the centre of mass is defined by the condition

$$m_1(v - V) = m_2V$$

or

$$V = \left( \frac{m_1}{m_1 + m_2} \right) v \quad \dots(1)$$

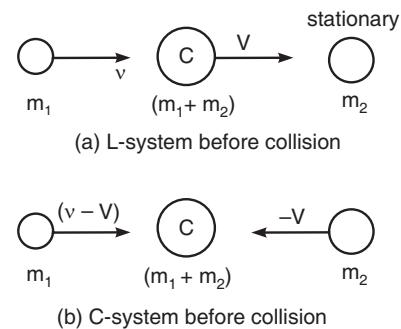
In the *L*-system, the total kinetic energy is that of the incident particle only.

$$K_{lab} = \frac{1}{2}m_1v^2 \quad \dots(2)$$

In the *C*-system,  $m_1$  has a velocity of  $(v - V)$  towards the right and  $m_2$  has a velocity of  $V$  towards the left. Therefore, the total kinetic energy in this reference frame is

$$\begin{aligned} K_{cm} &= \frac{1}{2}m_1(v - V)^2 + \frac{1}{2}m_2V^2 \\ &= \frac{1}{2}m_1v^2 - \frac{1}{2}(m_1 + m_2)V^2 \\ &= K_{lab} - \frac{1}{2}(m_1 + m_2)V^2 \\ \therefore K_{lab} &= K_{cm} + \frac{1}{2}(m_1 + m_2)V^2 \end{aligned} \quad \dots(3)$$

$$\text{Total K.E. in lab system} = \text{Total K.E. in cm system} + \text{K.E. associated with cm}$$



**Fig. 21.7**

When the particles collide, the maximum amount of kinetic energy that can be converted to excitation energy of the resulting compound nucleus while still conserving momentum is  $K_{cm}$ .

**EXAMPLE 1.** Find the excitation energy of  $^{92}U^{236}$  nucleus when  $U^{235}$  absorbs a thermal neutron of energy  $E = 0.025 \text{ MeV}$ . Mass of  $U^{235} = 235.0439 \text{ u}$  and those of  $U^{236}$  and Neutron are  $236.0457 \text{ u}$  and  $1.0087 \text{ u}$  respectively.  
(Osmania University, Oct. 93)

**SOL.** Mass of  $U^{235} = 235.0439 \text{ u}$

Mass of neutron =  $1.0087 \text{ u}$

Total mass of target and projectile =  $235.0439 + 1.0087 = 236.0526 \text{ u}$

Mass of  $U^{236} = 236.0457 \text{ u}$

Mass defect  $\Delta m = (236.0526 - 236.0457) = 0.0069 \text{ u}$

The equivalent energy =  $0.0069 \times 931 = 6.424 \text{ MeV}$ .

The excitation energy = Average binding energy per nucleon + KE of captured neutron

$$\left. \begin{array}{l} \text{Average binding energy} \\ \text{per nucleon of } U^{236} \end{array} \right\} = \frac{6.424}{236} = 0.02723 \text{ MeV}$$

KE of projectile neutron =  $0.025 \text{ MeV}$

Excitation energy =  $(0.02723 + 0.025) = 0.05223 \text{ MeV}$ .

**EXAMPLE 2.** The precise masses in the reaction



have been determined by mass spectrometer and are  $m(H) = 1.007825 \text{ u}$ ;  $m(He) = 4.002603 \text{ u}$ ;  $m(F^{19}) = 18.998405 \text{ u}$ ;  $m(O^{16}) = 15.994915 \text{ u}$ . Determine the  $Q$  and the nature of the reaction.

(Nagpur University, Oct. 92)

**SOL.**



$$1.007825 + 18.998405 \rightarrow 4.002603 + 15.994915 + Q$$

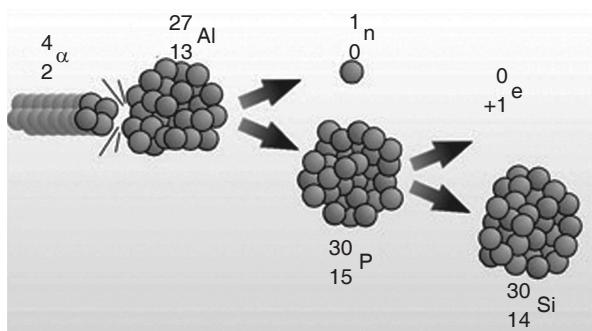
$$\begin{aligned} Q &= (+20.00623 - 19.997518)u = +0.008712 \text{ u} \\ &= +0.008712 \times 931 \text{ MeV} = +8.11 \text{ MeV} \end{aligned}$$

As  $Q$  is +ve, the reaction is exoergic.

## ARTIFICIAL RADIOACTIVITY

### 21.6 DISCOVERY

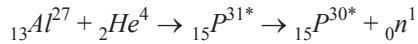
Artificial or induced radioactivity was discovered by Curie and Joliot in 1934. They were studying the disintegration of light elements by  $\alpha$ -particles. When boron and aluminium were bombarded with  $\alpha$ -particles, the target continued to emit radiations even after the source of  $\alpha$ -particles had been removed. With the help of magnetic deflection experiments and ionisation measurements it was found that the radiations consisted of particles having a positive charge and mass equal to electrons. These particles



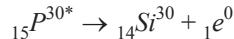
In 1934, Joliot-Curie Succeeded in Making the First Artificial Radioactive Nuclide.

are known as positrons. Curie and Joliot explained the phenomenon as follows: When the above elements are bombarded by  $\alpha$ -particles, an unstable nucleus was formed and this nucleus disintegrated spontaneously.

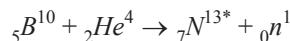
(i) In the case of aluminium, the nuclear reactions are



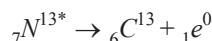
$P^{30}$  is radioactive. The half-life period of radiophosphorus is about 3 minutes and it disintegrates producing a stable isotope of silicon and ejecting a positron.



(ii) In the case of boron,



The nitrogen is radioactive and decays with a half-life of 10.1 minutes into a stable isotope of carbon with the emission of a positron.



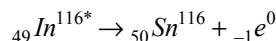
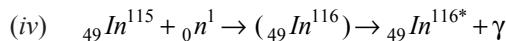
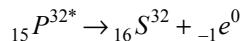
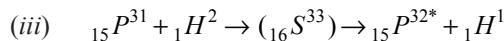
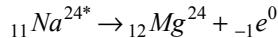
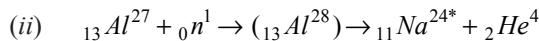
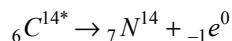
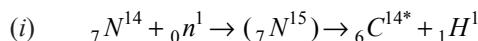
Curie and Joliot verified their explanation by chemical analysis. In the boron reaction, the target used was boron nitride ( $BN$ ). After irradiation with  $\alpha$ -particles for a few minutes, it was heated with caustic soda. One of the products of this chemical reaction was gaseous ammonia  $NH_3$ . When the various products were tested for positron activity, it was found that only  $NH_3$  had it. This indicated that nitrogen was the radioelement produced in the experiment. Its half-life was found to be the same as that produced in other irradiated boron targets.

This phenomenon, by which even light elements could be made artificially radioactive, is called *induced radioactivity*. The artificial radioactivity follows the same laws of decay as natural radioactivity. Unlike natural radioactive substances which emit  $\alpha$ ,  $\beta$  and  $\gamma$ -rays, artificial radioactive substances emit electrons, neutrons, positrons, or  $\gamma$ -rays.

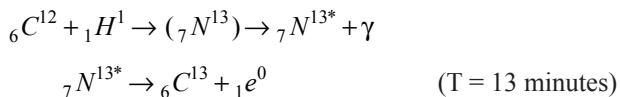
Practically all elements can be made artificially radioactive by bombarding them with various particles.

(i) Electron emission takes place in artificial radioelements produced by  $(n, p)$ ,  $(n, \alpha)$ ,  $(d, p)$  and  $(n, \gamma)$  reactions.

### EXAMPLES

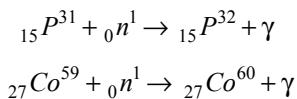


(ii) Positron emission takes place in artificial radioelements produced by  $(p, \gamma)$ ,  $(p, n)$ ,  $(\alpha, n)$ ,  $(d, n)$  and  $(\gamma, n)$  reactions.

**EXAMPLE**


## 21.7 PREPARATION OF RADIODEMENTS

(1) Artificially radioactive elements are now being produced by placing the elements in nuclear reactors. The large number of neutrons available in the nuclear reactor bombard these elements. Thus radiophosphorus can be obtained by neutron bombardment of  $P^{31}$ . A large number of isotopes are obtained from nuclear reactors by the neutron radiative capture  $(n, \gamma)$  process.



The most important radioisotopes so produced are Tritium,  $H^3$ , Radiosodium,  $Na^{24}$ , Radio-phosphorus,  $P^{32}$ , Radio-cobalt,  $Co^{60}$ , Radio-carbon,  $C^{14}$ , Radio-Iodine,  $I^{131}$  and Radio-sulphur,  $S^{35}$ .

(2) Another method of preparing a radio-element is to bombard the element with deuterons from a cyclotron. Radio-sodium can be obtained by bombarding  $Mg^{24}$  with deuterons. The reaction in this case is  $(d, \alpha)$  reaction.

The half-life of the radio-isotopes ranges from a fraction of a second to several thousand years.

## 21.8 APPLICATIONS OF RADIOISOTOPES

**Tracer technique.** As tracers, radio isotopes have found wide application in agriculture, medicine, industry and research. To study the dynamics of a particular system, a radio-isotope is added and the course of the isotope is studied to understand the whole system. For instance,  ${}_8O^{15}$  mixed with normal air breathed by a patient, enables radio-detection of the path of inhaled oxygen in the lungs. Phosphorus uptake by plants from the soil is studied using a phosphate fertilizer containing the radio-isotope  $P^{32}$ . Brain and thyroid tumours are detected using  $I-131$ .

(1) **Medical Applications. (Medical Radiation Physics)**  
Medical applications of radio-isotopes are divided into two parts:  
(i) *diagnosis* and (ii) *therapy*.

(i) **Diagnosis**

- Radioactive-isotopes can be introduced into the body in chemical forms that have an affinity for certain organs, such as bone or the thyroid gland. A sensitive detector (called a “gamma-ray camera”) can observe the radiations from the isotopes that are concentrated in the organ and can produce an image that shows how the activity is distributed in the patient. These detectors are capable of determining where each gamma-ray photon originates in the patient.



Rosalind Yalow (1921-, United States). After receiving her Ph.D. in nuclear physics, she researched the medical applications of radioactive isotopes. She developed the technique of radioimmunoassay, which uses radioactive tracers to measure small amounts of substances in the blood or other fluids. She first applied this technique to study insulin in the blood of diabetics. Her development of this technique was recognized with the award of the Nobel prize in medicine in 1977.

- Radio-isotopes are used to diagnose the nature of blood circulatory disorders, defects of bone metabolism, to locate tumours etc. For the study of the circulatory disorder in blood vessels, radio-sodium is injected. The constriction in a blood vessel is detected with a counter by surveying the area of the vessel.

Let us consider another example. Iodine taken into the body concentrates strongly in the thyroid gland, and an overactive thyroid holds it much longer than a normal thyroid does. If a patient drinks a glass of water containing  $I^{131}$  (as  $NaI$ ), the count rate in a detector placed near his neck 24 hr later immediately determines whether his thyroid is normal, overactive or underactive.

A similar technique is used to diagnose anemia. An anemic patient retains iron in the blood longer than a normal patient. So by injecting a radioactive iron isotope and later withdrawing blood and measuring its activity, anemia can be detected.

Radio-sodium has been largely used in medical research to study the action of various medicines. A certain amount of  $NaCl$  having radioactive sodium in it is mixed with the medicine and administered to the patient. With the help of radiations from radio-sodium, the circulation of the medicine in the body can be detected and hence its effects can be studied.

- Another technique that reveals a wealth of information is *positron emission tomography* (PET). In PET, the patient is injected with a positron-emitting isotope that is readily absorbed by the body.

Examples of isotopes used as  $^{15}O$  ( $t_{1/2} = 2$  min),  $^{13}N$  ( $t_{1/2} = 10$  min),  $^{11}C$  ( $t_{1/2} = 20$  min), and  $^{18}F$  ( $t_{1/2} = 110$  min).

These isotopes are produced with a cyclotron, and because of the short half-lives the cyclotron must be present at the site of the diagnostic facility. When a positron emitter decays, the positron quickly annihilates with an electron and produces two 511-keV gamma rays that travel in opposite directions. By surrounding the patient with a ring of detectors, it is possible to determine exactly where the decay occurred. From a large number of such events, the physician can produce an image that reconstructs the distribution of the radioisotope in the patient.

### (ii) Therapy

Radiation therapy takes advantage of the effect of radiations in destroying unwanted tissue in the body, such as a cancerous growth or an overactive thyroid gland. The effect of the passage of radiation through matter is often to ionize the atoms. The ionized atoms can then participate in chemical reactions that lead to their incorporation into molecules and subsequent alteration of their biological function, possibly the destruction of a cell or the modification of its genetic material.

**Example 1.** In therapy, radio-iodine is used in the treatment of thyroid gland.

An overactive thyroid gland is treated by giving the patient radioactive  $I^{131}$ , which collects in the thyroid. The beta emissions from this isotope damage the thyroid cells and ultimately lead to their destruction.

**Example 2.** Radio-cobalt and radio-gold are used in the treatment of cancer.

Certain cancers are treated by implanting needles or wires containing radium or other radioactive substances. The decays of these radioisotopes cause localized damage to the cancerous cells.

Other cancers can be treated using beams of particles that cause nuclear reactions within the body at the location of the tumor. Pions and neutrons are used for this purpose. The absorption of a pion or a neutron by a nucleus causes a nuclear reaction. The subsequent emission of particles or decays by the reaction products again causes local damage that is concentrated at the site of the tumor, inflicting maximum damage to the tumor and minimum damage to the surrounding healthy tissue.

**Example 3.** Radio-phosphorus is used in the treatment of skin diseases.

(2) **Agriculture.** In agriculture, the radio-isotopes help to raise crop yields. For example, in plant physiology  $P^{32}$  is incorporated in the fertilizer and added to the soil. The plant and soil are measured from time to time for activity. Phosphorus is taken up by the plant during its growth. It is also possible to find out whether a particular plant requires root feeding or foliage feeding.

(3) In fundamental scientific research, the  $\alpha$ ,  $\beta$  and  $\gamma$  radiations emanating from radio-isotopes are used for testing of materials. The material under test is interposed between the source (radioisotope) and a radiation detector. The change in counting rate in the detector gives a measure of the physical properties and structure of the material.

(4) Radio-isotopes are widely used in molecular biology for producing destructive effects. Radiations sterilize pharmaceuticals and surgical instruments. Large doses of radiations kill bacteria and insects.

(5) Certain perishable cereals exposed to radiations remain fresh beyond their normal life span. Very small doses of radiation prevent sprouting and spoilage of onions, potatoes and gram. Storage time can be enhanced by prior irradiation of the food-item.

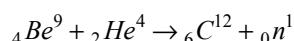
(6) **Radio-carbon Dating.** One of the interesting properties of radionuclides is that they have constant half-life irrespective of the state of chemical combination of the atoms in which they reside. This property finds applications in *dating or age estimation*.  ${}^6C^{14}$ , a carbon isotope is used in archaeological dating. This isotope has a half-life of 5730 years. The interaction of cosmic rays with the atmospheric gas atoms produces the radioactive isotope  $C^{14}$ . This is assimilated by living organism along with the more abundant  $C^{12}$ . The production of  ${}^6C^{14}$  and its subsequent decay are in equilibrium in the earth's atmosphere. Hence  ${}^6C^{14}$  is present in equilibrium concentration in all living plants. However, after the death of the creature or plant, the intake of  ${}^6C^{14}$  stops. With the death of organism,  $C^{14}$  begins to decay. Hence the number of  $C^{14}$  atoms to the number of stable  $C^{12}$  atoms in the object decreases with time after death. Thus the time elapsed since death can then be estimated by measuring this ratio. Using this procedure, ages of wooden slabs from tombs, charcoal used for old paintings etc., have been studied.

(7) **Industry.** The tracer technique is used in industry, e.g., to know how much a given bearing will wear out under particular conditions of running. The bearing is made of radioactive steel. A measurement of radioactivity in the lubricating oil is a measure of the extent to which wearing out has taken place.

## THE NEUTRON

### 21.9 THE DISCOVERY OF THE NEUTRON

In 1932, James Chadwick observed that when beryllium was bombarded with  $\alpha$ -particles, particles with about the same mass as protons were emitted. These emitted particles carried no electrical charge. Hence they were called neutrons. The nuclear reaction is represented by



Here  ${}_0n^1$  represents a neutron with zero charge and mass number 1. These neutrons have nearly the same mass as protons. Chadwick confirmed his hypothesis by the following experiment.

$\alpha$ -particles from polonium deposited on a disc  $D$  bombard a beryllium target  $T$  (Fig. 21.8). Both the disc and the target are enclosed in an evacuated chamber  $B$ . The neutrons released from the reaction escape through the thin wall of this chamber and enter an ionisation chamber  $C$  through a

thin window  $W$ . The ionisation chamber is connected to an amplifier and a counter. Since neutrons are uncharged, neutrons by themselves cannot produce ionisation in the chamber directly. When the radiations from  $B$  entered  $C$ , very few pulses were registered per minute. The pulses are due to the ions ejected by the neutrons from the walls of the chamber.

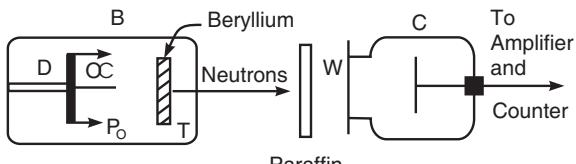


Fig. 21.8

Sir James Chadwick  
(1891 – 1974)

On placing a lead plate in front of  $W$ , the number of counts per minute did not increase appreciably. However when a paraffin slab was placed in front of  $W$ , the number of counts increased enormously. This is due to the fact that the neutrons eject fast moving protons from paraffin. These protons enter the ionisation chamber through  $W$  and are counted.

## 21.10 BASIC PROPERTIES OF THE NEUTRON

- (1) The neutron is one of the fundamental particles constituting the nuclei of atoms.
- (2) **Neutron Charge.** Neutrons carry no charge. As they pass through matter, they are neither affected by the positive charge on the nucleus nor by the negative charge on the extranuclear electrons. The neutron, therefore does not lose energy on its passage through matter and is readily able to collide with nuclei bringing about various types of nuclear reactions.
- (3) **Decay of the neutron.** In a free state, the neutron is unstable, unlike the proton and the electron. The neutron decays with the emission of a  $\beta$ -particle plus a neutrino, leaving a proton.

$$n \rightarrow p + e^- + \nu + 782 \text{ keV}$$

Neutron has a half-life period ( $T_{1/2}$ ) of about 13 minutes. The main difficulty that stands in the way of the experimental determination of the neutron half-life is the fact that the mean-life of free neutrons, before they are captured by nuclei, is short as compared with their half-life.

- (4) **Spin and Magnetic moment of the neutron.** Heisenberg suggested that the spin of a neutron is  $\frac{1}{2}\hbar$ . This has been confirmed by experiments. We would expect that the magnetic moment of the neutron, which is uncharged should be zero. Actually the magnetic moment of the neutron is about  $-1.9$  nuclear magneton. According to Yukawa's meson field theory, neutron transforms into a proton and a negative  $\pi$ -meson for a fraction of time. Since this meson is lighter than the proton, its motion will be faster and the resultant magnetic moment is that of a revolving negative charge. Hence it is negative. The proton and  $\pi$ -meson also possess spin. Therefore, the proton and  $\pi$ -meson have intrinsic magnetic moments, that of the meson being large due to its small mass. This situation also contributes to a negative magnetic moment of the neutron.

### (5) Neutron diffraction.

**Principle.** According to de-Broglie's theory, all particles are associated with wave characteristics. This

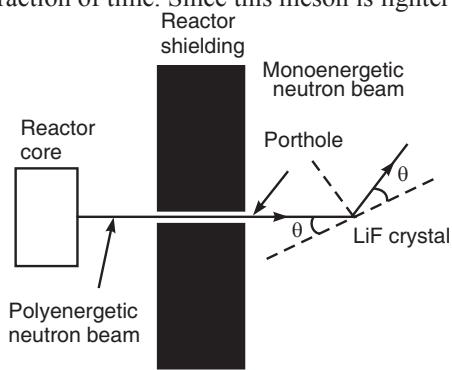


Fig. 21.9

wavelength is given by  $\lambda = h/mv$  where  $h$  is Planck's constant and  $mv$  is the momentum of the particle. Thermal neutrons from nuclear reactors have de-Broglie wavelengths of the order 1.82 Å, and are especially suited for crystal structure studies. Neutrons are diffracted by crystals similar to X-rays. Experimental techniques developed for X-ray diffraction can be adopted for this type of work.

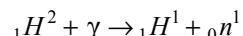
**Experimental arrangement.** Experiments use mono-energetic beams of neutrons and analyse their passage through the crystal. The neutrons emerging from a nuclear reactor through a porthole (Fig. 21.9) have a wide spectrum of energy. The neutron beam is not monochromatic. The neutron beam contains a spectrum composed of many de Broglie wavelengths. When the neutron beam from the reactor falls on the  $LiF$  crystal, the neutrons observed in the symmetric direction correspond only to the wavelength  $\lambda$  given by Bragg's condition  $2d \sin \theta = n\lambda$ . Therefore they have a well-defined energy and momentum. The crystal then acts as an *energy filter* or *monochromator*. The monoenergetic neutron beam is in turn used to study other materials, by diffraction, or to analyse nuclear reactions involving neutrons of definite energy.

#### Advantages

- (1) Neutron diffraction is one of the most powerful means of studying crystal structure.
- (2) Neutron diffraction has proved to be a very valuable technique for the study of molecular structure. X-ray and electron diffraction phenomena depend mainly on the number of orbital electrons of the scattering atom. Hence these techniques are not useful to locate the position, in the crystal lattice, of hydrogen atoms, which possess only single orbital electrons.

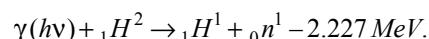
On the other hand, the diffraction of neutrons results from their scattering by the nucleus and the effect due to this, even with light atoms like hydrogen or beryllium, is quite appreciable. Therefore, neutron diffraction studies make it possible to identify the position of such light atoms in crystals.

(6) **Neutron mass.** Since the neutron is not a charged particle, its mass could not be determined from deflection measurements in electric and magnetic fields. A precise estimate of the mass of the neutron is obtained from the photo-disintegration of the deuteron discovered by Chadwick and Goldhaber. This method is based on the reaction



The minimum energy  $h\nu$  of the  $\gamma$  photon required to break up the deuteron into a proton and a neutron is known to be 2.225 MeV. The masses of the deuteron and the proton are accurately known from mass spectrograph measurements. Hence  $m_n$  can be calculated from the above reaction.

**EXAMPLE.** In the photodisintegration equation for the deuteron the threshold energy is 2.227 MeV, viz.

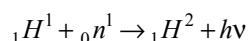


From the mass spectrometer the mass difference between the doublet  $2(_1H^1)$  and  $D(_1H^2)$  is  $1.5380 \times 10^{-3} u$ . If the mass of hydrogen atom is  $1.007825 u$ , calculate the mass of the neutron.

**SOL.** The mass of the neutron is given by

$$\begin{aligned} M(_1H^2) - M(_1H^1) + \frac{2.227}{931}u &= M(_1H^2) - 2M(_1H^1) + M(_1H^1) + 0.002392u \\ &= -0.001538 + 1.007825 + 0.002392 = 1.008679u. \end{aligned}$$

Bell and Elliot used a more accurate method to measure the mass of the neutron. This is based on the reaction



*i.e.*, when a proton captures a neutron to produce a deuteron nucleus, a  $\gamma$ -ray of energy equal to the binding energy is emitted. The  $\gamma$ -ray energy was measured accurately which was found to be 2.230 MeV = 0.002395 amu. We know the masses of  ${}_1H^1$  and  ${}_1H^2$  in mass units and the mass of neutron could be evaluated as follows in atomic mass units

$${}_0n^1 = {}_1H^2 - {}_1H^1 + h\nu = 2.014735 - 1.008142 + 0.002395 = 1.008988 \text{ amu.}$$

In the above measurement, the neutron and proton are initially at thermal energies and no measurement of energy of incident particles is involved.

(7) **Biological effects of neutrons.** It has been found that neutrons produce intense biological effects, even greater than X-rays or  $\gamma$ -rays on normal and tumour tissues. Neutrons are absorbed by fleshy tissues and their destructive action on malignant tissues is very much greater than on normal tissues. It is found that cancerous tissues are destroyed by neutron dosages which are harmless to normal tissues. Hence neutrons are being applied to cancer therapy and patients suffering from cancer are now being treated regularly with neutrons.

(8) **Absorption of neutrons by matter.** The absorption of neutrons by matter is mostly due to their “short range interactions” with nuclei. This interaction may be regarded as a collision which may be either *elastic* or *inelastic*.

(i) **Elastic collision.** In an elastic collision, the neutrons are scattered according to the laws of conservation of energy and momentum. A part of the energy of the striking neutron is transferred to the struck nucleus, the lighter the nucleus, the greater being the energy imparted to it. Thus fast neutrons will slow down rapidly when they collide with light nuclei. This explains the slowing down action of hydrogenous materials. Light elements are therefore used to slow down neutrons and these are called *moderators*.

(ii) **Inelastic collisions.** This may be subdivided into two categories :

(a) *Partial inelastic collisions.* Here the neutron is scattered anomalously and the struck nucleus undergoes internal changes in energy and is raised to an excited state of higher energy. This is very prominent excepting in the case of some very heavy nuclei. This type of scattering may be detected either by measuring the energy of the scattered neutrons or by the  $\gamma$ -rays that are emitted by the nucleus when it returns to a state of lower energy.

(b) *Total inelastic collisions.* Here the neutron is captured by the nucleus leading to a real transformation. The capture takes place when the energy of the neutron corresponds to the energy of an excited state of the nucleus. The nucleus and the neutron are now said to be in tune and the capture is called a *resonance capture*. This kind of inelastic collision, known as the *capture process*, plays a prominent role in transmutations produced by neutrons.

### 21.10.1. Classification of Neutrons

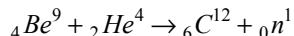
Neutrons are classified according to their kinetic energy.

(1) **Slow neutrons.** Neutrons with energies from zero to about 1000 eV are usually included in this category. Slow neutrons which are in thermal equilibrium with the medium through which they pass are called *thermal neutrons*. They exhibit a Maxwellian distribution of energy characteristic of the temperature of the medium.

(2) **Fast neutrons.** Neutrons with energies having range between 0.5 to 10 MeV are called fast neutrons. When neutrons pass through a material, they are slowed down by collisions with the nuclei of the material and lose a part of their energy. For light nuclei the loss is large and neutrons are slowed down quickly. Materials rich in hydrogen are very efficient in slowing neutrons and are known as *moderators*. Cadmium has the property of absorbing neutrons and is known as *absorber*.

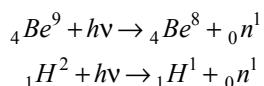
## 21.11 NEUTRON SOURCES

**(i) Radioactive sources.** Several ( $\alpha$ ,  $n$ ) reactions using  $\alpha$ -particles from natural radioactive substances are used to produce neutrons. The source consists of an intimate mixture of about 5 parts of fine beryllium powder to 1 part of radium. The majority of the neutrons is produced by the ( $\alpha$ ,  $n$ ) reaction given below :



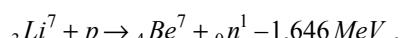
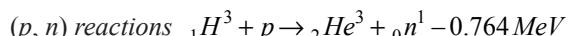
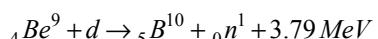
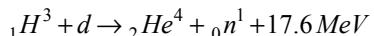
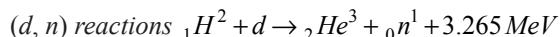
This neutron source is isotropic and produces neutrons over a wide energy range (0–13 MeV). Since the half-life of radium is 1600 years, this source is quite steady and gives a long time service.

**(ii) Photo-neutron Sources.**  $\gamma$ -rays from some radioactive substances have been used to produce neutrons by ( $\gamma$ ,  $n$ ) reactions in light elements. Neutrons can be produced by the disintegration of beryllium or deuterium by  $\gamma$ -rays by the reactions given below :



The special feature of these reactions is that they yield mono-energetic neutrons, unlike ( $\alpha$ ,  $n$ ) reactions from natural  $\alpha$ -emitters and produce a flux of  $10^7$  neutrons/sec. The energy of neutrons is, however, low and is not much greater than 1 MeV.

**(iii) Production of higher energy neutrons by accelerating machines.** Neutrons are produced by ( $d$ ,  $n$ ) and ( $p$ ,  $n$ ) reactions in particle accelerators. Van de Graffe machine, cyclotron and other charged particle accelerators serve as efficient neutron generators. Some of the reactions used for production of neutrons are :



**(iv) Reactor sources.** The most powerful source of neutrons available in these days is a *nuclear reactor*. The fission of each heavy nucleus by slow neutron capture is accompanied by the release of about 2.5 neutrons on the average. The neutrons produced from the fission of uranium atoms are fast and are slowed down in the moderator. The disadvantages of reactor sources are the wide spread in neutron energy and the impossibility of directly pulsing the neutron source.

### 21.11.1. Neutron Collimator

Dunning devised an arrangement known as *neutron howitzer* for collimating neutrons in a beam. He used the property of slowing down of neutrons in hydrogenous material like paraffin. The neutron source  $S$  (Ra + Be) is placed at the bottom of a barrel inside a paraffin block (Fig. 21.10). A cylindrical slit is cut in this block which is lined from all sides by cadmium sheet. Neutrons emitted in different directions are slowed down by paraffin. When they get slowed down they are effectively absorbed by Cd sheet which is very good absorber of slow neutrons. The Cd layer is covered by a Pb sheet just to absorb  $\gamma$ -rays which are emitted by Cd after neutron capture. However, the fast neutrons emitted along the axis of the barrel move outward in a collimated form.

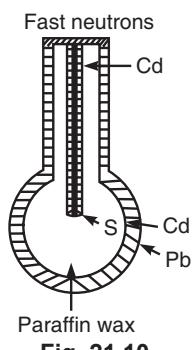
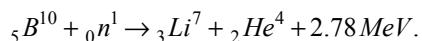


Fig. 21.10

## 21.12 NEUTRON DETECTION

Neutrons are uncharged particles and therefore produce negligible ionisation in passing through matter. They cannot be detected by G.M. counters or ionisation chambers or similar other devices which are based on detection by ionisation. Detectors used for neutrons use some indirect process in which neutrons produce charged particles. Also, fission may be utilized for detection of neutrons by placing fissionable material inside an ionization chamber and observing the ionization generated by the fission fragments.

**(1) Slow neutron detection—Boron detectors.** The neutrons are directed into a chamber filled with boron trifluoride gas. Alternatively, the walls of the chamber are lined with a compound of boron. The boron nucleus absorbs the neutrons and emits  $\alpha$ -particles which can be counted. The detection process is based on the reaction,



The cross-section for the above reaction follows the  $1/v$  law and falls off with increasing energy. Hence these counters are useful for slow neutrons. Ionisation chambers with boron coated electrodes or proportional counters filled with boron trifluoride vapour (enriched in  $B^{10}$ ) are in common use.

**(2) Fast neutron detection-Proton-recoil detector.** This method is based on the observation of ionisation produced by the recoiling protons in the elastic scattering of neutrons by hydrogenous materials. For this purpose, a proportional counter is filled with argon and a thin sheet of a hydrogenous material such as paraffin is placed at one end of the chamber. Fast neutrons striking the target (paraffin) cause the ejection of protons. The protons produce ionisation in their paths through the counter and so can be detected.

# EXERCISE

- For an exoergic reaction  
 (a)  $Q > 0$       (b)  $Q < 0$       (c)  $Q = 0$       (d) kinetic energy = 0  
 (B.U. M.Sc. 2014)
  - In exothermic nuclear reaction, energy is  
 (a) absorbed      (b) evolved      (c) absorbed or evolved      (d) no change  
 (B.U. 2006)
  - The total electric charge on either side of a nuclear reaction  
 (a) unequal      (b) equal      (c) zero      (d) none of the above  
 (B.U. M.Sc. 2006)
  - Identify X in the following reaction.  

$${}^5\text{B}^{10} + {}^2\text{He}^4 \rightarrow {}^6\text{C}^{13} + \text{X}$$
  
 (a)  ${}^0n^1$       (b)  ${}^1H^1$       (c)  ${}^1H^2$       (d)  ${}^1e^0$ .  
 (B.U. 2006)
  - The inverse of stripping reaction is  
 (a) compound nucleus reaction      (b) direct reaction  
 (c) knock-out reactions      (d) pick up reactions  
 (B.U. M.Sc. 2015)  
 [Ans. 1. (a), 2 (b), 3. (b), 4. (b), 5. (d)]
  - Write about the discovery of artificial transmutation by Rutherford.      (B.U. 2013)

7. Give an account of Bohr's compound nucleus formation hypothesis for nuclear reactions.  
*(Meerut 2003)*
8. Define threshold energy of a nuclear reaction. State the conservation laws that can be applied to a nuclear reaction and derive an expression for the  $Q$ -value.  
*(Osmania Univ. 2006)*
9. Derive  $Q$ -value equation in nuclear reaction.  
*(B.U. M.Sc. 2015)*
10. Briefly explain the different types of nuclear reaction.  
*(B.U. M.Sc. 2014)*
11. Explain about energy balance of nuclear reactions.  
*(B.U. 2014)*
12. Write the nuclear transmutation with  $\alpha$  particles, protons and neutrons.  
*(B.U. 2006)*
13. What is artificial radioactivity? How are radio elements prepared? List any four applications of radio isotopes.  
*(B.U. 2013)*
14. Write a concise account of the discovery, production, detection, properties and uses of neutrons.

*Uses of neutrons*

1. The neutrons produce intense biological effects even more than X-rays and  $\gamma$ -rays. That is why, they are used in medicine, specially in the treatment of cancer.
2. Fast and slow neutrons are used for artificial disintegration of nuclei and producing radio-isotopes.
3. Slow neutrons disintegrate Uranium-235 and Pu-239 in nuclear fission.
4. Neutrons are extremely useful for studying the properties of solids. Neutron diffraction studies are superior to those of X-ray diffraction.

## NUCLEAR FISSION AND FUSION

**After reading this chapter, you should be able to**

- ◆ Discuss how fission fuel reacts and describe what it produces
- ◆ Explain controlled and uncontrolled chain reactions
- ◆ Discuss controlled nuclear fission reactions that take place in nuclear reactors and the enrichment of nuclear fuels
- ◆ Discuss processes to achieve practical nuclear fusion energy generation
- ◆ Explain the working principles of fission bomb and hydrogen bomb
- ◆ Explain why a heavy nucleus is unsuitable as a moderator in a reactor
- ◆ Explain why a fusion reactor requires a high particle density, a high temperature, and a long confinement time.

## 22.1 NUCLEAR FISSION

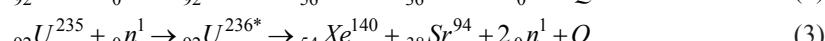
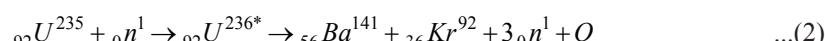
The process of breaking up of the nucleus of a heavy atom into two, more or less equal fragments with the release of a large amount of energy is called fission.

When uranium is bombarded with neutrons, a uranium nucleus captures a slow neutron, forming an unstable compound nucleus. The compound nucleus splits into two nearly equal parts. Some neutrons are also released in this process.

The schematic equation for the fission process is



${}_{92}U^{236*}$  is a highly unstable isotope, and  $X$  and  $Y$  are the fission fragments. The fragments are not uniquely determined, because there are various combinations of fragments possible and a number of neutrons are given off. Typical fission reactions are



Here,  $Q$  is the energy released in the reaction.

According to Eqn. (2), when  ${}_{92}U^{235}$  is bombarded by a slow moving neutron, the nucleus becomes unstable ( ${}_{92}U^{236*}$ ) and splits into  ${}_{56}\text{Ba}^{141}$  and  ${}_{36}\text{Kr}^{92}$  releasing 3 neutrons and energy  $Q$  (Fig. 22.1).

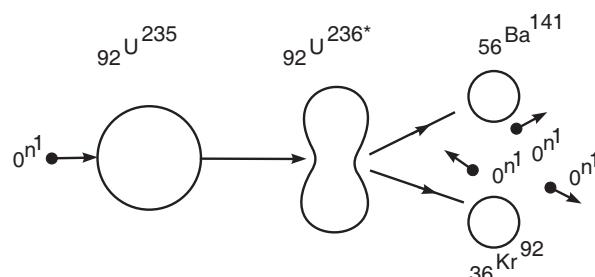


Fig. 22.1

### 22.1.1. Energy Released in Fission

We know that in the process of nuclear fission a large amount of energy is released. This energy is produced because the original mass of the nucleus is greater than the sum of the masses of the products produced after fission. The difference between these masses before and after fission is converted into energy according to Einstein's equation  $E = mc^2$ .

The energy liberated per fission can be calculated as follows :

Let us consider the fission of  ${}_{92}U^{235}$ . The fission reaction is shown in Fig. 22.1.

**Method I.** From the binding energy curve (Fig. 17.3), for a heavy nucleus (with  $A \sim 236$ ), the average binding energy per nucleon  $\bar{B}$  is  $\sim 7.6$  MeV. For a nucleus with a medium weight, it is about 8.5 MeV. Hence, the average binding energy per particle is about 0.9 MeV more in the fission products than in a compound nucleus (e.g.,  ${}_{92}^{236}\text{U}$ ). Thus, this excess of 0.9 MeV per nucleon must be liberated in the fission process. The total energy liberated per fission of  ${}_{92}^{236}\text{U}$  is given by

$$E_f = 236 (8.5 - 7.6) \text{ MeV} \sim 212 \text{ MeV.}$$

**Method II.** We shall calculate the total energy released by considering only the fission products (i.e., stable end products). Let us consider the reaction



The total mass of the nuclei before fission is

$$1.0087 + 235.0439 = 236.0526 \text{ u.}$$

The total mass of the nuclei after fission is

$$97.906 + 135.9072 + 2.0174 = 235.8306 \text{ u.}$$

The mass difference is 0.2220 u.

The energy release =  $0.2220 \times 931.3 = 206.7 \text{ MeV}$ .

#### **Energy released by 1kg of uranium.**

Number of atoms in 1kg of uranium =  $(6.023 \times 10^{26}) / 235$

Energy released in one fission = 200 MeV

$$\left. \begin{aligned} & \text{Energy produced} \\ & \text{by 1kg of uranium} \\ & \text{during fission} \end{aligned} \right\} = E = \frac{6.023 \times 10^{26}}{235} \times 200 = 5.128 \times 10^{26} \text{ MeV} \\ \therefore E = (5.128 \times 10^{26}) \times (1.6 \times 10^{-13}) \text{ J} \quad (\because 1 \text{ MeV} = 1.6 \times 10^{-13} \text{ J}) \\ = \frac{(5.128 \times 10^{26}) \times (1.6 \times 10^{-13})}{3.6 \times 10^6} \text{ k Wh} \quad (\because 1 \text{ k Wh} = 3.6 \times 10^6 \text{ J}) \\ = 2.26 \times 10^7 \text{ k Wh.} \end{aligned}$$

Thus the energy released by fission of 1 kg of  $U^{235}$  is  $2.26 \times 10^7 \text{ kWh}$ . Due to this reason, nuclear energy is being used for the generation of electricity. The most striking aspect of nuclear fission is the magnitude of the energy involved. Ordinary chemical reactions, such as those that participate in the combustion of coal and oil, liberate only a few electron volts per individual reaction. Most of the energy that is released during fission goes into the K.E. of the fission fragments. The emitted neutrons,  $\beta$  and  $\gamma$ -rays and neutrinos carry off perhaps 20% of the total energy.

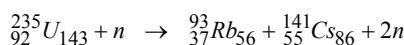
#### **22.1.2. Features of the Fission Reaction**

Three features of the fission reaction make it useful as a means to generate electrical energy:

(1) **Energy dissipation.** Most of the energy is released as kinetic energy of the fission fragments. These relatively heavy fragments do not travel very far through the reactor fuel element before they dissipate most of their kinetic energy in collisions with the atoms of the fuel element. The energy can be extracted as heat and used to boil water. The resulting steam can then be used in a conventional way to drive a turbine to generate electricity.

(2) **Neutron multiplicity.** The average number of neutrons produced is greater than one, making possible the chain reaction. How much greater than one it must be, in order to achieve a chain reaction, depends on the construction of the reactor.

(3) **Delayed neutrons.** The two neutrons emitted in the fission process



are *prompt neutrons*—they are emitted essentially at the instant of fission.

**Delayed neutrons.** About 1 percent of the neutrons in the fission process are *delayed neutrons* emitted following the decays of the heavy fragments. The delayed neutrons enable an operator or mechanical system to control the reaction and keep it from proceeding too rapidly.

**Bohr and Wheeler's theory of nuclear fission.** Bohr and Wheeler successfully explained the phenomenon of nuclear fission on *liquid drop model*. A liquid drop has a spherical form due to

internal molecular forces responsible for surface tension. According to their theory, an excited liquid drop may oscillate in a number of ways. On applying a large external force, the sphere may change into an ellipsoid. If the external force is sufficiently large, the ellipsoid may change into a dumb-bell shape and may even break at the narrow end into two portions.

The analogy may be extended to a nucleus which behaves like a liquid drop. When a nucleus absorbs a neutron, it forms a compound nucleus which is highly energetic (Fig. 22.2). The extra energy possessed by it comes mostly from the binding energy of the neutron absorbed by it. The extra energy may set up a series of rapid oscillations in the spherical compound nucleus shown as A in Fig. 22.2. As a result of these oscillations, the shape of the nucleus may change at times from spherical to ellipsoidal shown as B. If the extra energy is large, oscillations may be so violent that stage C and ultimately stage D may be approached. The nucleus is now dumbbell shaped (stage D). Each bell of the dumb-bell has now a positive charge and one repels the other. This results in a fission (stage E).

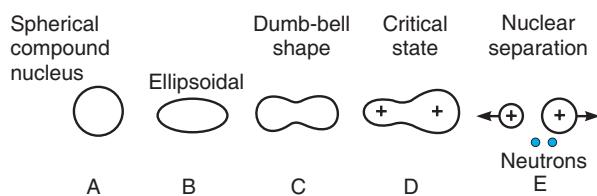
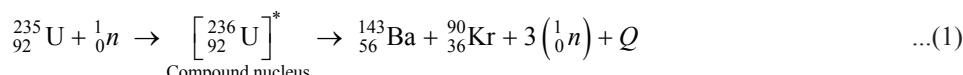


Fig. 22.2

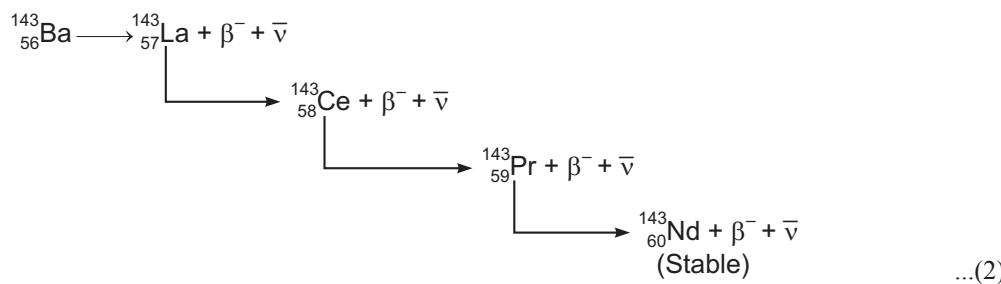
The nuclei that result from fission are called *fission fragments*. Usually fission fragments are of unequal size. A heavy nucleus undergoes fission when it acquires enough excitation energy to oscillate violently. Certain nuclei, notably  $^{92}_{\text{U}}\text{U}^{235}$  are sufficiently excited by the mere absorption of an individual neutron. Other nuclei, notably  $^{92}_{\text{U}}\text{U}^{238}$  require more excitation energy for fission than the binding energy released when another neutron is absorbed.  $^{92}_{\text{U}}\text{U}^{238}$  undergoes fission only by reaction with fast neutrons whose kinetic energies exceed about 1 MeV.

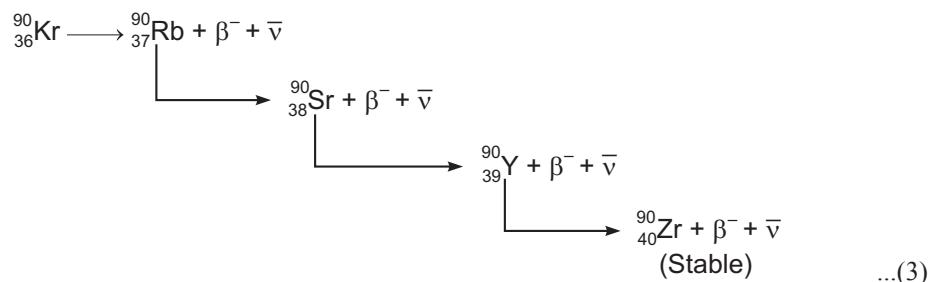
**EXAMPLE.** Explain mass yield in asymmetrical fission and discuss mass distribution of fission fragments for different fission chains of  $^{235}_{\text{U}}$ .  
(Mumbai, 2010)

**SOL.** Consider the following fission reaction:



The stable isotope of Barium with maximum  $A$  is  $^{138}\text{Ba}$ . The stable isotope of Krypton with maximum  $A$  is  $^{86}\text{Kr}$ . To get rid of excess neutrons, these fission fragments must exhibit much beta activity. Experimentally this is confirmed by the following beta decay observations:





Eq. (1) is only one of the many ways in which  ${}^{235}\text{U}$ -nucleus undergoes fission.

There are more than 30 modes of fission, in each of which a different pair of nuclei (fission-fragments) is formed. Eqs. (2) and (3) are examples of a *fission chain*. There are about 60 such chains known experimentally.

#### Mass Distribution of Fission Fragments for Different Fission Chains of ${}^{235}\text{U}$

The fission yield  $Y(A)$  is defined as the ratio of number of nuclei of mass number  $A$  formed in fission to the total number of fissions, i.e.,

$$Y(A) = \frac{\text{Number of nuclei of mass number } A \text{ formed in fission} (N_A)}{\text{Total number of fissions} (N_0)}$$

$$\text{Percentage fission yield } Y(A)\% = \frac{N_A}{N_0} \times 100\%$$

When this percentage of fission yield is plotted against mass number  $A$  for fission chains of  ${}^{235}\text{U}$ , we get a mass distribution of fission fragments (Fig. 22.3).

- The maxima lie near mass numbers 95 and 139 implying that fission is highly asymmetric.
- The yield is minimum for  $A \sim 117$ , which corresponds to symmetrical fission (equal masses).
- Some of the fragments may have as low a mass number as 72 or as high as 158.
- The *asymmetric mass distribution* of fission fragments involves the *shell structure* of nuclei and the existence of *magic numbers*. Corresponding to the peak ( $A \sim 95$  and 139) yields, the atomic numbers are  $Z \sim 42$  and 57. But the  $Z$ -value for minimum yield ( $A \sim 117$ ) is  $Z \sim 50$  that corresponds to a magic number. Hence its nuclear binding is strong, corresponding to a *closed nuclear shell*. Necessarily, therefore, the fission yield would be minimum. Thus, asymmetric fission is more probable for uranium.

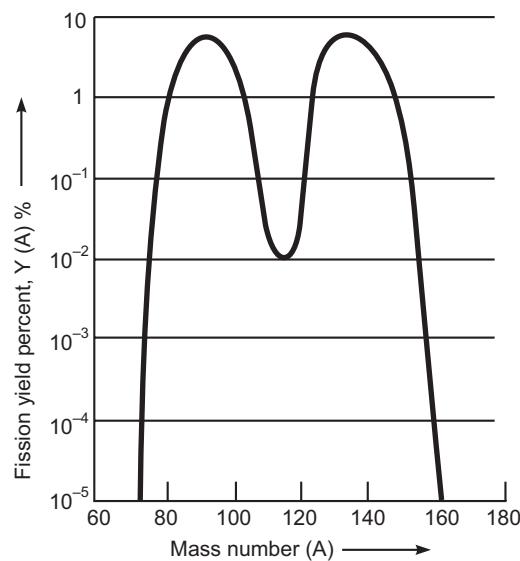


Fig. 22.3

## 22.2 CHAIN REACTION

*A chain reaction is a self-propagating process in which number of neutrons goes on multiplying rapidly almost in geometrical progression during fission till whole of fissile material is disintegrated.*

**Example :** Suppose a single neutron causing fission in a uranium nucleus produces 3 prompt neutrons. The three neutrons in turn may cause fission in three uranium nuclei producing 9 neutrons. These nine neutrons in turn may cause fission in nine uranium nuclei producing 27 neutrons and so on. The number of neutrons produced in  $n$  such generations is  $3^n$ . The ratio of secondary neutrons produced to the original neutrons is called the *multiplication factor* ( $k$ ).

Consider 1 kg of  $_{92}U^{235}$  which contains  $6.023 \times 10^{26}/235$  or about  $25 \times 10^{23}$  atoms. Suppose a stray neutron causes fission in a uranium nucleus. Each fission will release on the average 2.5 neutrons. The velocity of a neutron among the uranium atoms is such that a fission capture of a thermal neutron by the  $_{92}U^{235}$  nuclei takes place in about  $10^{-8}$  s. Each of these fissions, in turn, will release 2.5 neutrons. Let us assume that all these neutrons are available for inducing further fission reactions. Let  $n$  be the number of stages of fission captures required to disrupt the entire mass of 1 kg of  $_{92}U^{235}$ . Then

$$(2.5)^n = 25 \times 10^{23} \text{ or } n \approx 60.$$

The time required for 60 fissions to take place =  $60 \times 10^{-8}$  s =  $0.6 \mu$  s.

Since each fission releases about 200 MeV of energy, this means that a total of  $200 \times 25 \times 10^{23} = 5 \times 10^{26}$  MeV of energy is released in  $0.6 \mu$  s. The release of this tremendous amount of energy in such a short time interval leads to a violent explosion. This results in powerful air blasts and high temperature of the order of  $10^7$  K or more, besides intense radioactivity. The self-propagating process described here is called a *chain reaction*.

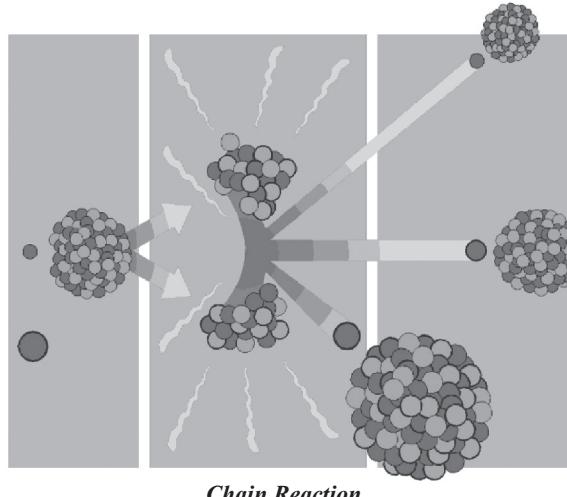
Two types of chain reaction are possible. In one, the chain reaction is first accelerated so that the neutrons are built up to a certain level and there after the number of fission producing neutrons is kept constant. This is *controlled chain reaction*. Such a controlled chain reaction is used in nuclear reactors. In the other type of chain reaction, the number of neutrons is allowed to multiply indefinitely and the entire energy is released all at once. This type of reaction takes place in atom bombs.

**Multiplication factor ( $k$ )**. The ratio of secondary neutrons produced to the original neutrons is called the multiplication factor. It is defined as

$$k = \frac{\text{Number of neutrons in any one generation}}{\text{Number of neutrons in the preceding generation}}.$$

The fission chain reaction will be “*critical*” or steady when  $k = 1$ , it will be building up or “*Supercritical*” when  $k > 1$  and it will be dying down or “*Subcritical*” when  $k < 1$ .

**Critical size for maintenance of chain reaction.** Consider a system consisting of uranium (as fissile material) and a moderator. Even though each neutron that produces fission ejects 2.5 neutrons on an average, all of them are not available for further fission. The maintenance of the chain reaction depends upon a favourable balance of neutrons among the three processes given below :



*Chain Reaction*

- (1) The fission of uranium nuclei which produces more neutrons than the number of neutrons used for inducing fission.
- (2) Non-fission processes, including the radiative capture of neutrons by the uranium and the parasitic capture by the different substances in the system and by impurities.
- (3) Escape or leakage of neutrons through the surface of the system.

If the loss of neutrons due to the last two causes is less than the surplus of neutrons produced in the first, a chain reaction takes place. Otherwise it cannot take place.

The escape of neutrons takes place from the surface of the reacting body and fission occurs throughout its *volume*.

$\therefore$  Escape rate varies as  $r^2$  and production rate varies as  $r^3$ .

$$\therefore \frac{\text{Escape rate}}{\text{Production rate}} \propto \frac{1}{r}.$$

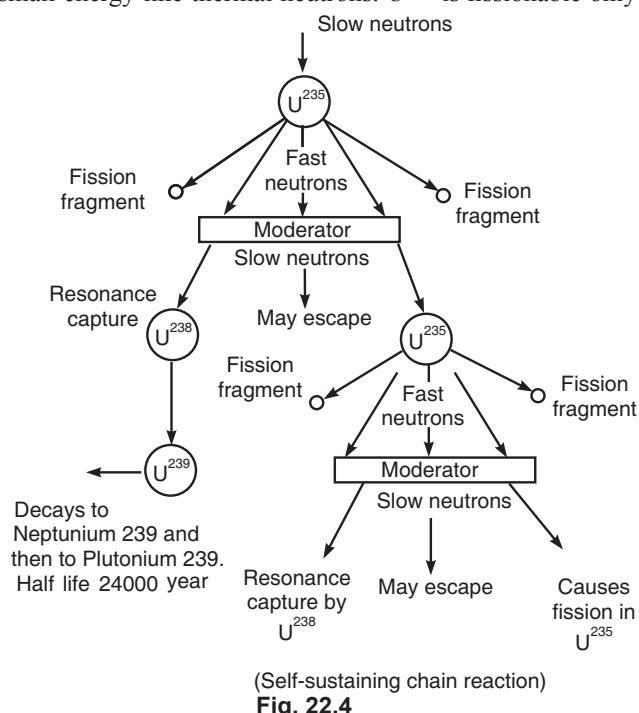
The larger the size of the body, the smaller is the escape rate. Thus it is clear that by increasing the volume of the system, the loss of neutrons by escape from the system is reduced. The greater the size of the system, the lesser is the probability of the escape of neutrons. In this case, the production of neutrons will be more than the loss due to other causes and a chain reaction can be maintained. Thus there is a critical size for the system.

**Definition.** Critical size of a system containing fissile material is defined as the minimum size for which the number of neutrons produced in the fission process just balance those lost by leakage and non-fission capture. The mass of the fissionable material at this size is called the *critical mass*.

- If the size is less than the critical size, a chain reaction is not possible.

**Natural uranium and chain reaction.** Natural uranium consists of 99.28% of  $U^{238}$  and 0.72% of  $U^{235}$ . As most of the mass of natural uranium consists of  $U^{238}$ , the neutrons released during nuclear fission will try to bombard the nuclei of  $U^{238}$  mostly and very few will bombard  $U^{235}$ .  $U^{235}$  undergoes fission even by neutrons of small energy like thermal neutrons.  $U^{238}$  is fissionable only with fast neutrons of energy 1 MeV or more. It has been found that very few neutrons can cause fission of  $U^{238}$  but neutrons of all possible energies can cause fission of  $U^{235}$ . Thus chain reaction is not possible in natural uranium.

A chain reaction can, however, be made to develop in natural uranium, if the fast neutrons from it are quickly reduced to thermal ones before they are lost through non-fission capture in the uranium, so that the chances of the thermal neutron fission of  $U^{235}$  go up. The neutron can be slowed down by distributing among lumps or rods of uranium a material called *moderator*. The moderators must not absorb the neutrons. The function of the moderator is to slow down the neutrons produced by fission by elastic collision. Materials used as



neutron moderators have a large inelastic scattering cross-section and, at the same time, a small neutron-capture cross-section. Commonly used moderators are graphite, heavy water ( $D_2O$ ), beryllium, beryllium oxide, hydrides of metals and organic liquids. The nuclei of these substances absorb neutrons only to a slight extent.

Fig. 22.4 shows a self-sustaining chain reaction.

A slow neutron bombards a  $U^{235}$  nucleus. The nucleus breaks into two fragments and in the fission process three fast neutrons are emitted. The neutrons are slowed down by the moderator. One neutron may escape. One neutron may be captured by  $U^{238}$  to form  $U^{239}$  which decays to  $Np^{239}$  and then to  $Pu^{239}$ . One neutron is still available for carrying on the chain reaction. It bombards  $U^{235}$  and the process is repeated.

#### 22.2.1. Atom Bomb

**Principle.** An atom bomb is a nuclear fission bomb. The principle of fission is made use of in the construction of the atom bomb. It is a device in which an *uncontrolled* chain reaction is built up in a fissionable material by means of fast neutrons. It then releases tremendous amount of energy in a very short time.

**Construction and working.** An atom bomb consists of two pieces of  $_{92}U^{235}$  (or  $_{94}Pu^{239}$ ) each smaller than the critical size and a source of neutrons. The two subcritical masses of  $U^{235}$  in the form of hemisphere are kept apart by using a separator aperture (Fig. 22.5).

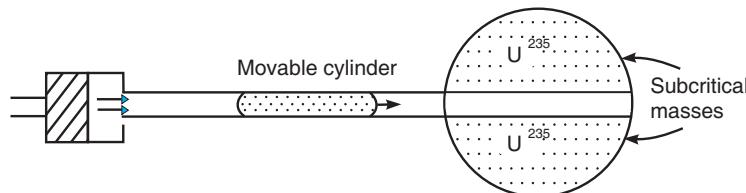


Fig. 22.5

When the bomb has to be exploded, a third well fitting cylinder of  $U^{235}$  (whose mass is also less than the critical mass) is propelled so that it will fit in or fuse together with the other two pieces. Now the total quantity of  $U^{235}$  is greater than the critical mass. Hence an uncontrolled chain reaction takes place resulting in a terrific explosion.

The explosion of an atom bomb releases tremendously large quantity of energy in the form of heat, light and radiation. A temperature of millions of degrees and a pressure of millions of atmospheres are produced. Such explosions produce shock waves. They are very dangerous because the waves spread radioactivity in air and cause loss of life. The release of dangerously radioactive  $\gamma$ -rays, neutrons and radioactive materials presents a health hazard over the surroundings for a long time. The radioactive fragments and isotopes formed out of explosion adhere to dust particles thrown into space and fall back to earth causing a radiation "fall-out", even at very distant places.

### 22.3 NUCLEAR REACTORS

We know that during the fission of  $U^{235}$  a large amount of energy is released. The atom bomb is due to an uncontrolled chain reaction. A very large amount of energy is liberated within an extremely small interval of time. Hence it is not possible to direct this energy for any useful purpose. But in a nuclear reactor, the chain reaction is brought about under controlled conditions. If the chain reaction is put under control, after some time a steady state is established. Under a steady state, the rate of energy production also attains a constant level. Such a device in which energy is released at a given rate is known as a *nuclear reactor*.

Nuclear reactors consist of five main elements:

- (1) The fissionable material called fuel,

- (2) Moderator,
- (3) Neutron reflector,
- (4) Cooling system, and
- (5) The safety and control systems.

**(1) The fissionable substance.** The commonly used fissionable materials are the uranium isotopes  $U^{233}$ ,  $U^{238}$ , the thorium isotope  $Th^{232}$ , and the plutonium isotopes  $Pu^{239}$ ,  $Pu^{240}$  and  $Pu^{241}$ .

**(2) Moderator.** The function of the moderator is to slow down the highly energetic neutrons produced in the process of fission of  $U^{235}$  to thermal energies. Heavy water ( $D_2O$ ), graphite, beryllium, etc., are used as moderators. Ideally, moderators have low atomic weight and low absorption cross-section for neutrons.

**(3) Neutron reflector.** By the use of reflectors on the surface of reactors, leakage of neutrons can be very much reduced and the neutron flux in the interior can be increased. Materials of high scattering cross-section and low absorption cross-section are good reflectors.

**(4) Cooling system.** The cooling system removes the heat evolved in the reactor core. This heat is evolved from the K.E., of the fission fragments when they are slowed down in the fissionable substance and moderator. The coolant or heat-transfer agent (water, steam,  $He$ ,  $CO_2$ , air and certain molten metals and alloys) is pumped through the reactor core. Then, through a heat exchanger, the coolant transfers heat to the secondary thermal system of the reactor.

**(5) Control and safety system.** The control systems enable the chain reaction to be controlled and prevent it from spontaneously running away. This is accomplished by pushing control rods into the reactor core. These rods are of a material (boron or cadmium) having a large neutron-absorption cross-section. These rods absorb the neutrons and hence cut down the reactivity. By pushing in the rods, the operation of the reactor can be made to *die down*, by pulling them out to *build up*. The safety systems protect the space surrounding the reactor against intensive neutron flux and gamma rays existing in the reactor core. This is achieved by surrounding the reactor with massive walls of concrete and lead which would absorb neutrons and gamma rays.

**Power reactor.** The heat generated in a nuclear reactor is used for producing power in a nuclear power plant. Fig. 22.6 shows the components of a power reactor. A quantity of enriched uranium in the form of pure metal or solution of a soluble salt in water constitutes the centre of the heat energy source. A large quantity of heat is produced in the fission process. The cadmium rods regulate the temperature, to a pre-determined value. If it is desired to bring down the temperature, the cadmium rods are pushed down further as to absorb more neutrons. If the temperature has to be raised, the cadmium rods are pulled up a little. A fluid is circulated through the shielded reactor and heat exchanger. The hot fluid, while flowing through the heat exchanger, converts water into steam. The steam produced runs conventional turbines to produce electricity.

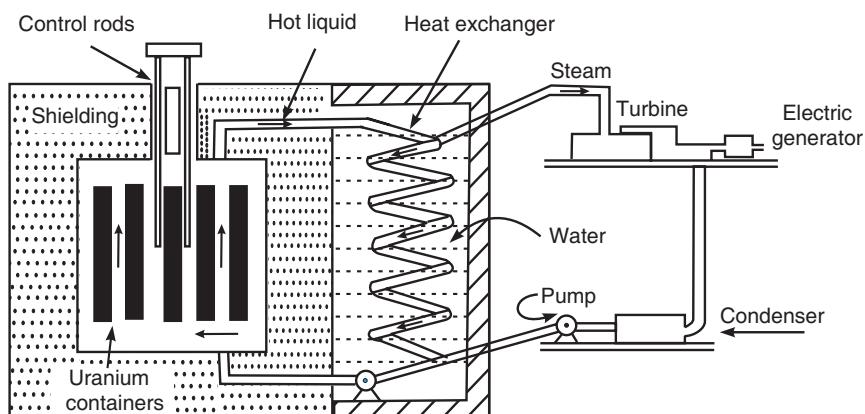
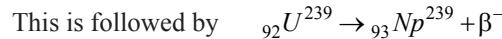
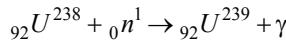
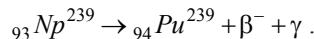


Fig. 22.6

**Breeder Reactor.** If a thermal reactor core with  $U^{235}$  fuel is surrounded by a blanket of a fertile material like  $U^{238}$ ,  $U^{238}$  can be converted into fissile fuel. Reactors of this type are called fuel producing reactors. The reactions are as follows :

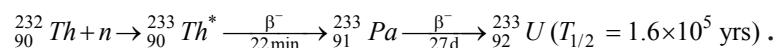


${}_{93}Np^{239}$  is also radioactive. It emits a  $\beta$ -particle to form plutonium.



This process of producing one type of fissionable material ( $Pu^{239}$ ) from a non-fissionable material ( $U^{238}$ ) is called *breeding* and the reactor a *breeder reactor*.

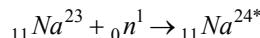
The breeding reactions for the fertile material  ${}_{90}Th^{232}$  are



### Uses of Nuclear Reactors.

(1) **Nuclear power.** Nuclear reactors are used in the production of electric energy.

(2) **Production of radioisotopes.** Nuclear reactors are useful in producing a large number of radio-isotopes. To produce radio-isotope, a suitable compound is drawn into the centre of the reactor core where the flux of neutrons may well be more than  $10^{16}/\text{m}^2/\text{sec}$ . Sodium-24 is manufactured in this way.



(3) **Scientific research.** Reactors produce a number of radioactive materials needed for research purposes. The reactors provide a huge source of neutrons. Using these neutrons, several useful radioisotopes have been artificially produced and several nuclear reactions have been studied. We may also study the effect of neutrons on biological tissues. Reactors may also be used to study radiation damage.

#### 22.3.1. Pressurized Water Reactor (PWR)

**Principle.** If the pressure on the water surface is increased, its boiling point increases so that it can have more quantity of heat energy per unit mass. This principle is utilised in PWR.

**Construction.** Fig. 22.7 shows a simple sketch of a PWR.

The fuel is in the form of uranium oxide ( $UO_2$ ). It is sealed in long, thin zirconium-alloy tubes that are assembled together with movable control rods into a core. The core is enclosed in a steel pressure vessel. Light water is used as coolant and moderator.

The water that circulates past the core is kept at a sufficiently high pressure, about 150 atm, to prevent boiling.

**Working.** The heat is extracted in a two-step process. Water circulates through the core under great pressure, to prevent

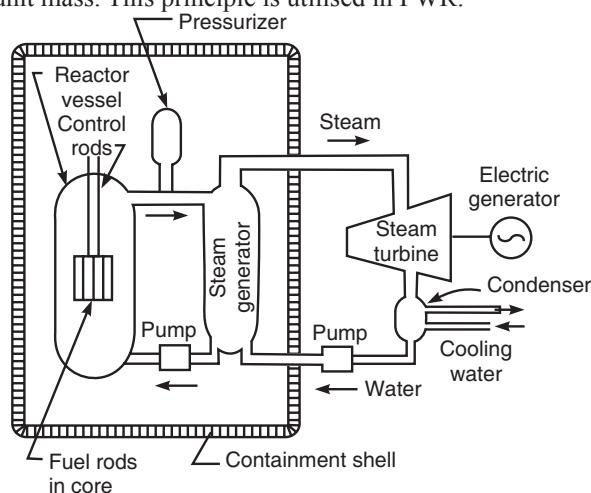


Fig. 22.7

its turning to steam. This hot water then in turn heats a second water system, which actually delivers steam to the turbine.

The water enters the pressure vessel at about  $280^{\circ}\text{C}$  and leaves at about  $320^{\circ}\text{C}$ . The water then passes through a heat exchanger (steam generator) which produces steam. The resulting steam then passes out of the containment shell. The containment shell serves as a barrier to protect the outside world from accidents to the reactor. The high pressure steam drives a turbine. The turbine drives an electric generator and produces electrical power. Steam after running the turbine gets condensed into water in the condenser. This water is again circulated through heat exchanger by means of a pump.

**Advantage.** Since the steam never enters the reactor core, it is not radioactive. So there is no radioactive material in the vicinity of the turbine.

**Disadvantages.** (1) The disposal of the exhaust or waste heat (primarily from the steam recondensing to water) generates considerable thermal pollution.

(2) Nuclear power plants are generally less efficient at converting fuel to electrical power compared with plants that burn fossil fuels, because nuclear plants operate at lower temperatures.

### 22.3.2. Boiling Water Reactor (BWR)

**Construction.** Fig. 22.8 shows a simple sketch of a BWR.

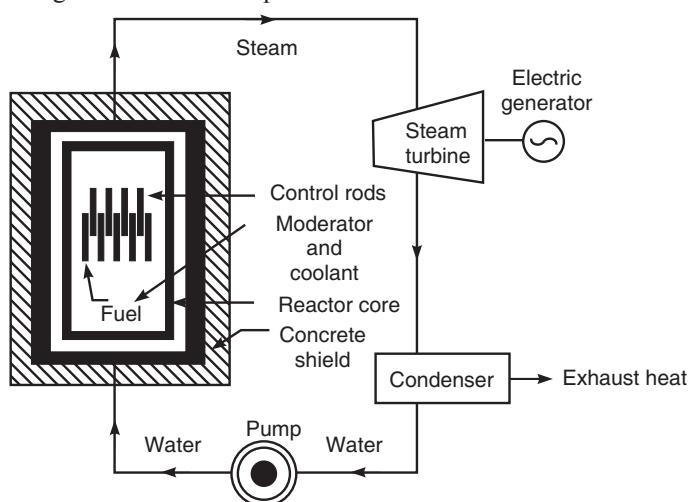


Fig. 22.8

Uranium is used as fuel in BWR. Water is used as a coolant and moderator.

**Working.** Circulating water absorbs heat produced by fission of the fuel and gets converted into high pressure steam in the reactor core itself. This steam runs the turbine which in turn produces electric power. Steam after running the turbine gets condensed into water in the condenser. This water is again circulated into the reactor core by using a pump.

**Disadvantage.** The steam leaving the reactor may be slightly radioactive. Therefore, the pipe lines and turbine assembly are properly shielded.

A rupture of the pipes near the turbines could result in a serious accident, with the spread of radioactive materials.

### 22.3.3. Fast Breeder Reactor

**Principle.** It is possible to run a reactor on fast neutrons by using a fuel with  $^{239}_{94}\text{Pu}$ . The fast neutrons released in the fission are used to transform  $^{238}_{92}\text{U}$  into  $^{239}_{94}\text{Pu}$ .

A fast breeder reactor is one designed to function with fissions induced by fast neutrons. The core of such a reactor will contain two materials, say, fissionable fuel  $^{239}_{94}\text{Pu}$  and the potential fuel  $^{238}_{92}\text{U}$ . Existing breeder reactors are all of the fast type. They use a mixture of  $\text{UO}_2$  and  $\text{PuO}_2$  as fuel. The coolant is liquid sodium.

**Construction and working .** Fig. 22.9 shows a liquid-sodium-cooled fast breeder reactor.

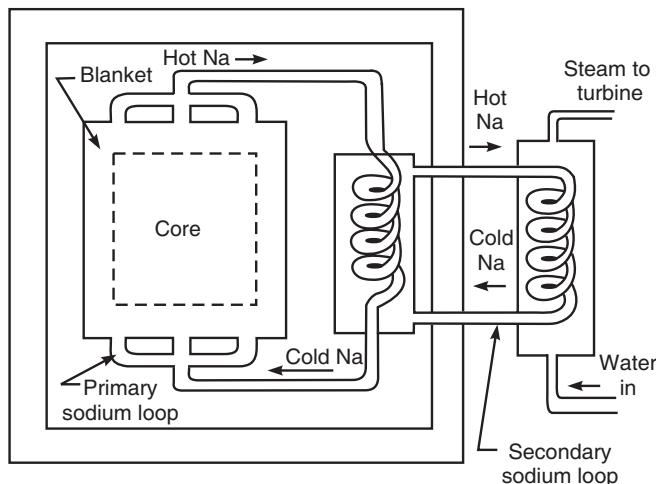


Fig. 22.9

The core may consist of  $^{235}\text{U}$  and  $^{239}\text{Pu}$ , while the blanket contains the fertile  $^{238}\text{U}$  that will breed into fissionable material. Liquid sodium is very effective as a medium for heat transfer. But its chemical reactivity and the radio-activity induced in it by neutron irradiation pose serious problems. To prevent contaminating the steam-generating system in the event of a sodium leak, an intermediate liquid sodium cooling loop is used between the primary loop that passes through the reactor core and the steam generator.

## 22.4 NEUTRON CYCLE IN A THERMAL NUCLEAR REACTOR

Let us study the behaviour of neutrons in a reactor assembly consisting of enriched uranium (a mixture of the uranium isotopes  $^{235}_{92}\text{U}$  and  $^{238}_{92}\text{U}$ ) and a moderator. For simplicity we assume that the reactor is infinitely large so that there is no leakage of neutrons through its surface. Let us start with the fission of a  $^{235}_{92}\text{U}$  nucleus by a thermal neutron. In this fission, the number of fast neutrons produced is  $v$ . Of these, some will cause a fission in  $^{238}_{92}\text{U}$ , and therefore create additional fast neutrons. We take account of this small increase in the neutrons by introducing a factor, namely,  $\varepsilon$ , called the *fast fission factor*. The number of fast neutrons available is now  $v\varepsilon$ , where  $\varepsilon > 1$ .  $\varepsilon$  is usually about 1.03.

These  $v\varepsilon$  neutrons diffuse through the pile and are slowed down by collisions with moderator nuclei. However, a few of them are captured by  $^{238}_{92}\text{U}$  before they are slowed down to thermal energies. This is known as resonance absorption. Therefore, we define *resonance escape probability*,  $p$ , as the fraction of neutrons escaping resonance absorption. Hence, the number of neutrons surviving thus far and reaching thermal energies is  $v\varepsilon p$ .  $p$  is usually 0.95.

Of these  $v\varepsilon p$  thermal neutrons, a fraction  $f$  called the *thermal utilisation factor* is absorbed in  $^{235}_{92}\text{U}$  nuclei while the rest are absorbed in moderator and other structural materials.  $f$  is always less than 1. Thus the number of  $^{235}_{92}\text{U}$  nuclei undergoing fission is  $v\varepsilon p f$ .

When these neutrons are absorbed by  $^{235}_{92}\text{U}$  in the fuel, some cause fission, whereas some produce other kinds of reactions such as capture. The fraction of neutrons causing fission is given by the ratio of fission cross-section ( $\sigma_f$ ) to the total absorption cross-section ( $\sigma_a$ ).

$\therefore$  the number of second generation thermal neutrons causing fission of  $^{235}_{92}\text{U} = \nu \epsilon p f \frac{\sigma_f}{\sigma_a}$ .

Now,  $\nu \frac{\sigma_f}{\sigma_a} = \eta = \text{the average number of fission neutrons released per neutron absorbed by a fissionable nuclide.}$

The ratio  $K$  of the number of fissions produced to a single neutron is

$$K = \eta \epsilon p f.$$

This is the *multiplication factor*. We have assumed that the reactor assembly is of an infinite size so that there is no leakage of neutrons. Hence

$$K_\infty = \eta \epsilon p f.$$

This relation is called the *four-factor formula*. In designing a nuclear reactor, the fundamental problem is to maintain  $K$  at 1 in order to achieve a self-sustaining chain reaction. Of the four factors,  $\eta$  and  $\epsilon$  are fixed for a given fuel. The other two factors  $p$  and  $f$  are made as large as possible by properly designing the reactor geometry, arranging the fuel, and choosing the moderator.

**EXAMPLE 1.** A reactor is developing energy at the rate of 3000 kW. How many atoms of  $U^{235}$  undergo fission per second? How many kilograms of  $U^{235}$  would be used in 1000 hours of operation assuming that on an average energy of 200 MeV is released per fission?

**SOL.** Rate of development of energy by the reactor = 3000 kW =  $3 \times 10^6 \text{ J s}^{-1}$

$$\begin{aligned} \text{Energy released per fission} &= 200 \text{ MeV} = 200 \times 1.6 \times 10^{-13} \text{ J} \\ &= 32 \times 10^{-12} \text{ J.} \end{aligned}$$

$$\therefore \text{No. of atoms undergoing fission per second} = \frac{3 \times 10^6}{32 \times 10^{-12}} = 9.4 \times 10^{16}$$

$$\begin{aligned} \text{No. of atoms undergoing fission in 1000 hours} &= (9.4 \times 10^{16}) \times (1000 \times 60 \times 60) \\ &= 3.384 \times 10^{23} \end{aligned}$$

According to Avogadro's hypothesis,  $6.025 \times 10^{26}$  atoms of  $U^{235}$  weigh 235 kg.

$$\begin{aligned} \therefore \text{Weight of } 3.384 \times 10^{23} \text{ atoms of } U^{235} &= \frac{235 \times (3.384 \times 10^{23})}{6.025 \times 10^{26}} \\ &= 0.1321 \text{ kg.} \end{aligned}$$

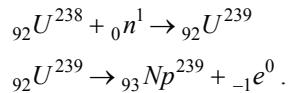
## 22.5 TRANSURANIC ELEMENTS

Elements with their atomic numbers greater than that of uranium ( $Z = 92$ ) are called transuranic elements. All these are man-made and radioactive. Some of these elements are fissionable and hence useful. The following is the list of transuranic elements.

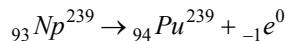
$Z = 93$	94	95	96	97	98
$Np$	$Pu$	$Am$	$Cm$	$Bk$	$Cf$
Neptunium	Plutonium	Americium	Curium	Berkelium	Californium
$Z = 99$	100	101	102	103	104
$Es$	$Fm$	$Md$	$No$	$Lw$	$Xv$
Einsteinium	Fermium	Mendelevium	Nobelium	Lawrencium	Kurchatovium

Such transuranic elements may be produced in the laboratory by the bombardment of certain heavy nuclides with neutrons. We give below typical methods of production, the reactions involved and the radioactive decays of two of these nuclides.

**(1) Neptunium ( $Z = 93$ ) :** When  $_{92}U^{238}$  is bombarded with slow energy neutrons, neptunium is formed according to the reaction



**(2) Plutonium ( $Z = 94$ ) :** Neptunium ( ${}_{93}Np^{239}$ ) is itself radioactive. It emits a  $\beta$ -particle and produces plutonium according to the reaction



Plutonium emits  $\alpha$ -particles and decays into  ${}_{92}U^{235}$  with a half-life of 24000 years.

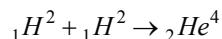


## 22.6 NUCLEAR FUSION

**Nuclear fusion.** In this process, two or more light nuclei combine together to form a single heavy nucleus. For example, when four hydrogen nuclei are fused together, a helium nucleus is formed. The mass of the single nucleus formed is always less than the sum of the masses of the individual light nuclei. The difference in mass is converted into energy according to Einstein's equation  $E = mc^2$ .

**EXAMPLE 1.** Consider a single helium nucleus formed by the fusion of two deuterium nuclei.

Mass of  ${}_1H^2 = 2.014102u$ ; mass of  ${}_2He^4 = 4.002604u$ .



The initial mass of 2 deuterium atoms  $= 2 \times 2.014102 = 4.028204 u$ .

Mass of helium atom  $= 4.002604 u$ .

Decrease in mass  $= 4.028204 - 4.002604 = 0.025600 u$ .

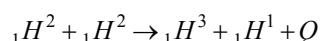
$\therefore$  Energy released  $= 0.025600 \times 931.3 \text{ MeV} = 23.84 \text{ MeV}$ .

Thus the energy released in fusion is 23.84 MeV.

**EXAMPLE 2.** A deuterium reaction that occurs in experimental fusion reactor is  $H^2(d, p) H^3$  followed by  $H^3(d, n) He^4$ . (a) Compute the energy release in each of these. (b) Compute the total energy release per gram of the deuteron used in the fusion. (c) Compute the percentage of the rest mass of deuteron released as energy. (d) Compare  $U^{235}$  fission with deuteron fusion as a source of energy release. Given

$H^2 = 2.014102 u$ ,  $H^3 = 3.016049 u$ ,  $H^1 = 1.007825 u$ ,  ${}_0n^1 = 1.008665 u$ ,  
 $He^4 = 4.002604 u$  and  $U^{235} = 235.0439 u$ .

**SOL.** (a) (i) The fusion reaction  $H^2(d, p) H^3$  is

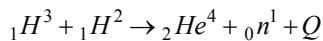


Mass decrease in the reaction

$$\Delta m = (2.014102 + 2.014102 - 3.016049 - 1.007825) u = 0.00433 u.$$

$\therefore$  Energy released  $= 0.00433 \times 931.3 \text{ MeV} = 4.032 \text{ MeV}$ .

(a) (ii) The reaction  $H^3(d, n)He^4$  is



$$\Delta m = (3.016049 + 2.014102 - 4.002604 - 1.008665) u = 0.018842 u.$$

$$\text{Energy released} = 0.018842 \times 931.3 = 17.58 \text{ MeV.}$$

$$\therefore \text{Total energy release} = 4.032 + 17.58 = 21.61 \text{ MeV.}$$

(b) This total energy release is from the fusion of 3  $_1H^2$  nuclei.

$$\therefore \text{Energy release per } H^2 = 21.61/3 = 7.203 \text{ MeV.}$$

$$\text{No. of nuclei in 1 gram of } H^2 = \frac{6.02 \times 10^{23}}{2.014102}$$

$$\therefore \text{Total release of energy from 1 gram of } H^2 = \frac{6.02 \times 10^{23}}{2.014102} \times 7.023 = 2.153 \times 10^{24} \text{ MeV.}$$

(c) Energy equivalent of one  $H^2$  nucleus =  $2.014102 \times 931.3$  MeV

$$\text{Average release of energy per } H^2 \text{ nucleus} = 7.203 \text{ MeV.}$$

$$\therefore \text{The percentage of the rest of mass of deuteron released as energy} = \frac{7.203}{2.014102 \times 931.3} \times 100 = 0.3840\%.$$

(d) In  $U^{235}$  fission, 200 MeV is released per uranium nucleus.

$$\therefore \text{The percentage of mass energy release in } U^{235} \text{ fission} = \frac{200}{235.0439 \times 931.3} \times 100 = 0.09137\%$$

$$\therefore \frac{\text{Energy release from } H^2 \text{ fusion}}{\text{Energy release from } U^{235} \text{ fission}} = \frac{0.3840}{0.09137} = 4.202.$$

### 22.6.1. Source of Stellar Energy

The temperature of the stars are very high and they radiate tremendous amount of energy. The sun is one of the innumerable stars. The sun radiates  $3.8 \times 10^{26}$  joules of energy each second. The origin of such a tremendous amount of energy is neither chemical nor gravitational. The fusion of protons is supposed to release the energy in the sun and in other stars. Bethe suggested the following carbon-nitrogen cycle as one of the most important nuclear reactions for release of energy by fusion.

**Carbon-Nitrogen Cycle.** The cycle is as follows (Fig. 22.10).



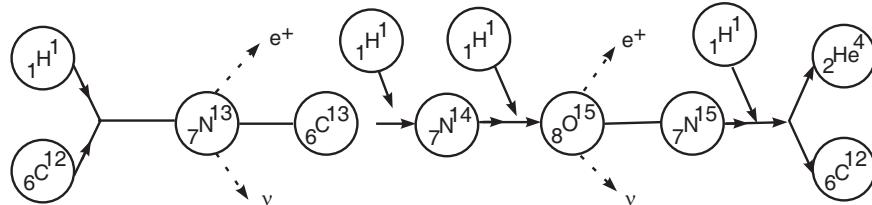
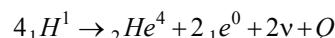


Fig. 22.10

In this cycle  $C^{12}$  acts like a catalyst.

The reaction cycle is essentially the reaction



The loss in mass is calculated as follows :

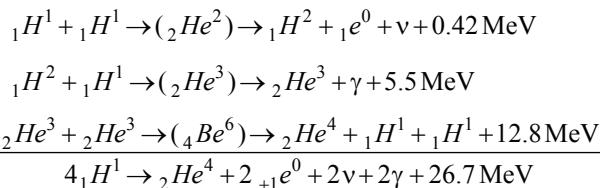
$$4 {}_1H^1 = 4.031300; {}_{\alpha}He^4 = 4.002603 \text{ and } 2 {}_1e^0 = 0.001098.$$

$$\therefore \text{Loss in mass} = 0.02756 u.$$

$$\therefore \text{Energy released} = 0.02756 \times 931 = 27.5 \text{ MeV}.$$

It is found that in one million years the sun loses about  $10^{-7}$  of its mass by the above process. Taking mass of the sun as  $2 \times 10^{30}$  kg and its present age as  $10^{10}$  years, it is estimated that the C-N cycle may keep going for another 30 billion years.

**Proton-Proton Cycle.** Recent modification of the estimates of the central temperature of the sun now favour the proton-proton chain. In the  $p-p$  chain, two protons first fuse to produce a deuterium nucleus which combines with another proton to yield  $He^3$ . Two  $He^3$  nuclei interact and form  $He^4$  and two protons. These reactions can be represented by the equations,



### 22.6.2. Thermonuclear Reactions

The source of stellar energy is fusion. This suggests that a large amount of energy can be obtained by nuclear fusion. But it is not easy to fuse the light nuclei into a single nucleus. The main difficulty in the fusion of nuclei is the electric force of repulsion between the positively charged nuclei. Fusion is possible when the K.E. of each of the nuclei is large enough to overcome the repulsion. Fusion reactions can take place only at very high temperatures (of the order of  $10^7$  to  $10^9$  K). Only at these very high temperatures, the nuclei are able to overcome their mutual Coulomb repulsion and enter the zone of nuclear attractive forces. Hence these reactions are called *thermonuclear reactions*.

A star is able to control thermonuclear fusion in its core because of its strong self-gravity. The thermonuclear reactions in the core of the sun cause high temperatures which generate strong outward pressures; these act against the sun's own gravity, preventing it from contracting, and holding it in equilibrium. The equation of stellar structure, set up by A.S. Eddington, relates the gravitational force in the star to the progressive changes of pressure from its centre outwards, the magnitude of pressure to density and temperature, and the fall of temperature outwards to the flow of energy from the interior to the surface. From these equations, stable models of stars emerge, with central temperatures high enough to start and sustain thermonuclear fusion. The key role, of course,

is played by the controlling force of gravity. The large mass of an astronomical system makes gravity the most important factor in determining its behaviour.

**Hydrogen bomb.** Hydrogen bomb is a device which makes use of the principle of nuclear fusion. The very high temperature required for an uncontrolled thermonuclear reaction is obtained by the detonation of an atom bomb. In this weapon, hydrogen is the core. The fission bomb produces a very high temperature, at which thermonuclear reactions start resulting in the fusion of hydrogen nuclei to form helium. Greater energy per unit mass is obtained from a hydrogen bomb than from a nuclear fission bomb.

**Controlled thermonuclear reactions.** A large amount of energy is released in a fraction of a second in a hydrogen bomb. If the thermonuclear reaction could be controlled to take place more slowly, the energy released can be used for constructive purposes. We know that very high temperatures are needed to bring about a nuclear fusion process. The main problem is to produce such a high temperature and to find a container for the gas which can stand this temperature. At this temperature the gas is highly ionised and is called *plasma*. One of the severe engineering problems is the design of a “*container*” in which a very hot plasma can be contained under high pressure to initiate a fusion reaction. Since almost any container would melt in the presence of a plasma, attempts are being made to contain and control plasmas trapped in a specially shaped magnetic field (Fig. 22.11). By increasing the field and changing the shape of the field, it is hoped that the plasma in this “*magnetic bottle*” can be raised to the required temperature and pressure for fusion reactions.

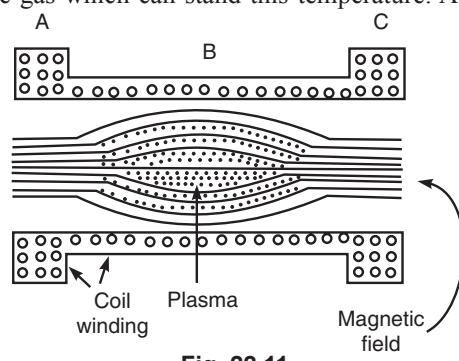


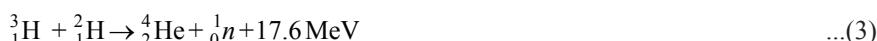
Fig. 22.11

Nuclear fusion as an energy source will be a boon to humanity because of the following reasons:

- (1) Hydrogen is available everywhere on this planet in various forms.
- (2) The lightness of the reactant nuclei makes the energy yield per unit mass of the reacting material much greater than that in nuclear fission process.
- (3) A fusion reactor does not leave behind as in fission reactor radioactive waste, the disposal of which poses a tremendous problem.

## 22.7 FUSION REACTORS

In the experiments conducted by bombarding the particles  ${}_1^2\text{H}$  and  ${}_1^3\text{H}$  with the deuterons, the following reactions have been found to occur:



Each of the deuteron-deuteron reactions (1) and (2) has a small cross-section. However, the deuteron-triton reaction, (3) is much more probable and also liberates a large amount of energy. The first fusion reactors are more likely to employ a deuterium—tritium mixture.

Three conditions must be met by a successful fusion reactor.

1. We know that for a fusion reaction to take place, the colliding particles should have enough energies to overcome the coulombian repulsive forces between them. This needs energies of the

order of 0.1 MeV which corresponds to a temperature of approximately  $10^8$  K. This means that for a fusion reaction to take place temperatures of this order should prevail initially. At such high temperatures, the atoms are ionized into positively charged ions (*i.e.*, nuclei) and electrons. This completely ionised and electrically neutral state of matter is called *plasma*.

2. There should be a fairly high density of the nuclei to ensure that collisions between nuclei are frequent. An important parameter for the fusion reaction is the *reaction rate* which can be represented as

$$R_{12} = n_1 n_2 \langle \sigma v_{12} \rangle \text{ reactions/m}^2/\text{sec.}$$

Here  $n_1, n_2$  represent particle densities of the colliding particles,  $v_{12}$ , the relative velocity of the particles and  $\langle \sigma v_{12} \rangle$  represents the average of the product of cross-section  $\sigma$  and the relative velocity of the particles.

If the colliding particles are identical, we have

$$R_{11} = \frac{1}{2} n^2 \langle \sigma v \rangle \text{ reactions/m}^2/\text{sec.}$$

3. The assembly of reacting nuclei must remain together for a long enough time to give off more energy than the reactor's operation takes.

It is difficult to achieve a high enough particle density  $n$  for a long enough time  $\tau$  in a hot plasma to obtain a net energy yield. According to the *Lawson criterion*, the product of the density and the confinement time (*i.e.*, the product  $n\tau$ ) of the plasma should be about  $2 \times 10^{22}$  particles  $\text{sm}^{-3}$  and  $10^{20}$  particles  $\text{sm}^{-3}$  for the deuteron-deuteron and the deuteron-triton fusion reactions respectively to occur in it.

### 22.7.1. Plasma Confinement

(i) **Magnetic confinement.** For the confinement of the plasma, one cannot use the walls of any vessel. Any contact with the wall will not only quickly cool the plasma but also cause the wall to evaporate. There are at present two schemes under investigation for confining the thermonuclear fuel: (i) *magnetic confinement* and (ii) *inertial confinement*.

In magnetic confinement, the plasma is confined by a carefully designed magnetic field. Two classes of field geometry are being studied. They are: (1) Magnetic bottle and (2) Tokamak.

1. **Magnetic bottle.** It is based on the reflection of an ion that moves in a magnetic field  $B$  whose lines of force converge. The magnetic force on an ion of velocity  $v$  is perpendicular to both  $v$  and  $B$ . So the force has a backward component in such a field as well as the inward component that causes the ion to follow a helical path around the lines of force (Fig. 22.12). If the backward force is powerful enough, the converging field acts as a mirror to reverse the ion's direction of motion. Thus, we can form a high density of magnetic field lines which reflects the particles back into the low-field region and is hence known as a *magnetic mirror*. A pair of magnetic mirrors constitutes a magnetic bottle (Fig. 22.13).

2. **Tokamak.** A 'tokamak' is a magnetic bottle in which the confining field is created by a combination of currents flowing in external coils and currents flowing in the plasma. It is a toroidal magnetic trap.

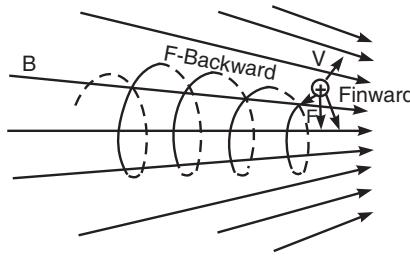


Fig. 22.12

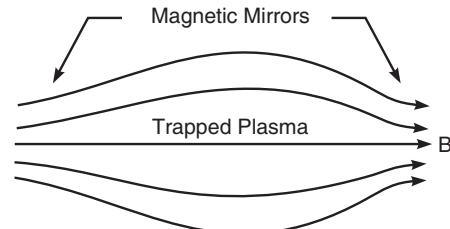
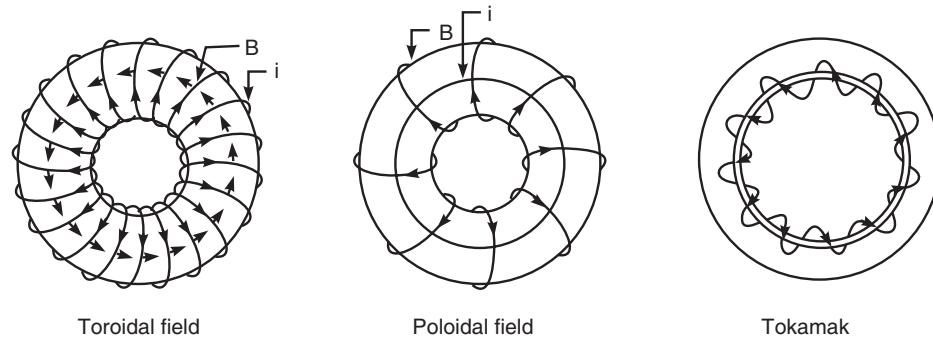


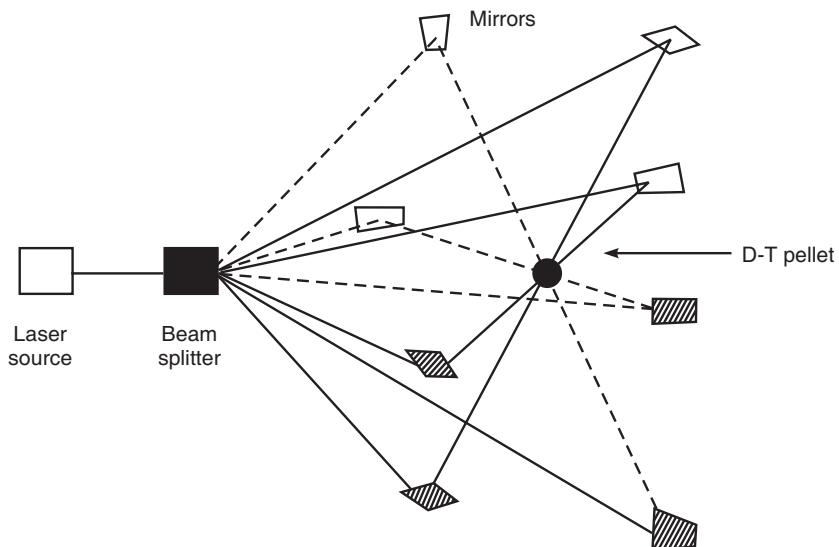
Fig. 22.13

The principle of tokamak method of magnetic confinement is shown in Fig. 22.14. A toroidal field is produced by a winding of coils, and a poloidal field is produced by an axial current. These two fields are combined in the tokamak design. (The current-carrying coils are not shown). The resulting field lines form a helix, through which the ions can travel in closed orbits.



**Fig. 22.14**

(ii) **Inertial confinement.** This scheme employs energetic beams to both heat and compress tiny deuterium-tritium pellets. Here, the fusion fuel (e.g., mixture of deuterium and tritium), in the form of a pellet, is imploded from all sides by energy sources such as laser beams, high energy electron or ion beams (Fig. 22.15). The intense compression pressures and the high temperatures produced in the pellet may produce conditions conducive to fusion. The difficulties in this approach are the low efficiencies of laser or other sources, and the need to produce stable symmetrical implosion.



**Fig. 22.15**

Both the above schemes are in advanced experimental stage. The experiments have shown that fusion reactions can be initiated but it has not been possible to extract the useful amounts of energy.

Fusion reactors have two important advantages over fission reactors as sources of mechanical and electrical power.

1. The fuel for fusion reactors (Deuterium and tritium) is much more abundant on earth than that for fission reactors (Uranium or Thorium).
2. Fusion reactors do not produce large amounts of radioactive residues which are the usual by products of the operation of fission reactors.

### EXERCISE

1. The average energy released per fission is  
 (a) 200 eV                         (b) 200 MeV                         (c) 20 MeV                         (d) 200 GeV  
 (B.U. 2015)
  2. Energy released by 1 kg of uranium is  
 (a)  $2.26 \times 10^7$  KWh      (b)  $2.26 \times 10^6$  KWh      (c)  $2.26 \times 10^5$  KWh      (d) 200 MeV  
 (B.U. 2015)
  3. Nuclear fission can be explained by  
 (a) shell model                     (b) liquid drop model      (c) quark model                         (d) Bohr atom model  
 (B.U. 2013)
  4. Fission fragments are generally radioactive as  
 (a) they have excess of neutrons  
 (b) they have excess of protons  
 (c) they are products of radioactive nuclides  
 (d) their total kinetic energy is of the order of 200 MeV  
 (GATE)
- [Hint. Usually the fission fragments are radioactive because they are products of radioactive nuclides]
5. The function of a moderator in a nuclear reactor is to:  
 (a) absorb  $\gamma$ -radiation  
 (b) react with uranium to release energy  
 (c) slow down fast neutrons so as to have greater probability for nuclear fission to occur  
 (d) provide neutrons for the fission process
  6. Atom bomb is based on  
 (a) controlled chain reaction                             (b) uncontrolled chain reaction  
 (c) thermo nuclear reaction                             (d) particle acceleration reaction  
 (B.U. 2013)
  7. The principle of \_\_\_\_\_ is used in the construction of atom bomb.  
 (a) Weiss                             (b) fusion                             (c) fission                             (d) Rutherford  
 (B.U. 2015)
  8.  ${}_1\text{H}^3 + {}_1\text{H}^2 \rightarrow {}_2\text{He}^4 + {}_0\text{n}^1 + \text{Energy}$ . This reaction takes place in  
 (a) atom bomb                             (b) hydrogen bomb      (c) proton-proton cycle      (d) carbon-nitrogen cycle  
 (B.U. 2011)
  9. The catalyst used in nuclear fusion reaction is  
 (a) nitrogen                             (b) helium                             (c) hydrogen                             (d) carbon  
 (B.U. 2013)
  10. Nuclear fusion produced by very high temperature is called  
 (a) chemical reaction      (b) physical reaction      (c) directional property      (d) thermonuclear reaction  
 (B.U. 2012)  
 [Ans. 1. (b), 2. (a), 3. (b), 4. (c), 5. (c), 6. (b), 7. (c), 8. (b), 9. (d), 10. (d)]
  11. With suitable example explain nuclear fission process. Also calculate the energy released during nuclear fission process.  
 (B.U. 2011)
  12. What do you mean by prompt and delayed neutrons?  
 (Rohilkhand 2007)

13. Explain nuclear fission on the basis of Bohr and Wheeler theory. (B.U. 2013)
14. What is chain reaction? Explain the critical size of the fissile material to maintain chain reaction. (B.U. 2006)
15. Draw a neat diagram of nuclear reactor and explain its working. (B.U. 2013)
16. Discuss the neutron cycle and possibility of a chain reaction in natural  $^{238}\text{U}$ . Obtain a four-factor formula and explain on what factors depend each term in the formula. What is the advantage of ‘lumping’ in a heterogeneous reactor? (Kerala)
17. Explain the working of atom bomb. (B.U. 2013)
18. Calculate the electrostatic potential energy between two equal nuclei produced by the fission of  $_{92}\text{U}^{235}$  at the moment of separation.

**Sol.** Before fission,  $_{92}\text{U}^{235} + {}_0\text{n}^1$ . i.e.,  $A = 235 + 1 = 236$  and  $Z = 92$ .

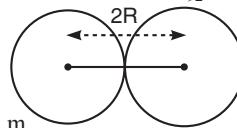
Two equal nuclei produced will each have  $A = 118$  and  $Z = 46$ .

The radius of the nucleus,  $R = R_0 A^{1/3} = (1.3 \times 10^{-15}) (118)^{1/3} = 6.376 \times 10^{-15} \text{ m}$

The electrostatic potential energy between the product nuclei is given by

$$E = \frac{1}{4\pi \epsilon_0} \frac{(Ze)(Ze)}{(2R)} = \frac{(46)^2 (1.6 \times 10^{-19})^2}{4\pi \times (8.854 \times 10^{-12}) \times 2 \times (6.376 \times 10^{-15})}$$

$$= 3.81952 \times 10^{-11} J = \frac{3.81952 \times 10^{-11}}{1.6 \times 10^{-13}} \text{ MeV} = 238.72 \text{ MeV}$$



19. Differentiate between nuclear fission and nuclear fusion. (Garhwal 97)
- [Ans. In fission, a large nucleus splits into smaller ones. In fusion, two or more light nuclei combine together to form a single heavy nucleus. In both processes, there is a decrease in mass in the nuclear reaction. The decrease in mass is converted into energy, according to Einstein’s equation  $E = mc^2$ .]
20. Explain the working of hydrogen bomb. (B.U. 2011)
21. What are thermonuclear reactions? Explain carbon-nitrogen cycle and proton-proton cycle as source of Stellar energy. (B.U. 2013)

## COSMIC RAYS

**After reading this chapter, you should be able to**

- ◆ Discuss properties of cosmic rays like energy, composition, latitude effect and altitude effect
- ◆ Distinguish between primary and secondary cosmic rays
- ◆ Explain why the mesons that reach the sea level are muons but not pions
- ◆ Explain why GM counter is quenched after every event
- ◆ Discuss the origin of cosmic rays in galaxy and solar system.

- Cosmic rays are highly penetrating radiations which are continually entering the earth's atmosphere in all directions from outer space. Cosmic rays consist of high energy charged particles. Most of these particles have energy of the order of 15 GeV.

### 23.1 LATITUDE EFFECT

The variation of cosmic ray intensity with latitude is shown in Fig. 23.1. The cosmic ray intensity is maximum at the geomagnetic poles ( $\lambda = 90^\circ$ ) and minimum at the geomagnetic equator ( $\lambda = 0$ ). The intensity remains constant between  $42^\circ$  and  $90^\circ$ . This variation of cosmic ray intensity with geomagnetic latitude is called latitude effect.

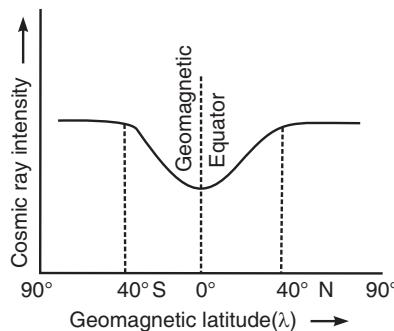


Fig. 23.1

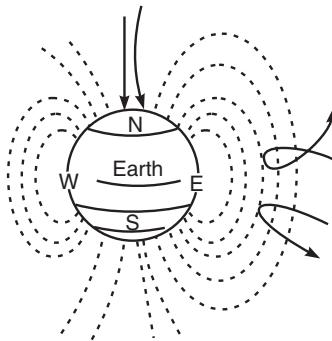


Fig. 23.2

**Explanation of latitude effect.** The latitude effect may be explained on the basis of *effect of earth's magnetic field on cosmic rays*. The presence of such an effect indicates that the cosmic rays are charged particles. The earth's magnetic field is directed from south to north. The earth's magnetic field at the equator is perpendicular to the direction of travel of charged cosmic ray particles. Therefore, it exerts maximum force ( $F = qv B \sin 90^\circ$ ) upon the particles which consequently suffer maximum deflection. They are deflected away from the earth (Fig. 23.2). The intensity of cosmic rays is therefore, minimum at the equator. At poles, cosmic ray particles move parallel to the earth's field. They, therefore, suffer minimum deflection. Hence the intensity of cosmic rays is maximum at the poles.

Mathematical analysis shows that for a given latitude ( $\lambda$ ), there is a minimum momentum ( $p_{\min}$ ) below which no particle can reach the earth. The minimum momentum  $p_{\min}$  at zenith is given by

$$p_{\min} = 15 \cos^4 \lambda \text{ (Be V/c)}$$

### 23.2 THE EAST-WEST EFFECT (AZIMUTH EFFECT)

It is found that the *number of cosmic ray particles coming from the west is greater than those coming from the east*. This effect is called east-west effect and is maximum at the equator. At the equator, the number of particles coming from west is 14% more than the particles coming from east. This phenomenon gave an evidence of the fact that *cosmic rays are composed predominantly of positively charged particles*.

**Explanation :** The charged particles approaching the earth's atmosphere are deflected by the earth's magnetic field in a direction perpendicular to the magnetic field and to the direction of their motion. According to Fleming's left hand rule, the positively charged particles are deflected towards the east by earth's magnetic field and thus appear to come from the west of the vertical (Fig. 23.3). Thus at any azimuth, more particles approach the earth from the west than from the east.

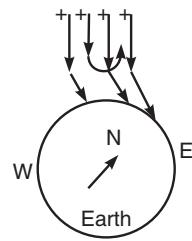


Fig. 23.3

If  $I_W$  and  $I_E$  are the intensities of cosmic rays coming from the west and east respectively, then the east-west asymmetry is defined as

$$\frac{I_W - I_E}{[(I_W + I_E)/2]}.$$

### 23.3 ALTITUDE EFFECT

The intensity of cosmic rays was measured by measuring the ionization produced by cosmic rays at different altitudes. The variation of cosmic ray intensity with altitude is shown in Fig. 23.4. It is observed that the intensity rises slowly upto a height of about 8 km, after which the rise becomes fast upto about 19 to 24 km. At heights above 24 km, the intensity starts decreasing gradually. The experiments were conducted at  $3^\circ$ ,  $38^\circ$  and  $51^\circ$  N and the results are similar.

The maxima of intensity is not at the top of the atmosphere but well below it. The reason is that at this height the primary cosmic rays produce quite a good number of secondary particles due to interaction with nuclei of atmospheric gases. Thus both primary and secondary rays are present in abundance at that height. With decrease of altitude the absorption increases and, therefore, intensity falls.

#### Longitude Effect

The intensity of cosmic rays also depends upon the longitude of the point of observation. It is called longitude effect. The intensity of cosmic radiation along the equator varies at different longitudes. This equatorial variation is attributed to the fact that the earth's magnetic field is not symmetrical about its axis.

#### Seasonal and Diurnal changes

There is a small seasonal change in cosmic ray intensity. This was established by Compton and Turner from their experimental data on the latitude effect. They found that the shape of the intensity-latitude curve depended slightly upon the season in which the data were taken. They were able to resolve the observed intensity-latitude variation into two components, viz., the *atmospheric effect* and the *pure magnetic effect*. Of the two components, the atmospheric effect alone showed seasonal variations. This change is attributed to the temperature variation in the meson-producing layers in the atmosphere.

There exists a variation of cosmic ray intensity with *solar* time of the day. This is called *diurnal change*. Experiments were conducted to study separately the soft and hard components of cosmic rays. The results obtained are:

- (1) There is a small diurnal variation in the softer component. The intensity reaches a maximum at about midday and a minimum at about midnight. But the intensity of the penetrating component is constant.
- (2) This variation of intensity with solar time appears to be independent of both latitude and altitude. Besides these, intensity-time variations of cosmic rays have been found to occur with 27 day, annual, 11 year etc., periodicity.

### 23.4 PRIMARY COSMIC RAYS

The cosmic rays which are just entering our earth's atmosphere from outer space are called primary cosmic rays. Primary cosmic rays consist mainly of positively charged atomic nuclei with

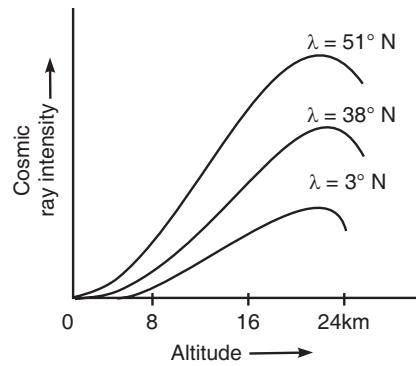


Fig. 23.4

Z upto about 40. About 90% of the primaries are protons, 9% helium nuclei and the remaining heavy nuclei. The energies of primary cosmic rays range from 1 MeV to  $10^{14}$  MeV.

### 23.5 SECONDARY COSMIC RAYS

When primary cosmic rays interact with the nuclei of atmospheric gases, secondary cosmic rays are produced. Below an altitude of 20 km, all cosmic radiation is secondary. On entering the atmosphere, the primary cosmic rays collide with air nuclei and produce mostly  $\pi$ -mesons (positive, negative and neutral) and some hyperons. The  $\pi$ -mesons so produced carry sufficient energy and decay into lighter particles :  $\mu$ -mesons, electrons, positrons, neutrinos and photons. All such particles constitute the secondary cosmic rays. At sea level the secondary cosmic rays contain nearly 70%  $\mu$ -mesons, 29% electron-positron pairs and 1% heavy particles. The mesons in the secondary cosmic rays constitute the *hard* component and the electrons, positrons and photons constitute the *soft* component.

### 23.6 COSMIC RAY SHOWERS

Cosmic ray shower is the name given to a grouping of cosmic ray particles that have been produced by some common cause and not by mere chance.

Rossi investigated the phenomenon by using an arrangement of three coincidence counters in a triangular pattern with a lead plate above the counters (Fig. 23.5). A coincidence discharge of the three counters can be produced only by the simultaneous passage of at least two particles, including the incident one. Rossi observed that an appreciable number of coincidence counts were registered. He found that the number of coincidences increased as the thickness of the lead plate was increased. From this result he inferred that the three counters were affected by two or more secondary particles produced simultaneously by a single cosmic ray as it penetrated the lead plate. Thus the production of showers in the lead plate was detected. Rossi's results were explained by Bhabha and Heitler and, independently, by Carlson and Oppenheimer. They gave the cascade theory of cosmic ray showers.

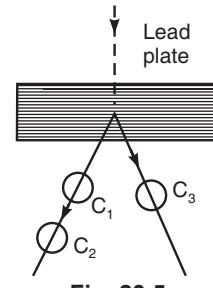


Fig. 23.5

**Cascade theory of showers.** Shower production involves two processes, *viz.*, *radiative collision* and *pair production*.

A high energy electron (or positron), present in cosmic rays, loses energy, when it encounters the atomic nuclei in earth's atmosphere. The energy appears as high energy photon. The photon interacts with the electric field of an atomic nucleus and is completely absorbed, resulting in the production of an electron-positron pair. The energy required for pair production is more than 1 MeV. The electron and positron so produced have sufficient energy to produce more photons on interaction with nuclei. These photons are further capable of bringing about pair production. The result is the generation of a large number of photons, electrons and positrons having a common origin (Fig. 23.6). The multiplication will continue until the initial energy becomes divided between a large number of pairs and the individual energies of the particles fall below the "critical energy" when photon emission and pair-production can no longer occur.

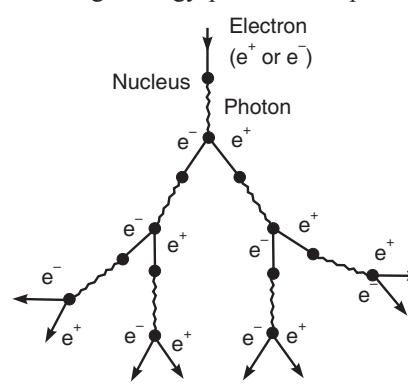


Fig. 23.6

**Pair Production.** The conversion of a *photon* into an *electron* and a *positron*, when the photon traverses the strong electric field surrounding a nucleus, is called pair production. The electric

charge is conserved, since the electron and positron have charges of equal magnitude and opposite sign. The energy of the photon provides :

- (i) the rest energy of the electron ( $m_0 c^2$ )
- (ii) the rest energy of the positron ( $m_0 c^2$ )
- (iii) the K.E. of the electron ( $E^-$ ) and
- (iv) the K.E. of the positron ( $E^+$ )

If  $h\nu$  is the energy content of the radiation, then

$$h\nu = 2m_0 c^2 + E^+ + E^-$$

Pair production is *materialization of radiant energy*. It is clear from the equation that a photon cannot create an electron-positron pair unless its energy exceeds  $2m_0 c^2$ . This gives the *threshold energy* for pair production.

**Annihilation of Matter.** The converse of the materialization of energy (production of electron-positron pair by photons) is the annihilation of matter. When a positron combines with an electron both disappear producing two quanta of  $\gamma$  rays:  $[e^+ + e^- \rightarrow 2\gamma]$ . This process is called annihilation of matter.

### 23.7 DISCOVERY OF POSITRON

The positron is the *exact counterpart of the electron*, having the same mass and having one unit of positive charge. It was first discovered by Anderson in 1932.

Anderson was making cosmic ray investigations with a cloud chamber placed in a strong magnetic field. In one of the photographs, he observed two tracks of the same curvature but in opposite directions (Fig. 23.7). The curve A was identified as due to electron. The track B was at first thought to be due to an electron moving from above downwards. But measurements of the energy of the particles on the two sides of the lead plate showed that the energy above the plate was less than that below the plate. This showed that the particle is moving from below upwards, just like the electron causing the track A. It was, therefore, deduced that the track B was caused by a particle of the same mass as the electron, but with a *positive charge*  $e$ . The particle was called a *positron*.

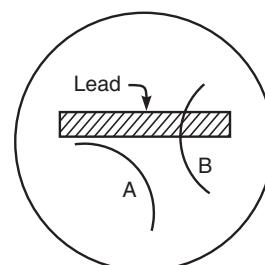


Fig. 23.7

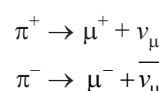
### 23.8 THE MESONS

The meson is a fundamental particle, possessing a mass intermediate between the electron and proton. The known varieties of mesons are:

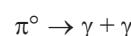
- (i)  $\pi$ -mesons or pions      (ii)  $K$ -mesons or kaons      (iii)  $\eta$ -mesons or eta.

**(i)  $\pi$ -mesons or Pions:**  $\pi$ -mesons were discovered in 1947 in the cosmic rays. They can exist in three states:  $\pi^+$ ,  $\pi^-$  and  $\pi^0$ . The  $\pi^+$  and  $\pi^-$  are antiparticles of each other.  $\pi^+$  and  $\pi^-$  mesons have a rest mass of  $273 m_e$ . The rest mass of  $\pi^0$  meson is  $264 m_e$ .

The pion is unstable with a mean life of  $2.6 \times 10^{-8}$  s. The charged pions decay into corresponding muon and a  $\mu$ -neutrino :



The neutral pion has a mean life of  $8.7 \times 10^{-17}$  s and decays into two gamma rays :



$\mu^+$  and  $\mu^-$  (positive and negative muons) have the same rest mass of  $106 \text{ MeV}/c^2$  ( $207 m_e$ ) and the same spin of  $1/2$ . Both decay with a relatively long mean life of  $2.2 \times 10^{-6} \text{ s}$  into electrons and neutrino-antineutrino pairs.

$$\begin{aligned}\mu^+ &\rightarrow e^+ + \nu_e + \bar{\nu}_\mu \\ \mu^- &\rightarrow e^- + \nu_\mu + \bar{\nu}_e\end{aligned}$$

(ii) **K-Mesons (or Kaons):** *K-meson is a heavier class of mesons.*  $K^+$  has a rest mass of  $966 m_e$  and mean life of  $1.2 \times 10^{-8} \text{ s}$ . There are two varieties of neutral K-mesons  $\rightarrow K_1^0$  and  $K_2^0$ . Both of them have rest masses of  $974 m_e$  but their mean lives are  $9 \times 10^{-11} \text{ s}$  and  $5 \times 10^{-8} \text{ s}$  respectively.

(iii)  **$\eta$ -Meson:** The neutral eta-meson ( $\eta^0$ ) was discovered in 1961. It has a rest mass of  $1073 m_e$  and a zero spin value (boson). Its mean life is  $6 \times 10^{-19} \text{ s}$ .

### 23.9 VAN ALLEN BELTS

Artificial satellites helped to discover the presence of two strong radiation belts, called Van Allen belts. These belts are formed around the earth except at magnetic poles, at heights of  $3200 \text{ km}$  and  $16000 \text{ km}$  above the equator. Each belt contains two localized zones:

1. **Inner Zone:** It consists of protons of high energy of the order of  $100 \text{ MeV}$  and electrons of low energy of the order of  $1 \text{ MeV}$ .
2. **Outer Zone:** It is less intense than the inner zone and consists of only low energy electrons, the energy being of the order  $0.1 \text{ MeV}$ .

The Van Allen belts consist of high energy protons and electrons originating from the sun and cosmic rays that have been captured by the earth's magnetic field. The existence of such a belt proved beyond doubt that high intensity particles are trapped by the earth's magnetic field. Radiation belts are present around all celestial bodies which have a magnetic field. The moon has no magnetic field of its own and no radiation belt.

### 23.10 ORIGIN OF COSMIC RAYS

Several theories have been put forward regarding the origin of *primary* cosmic rays, and the mechanism by which they acquire enormous energies.

(i) **Explosion theory.** Lemaitre, Regener and others have suggested that once upon a time the whole mass of the universe was concentrated in a single nucleus. About three billion years ago, this universe *exploded*, forming the galaxies which are still running away from each other. During the explosion, a fantastically great amount of radiation was formed. Protons and other nuclei were shot out in all directions with all energies and in sufficient numbers. The cosmic rays are simply the debris, the dust of the explosion. This view does not explain the presence of the *heavy nuclei* in primary rays. Such a tremendous explosion should have broken up the *heavy nuclei* into their constituent particles, *viz.*, protons and neutrons.

(ii) **Origin from sun.** One of the views is that the *sun may be the source of at least some of the cosmic rays*. According to it, at the times of solar activity violent eruptions occur and ionized gases shoot out from the sun. Thus some of the protons in the sun acquire high energies and thrown out into interplanetary space. The points in favour of this assumption are that the cosmic ray intensity increases during solar flares and varies slightly with the rotation of the sun. But since the cosmic ray intensity remains almost uniform at all hours of the day and night, sun can not be thought to be responsible for the majority of the primary cosmic ray particles. It may be the source of the small fraction of the low energy primaries.

**(iii) Origin from cosmic ray stars.** Another view is that the *cosmic rays come from the so called cosmic ray stars* which are more active than the sun. Our galaxy has about  $10^{11}$  stars including double stars, variable stars, novae and supernovae. All these may be the possible origins of cosmic rays.

The current view of the origin of cosmic rays is that the *sun emits low energy cosmic rays while high energy cosmic rays are emitted by cosmic ray stars within our galaxy*.

Regarding the *tremendous energies* of cosmic rays, most scientists believe that *protons acquire acceleration in interstellar magnetic fields*.

## EXERCISE

## ELEMENTARY PARTICLES

**After reading this chapter, you should be able to**

- ◆ Outline the differences between hadrons and leptons
- ◆ Identify elementary particles and the type of fundamental interaction responsible for a given elementary process
- ◆ Outline the quark composition of hadrons
- ◆ Explain why a free quark is not observed

## 24.1 INTRODUCTION

After studying the structure of atoms, one gets the impression that perhaps electron, proton and neutron are the only building blocks of matter. Recently the extensive studies made partly on high energy cosmic ray particles and even more, with the help of high energy accelerators, have revealed the existence of numerous new nuclear particles.

The subatomic or elementary particles discovered so far, form a long list (around 200). These particles are elementary in the sense that they are structureless, *i.e.*, they cannot be explained as a system of other elementary particles. Table 24.1 shows the elementary particles having relatively long life. Their classification is as follows :

**(A) Baryons (or heavy particles).** Proton and particles heavier than proton form this group. Proton and neutron are called *nucleons* and the rest are called *hyperons*. Every baryon has an *antiparticle*. If a number, called the baryon number + 1 is assigned to baryons and a number – 1 is assigned to antibaryons, then in any closed system interaction or decay, the baryon number does not change. This is the law of conservation of baryons.

**Hyperons** is a special class of baryons characterised by a time decay of the order of  $10^{-10}$  seconds and mass value intermediate between those of the neutron and deuteron. Their decay time is very much greater than the time of their formation ( $10^{-3}$  sec). It is because of this unsolved problem, that these particles, along with the *K*-mesons, are called *strange particles*. There are four types of hyperons; Lambda, Sigma, Xi and Omega.

Table 24.1. Elementary particles

	Name	Symbol	Rest Energy $M_0/\text{MeV}$	Mean lifetime $\tau/\text{s}$	Common decay modes
Leptons	Photon	$\gamma$	0	stable	
	Neutrino	$\nu_e$	0	stable	
		$\nu_\mu$	0	stable	
	Electron	$e^\pm$	0.511004(2)	stable	
	Muon	$\mu^\pm$	105.659 (2)	$2.1994 (6) \times 10^{-6}$	$e \nu \bar{\nu}$
Mesons	Pion	$\pi^\pm$	139.576 (11)	$2.602 (2) \times 10^{-8}$	$\mu\nu$
		$\pi^0$	134.972 (12)	$0.84 (10) \times 10^{-16}$	$\gamma\gamma (99\%) \gamma e^+ e^- (1\%)$
	Kaon	$K^\pm$	493.82 (11)	$1.235 (4) \times 10^{-8}$	$\mu\nu (64\%) \pi^\pm \pi^0 (21\%)$
		$K^0$	497.76 (16)	50% $K_1$ , 50% $K_2$	$3\pi (5\%)$
		$K_1$		$8.62 (6) \times 10^{-11}$	$\pi^+ \pi^- (69\%) 2\pi^0 (31\%)$
		$K_2$		$5.38 (19) \times 10^{-8}$	$\pi e \gamma (39\%) \pi \mu \nu (27\%)$
	Eta				$3\pi^0 (21\%) \pi^+ \pi^- \pi^0 (13\%)$
		$\eta^0$	548.8 (6)		$\gamma\gamma (3.8\%) \pi\gamma\gamma (2\%) 3\pi^0 (31\%) \pi^+ \pi^- \pi^0 (23\%)$
					$\pi^+ \pi^- \gamma (5\%)$

Baryons	Proton	$P^\pm$	938.256 (5)	Stable	
	Neutron	n	939.550 (5)	$9.32 (14) \times 10^2$	pev
	Lambda	$\Lambda^0$	1115.60 (8)	$2.51 (3) \times 10^{-10}$	$p\pi^2$ (65%) $n\pi^0$ (35%)
	Sigma	$\Sigma^+$	1189.4 (2)	$8.02 (7) \times 10^{-11}$	$p\pi^0$ (52%) $n\pi^+$ (48%)
		$\Sigma^0$	1192.46 (12)	$< 10^{-14}$	$\Lambda\gamma$
		$\Sigma^-$	1197.32 (11)	$1.49 \times 10^{-10}$	$n\pi^-$
	Xi	$\Xi^0$	1314.7 (7)	$3.03 (18) \times 10^{-10}$	$\Lambda\pi^0$
		$\Xi^-$	1321.25 (18)	$1.66 (4) \times 10^{-10}$	$\Lambda\pi^-$
	Omega	$\Omega^-$	1672.5 (5)	$1.3 (4) \times 10^{-10}$	$\Xi^0\pi^-$ , $\Xi^-\pi^0$ , $\Lambda K^-$ (?)

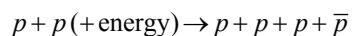
**(B) Leptons.** This group contains electron, photon, neutrino and muon.

**(C) Mesons.** The rest mass of these particles varies between about  $250m_e$  and  $1000m_e$ . The mesons are the agents of interaction between particles inside the nucleus. Baryons and mesons, are jointly called *hadrons*, and are the particles of strong interaction.

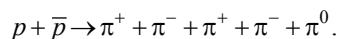
## 24.2 PARTICLES AND ANTI-PARTICLES

**Electron and Positron.** The positron and the electron are said to be antiparticles. They have the same mass and the same spin but opposite charge. They annihilate each other with the emission of photons, when they come in contact with each other. The existence of an antiparticle for the electron was actually predicted by Dirac, because of a symmetry of the equations of the relativistic quantum theory of the electron. Positron was discovered by Anderson in 1932.

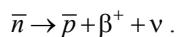
**Proton and antiproton.** Dirac's theory, anticipating the positron, could be interpreted as implying a particle identical to the proton, except for a negative charge. The existence of this particle, the *antiproton*, was established in 1955 by Segre, Chamberlain, and their collaborators. Antiprotons were produced by bombarding protons in a target with 6-GeV protons, thereby inducing the reaction



The K.E. of the bombarding proton is converted to a proton-antiproton pair plus the K.E. of the four residual particles. Antiprotons interact strongly with matter and annihilate with protons. In a typical annihilation reaction, the rest mass of the annihilating pair appears as five pions and their K.E.



**Neutron and antineutron.** The antiparticle of neutron, *antineutron*, was discovered in 1956 by Cork, Lamberton and Wenzel. The nature of the antineutron is not very well known. Both neutron and antineutron have zero charge and the same mass. However, since neutron is supposed to have a certain internal charge distribution, it is expected that the antineutron has an internal charge distribution opposite to that of the neutron. Antineutron is quickly annihilated, either by a proton or a neutron, usually with the production of several pions. If an antineutron is not annihilated by a nucleon, it decays by the reaction



**Neutrino and antineutrino.** As discussed in chapter 20, it is possible to avoid violations of the law of conservation of energy, linear and angular momenta and spin by assuming the existence of a particle, called a *neutrino*. It has *charge zero, spin 1/2, zero rest mass (rest mass very small in comparison with that of the electron) and magnetic moment smaller than  $10^{-8}$  Bohr magneton or nearly zero*. The neutrino has a finite energy and momentum in flight. It travels with the speed of light  $c$ . It does not cause ionization on passing through matter.

The antiparticle of neutrino is *antineutrino*. The distinction between the neutrino  $\nu$  and antineutrino  $\bar{\nu}$  is a particularly interesting one. The spin of the neutrino is opposite in direction to the direction of its motion; viewed from behind, the neutrino spins-counterclockwise. But the spin of the antineutrino is in the same direction as its direction of motion; viewed from behind, it spins clockwise. Thus the neutrino moves through space in the manner of a left-handed screw, while the antineutrino does so in the manner of a right-handed screw. Thus a neutrino possesses a “left-handed” helicity; the antineutrino possesses a “right-handed” helicity, *i.e.*, *a neutrino and antineutrino differ only in the sense of their helicity*.

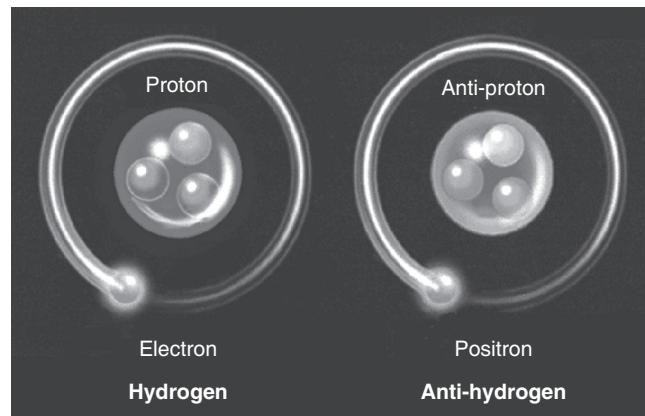
It is customary to call the particle accompanying a positron a neutrino,  $\nu$ , while that accompanying an electron is called an antineutrino,  $\bar{\nu}$ . It is important to remember that positrons and electrons are never ejected alone. It is clear from experiments that neutrinos emitted in  $\beta$  decay (positron decay and  $K$ -capture) have a negative helicity, *i.e.*, they are longitudinally polarized with their spin axes antiparallel to their direction of travel.

$$\mathbf{H} = +1 \text{ for } \bar{\nu}; \mathbf{H} = -1 \text{ for } \nu.$$

Because of its lack of charge and magnetic moment, a neutrino has essentially no interaction with matter, except that which leads to inverse  $\beta$  decay. This interaction is extremely weak. The cross section for this process is only  $\sigma \approx 10^{-48} \text{ m}^2 \approx 10^{-20} \text{ barn}$ . Matter is almost totally transparent to neutrinos.

### 24.3 ANTIMATTER

In Atomic Physics, it has long been useful to consider an atom as composed of extranuclear electrons and a nucleus formed of protons and neutrons. There is reason to believe that a *positron* and an *antiproton* could form an atom of *antihydrogen*. Antihydrogen would have a spectrum similar to that of ordinary hydrogen. Indeed, from a collection of antiprotons, antineutrons, and positrons, a world of antimatter might be constructed, which could be indistinguishable from our world so long as everything were made of antiparticles. However, if some of this *contraterrene* matter were to come in contact with the ordinary matter, particle-antiparticle annihilation would occur with a tremendous release of energy.



### 24.4 THE FUNDAMENTAL INTERACTIONS

Four kinds of interaction between elementary particles can account for all known processes in the physical universe on all scales of size. In Table 24.2 the four basic interactions are summarised.

**Table 24.2. The four fundamental interactions**

<i>Interaction</i>	<i>Particles Affected</i>	<i>Range</i>	<i>Relative Strength</i>	<i>Particles Exchanged</i>
Strong Electro-magnetic	Hadrons	$\sim 10^{-15}$ m	1	Mesons
	Charged Particles	$\infty$	$\sim 10^{-2}$	Photons
Weak	Hadrons and leptons	$\sim 10^{-17}$ m	$\sim 10^{-13}$	Intermediate bosons
Gravitational	All	$\infty$	$\sim 10^{-40}$	Gravitons

(1) **Strong interaction.** A familiar example of strong interaction is the forces which hold nucleons together (nuclear forces) in the atomic nucleus. The strong nuclear interaction is independent of the electric charge. The range of these interactions is about  $10^{-15}$  m. Time interval of such an interaction is roughly  $10^{-23}$  s.

(2) **Electromagnetic interaction.** It operates on all charged particles. Thus electromagnetic interactions are charge dependent (attractive as well as repulsive). The range is infinite and the interaction works through the *photon*. The formation of electron-positron pair from  $\gamma$ -ray is an example of electromagnetic interaction.

(3) **Weak interaction.** All strong interactions take place in times of about  $10^{-23}$  s (characteristic time). Yet it has been observed that some of the resulting particles, although energetically unstable, suffer no decay until a time  $10^{13}$  times greater than  $10^{-23}$  is reached. That is, their decay takes place in time of about  $10^{-10}$  s. For example,  $\beta$ -decay of radioactive nuclei does not take place until a time  $10^{13}$  times greater than that involved in strong interactions has approached. Had there been strong nuclear or electromagnetic interactions, there would have been no such delay in the decay process. Therefore, this delay in the decay process suggests that either these particles are not subjected to strong interacting forces or there is some new conservation law or prohibition which forbids the decay. But since most of the particles involved are subjected to either nuclear force or have electric charge or both, there must be some rule which stops the process. But eventually the decays do happen, there must be some other type of interaction as predicted by Fermi to explain  $\beta$ -decay. Since particles take long time to respond to such an interaction, force involved must be very weak compared with strong nuclear forces. The range of such an interaction is less than  $10^{-17}$  m. The characteristic time of this interaction is  $\approx 10^{-10}$  s. Yukawa, in 1938, suggested that there should be a field quantum for the weak interaction, corresponding to the photon and pion. This so called *intermediate vector boson* has not been experimentally detected as a free particle. The weak interaction is responsible for the decay of strange and non-strange particles and for non-leptonic decays of strange particles.

(4) **Gravitational interaction.** It is the weakest of the four types of interactions. It has infinite range. Although gravity has a measurable influence on macroscopic bodies, its interaction with subnuclear particles is very small. Gravitation can be explained in terms of the interactions of '*gravitons*'. Their mass must be zero, and therefore, their velocity must be that of light. As the gravitational field is extremely weak, the gravitons can not be detected in the laboratory.

Of the four basic forces, only gravitational force is universal. Weak forces affect every particle other than photon. Electromagnetic force is confined to charged particles. Strong forces are the most selective and they serve as the criteria for classifying all known particles other than photons into two broad categories, the leptons and the hadrons. Leptons and photons are light particles and do not feel the strong forces. But hadrons feel the strong forces and participate in strong interactions besides taking part in other types of interactions as well. Thus the proton is a strongly interacting nuclear

constituent and is therefore, a hadron. At the same time, because of its charge and mass, it must also experience electromagnetic and gravitational forces. The fact that it can be created by  $\beta$ -decay of neutron shows that the proton is involved in weak interactions as well.

## 24.5 ELEMENTARY-PARTICLE QUANTUM NUMBERS

In classifying the various elementary particles, several discrete quantum numbers are used. We are already familiar with two such quantum numbers, namely those that describe a particle's *charge* and *spin*. These quantum numbers specify measurable physical properties and are always conserved. We know that all elementary charges are 0 or  $\pm 1$ . The charge is conserved in all processes and no exceptions are known. The spin quantum number  $J$  is either an integer or an half odd integer for the particles so far detected. Particles with integer spin obey the Bose-Einstein statistics and are called *bosons*. Particles with half odd integer spins obey the Fermi-Dirac statistics and are called *fermions*. The other quantum numbers are more abstract and it is not always clear precisely to what aspect of physical reality they refer.

(a) **Baryon number.** Each baryon is given a baryon number  $B = 1$ , each corresponding antibaryon is given a baryon number  $B = -1$ . All other particles have  $B = 0$ . The law of conservation of baryons states that the sum of the baryon numbers of all the particles after a reaction or decay must be the same as their sum before. This rule ensures that a proton cannot change into an electron, even though a neutron can change into a proton. Baryon conservation ensures the stability of the proton against decaying into a particle of smaller mass.

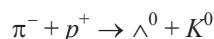
(b) **Lepton number.** Leptons are supposed to possess a property called *Lepton number* ( $L$ ). Since the neutrinos associated with electrons and with muons are recognised as different, we introduce two lepton numbers  $L_e$  and  $L_\mu$ , both of which must be conserved separately in particle reactions and decays. The number  $L_e = 1$  is assigned to the electron and the e-neutrino and  $L_e = -1$  to their antiparticles. All other particles have  $L_e = 0$ . Also, the number  $L_\mu = 1$  is assigned to the muon and the  $\mu$ -neutrino and  $L_\mu = -1$  to their antiparticles. All other particles have  $L_\mu = 0$ . An example of particle-number conservation is the decay of the neutron, in which  $B = 1$  and  $L_e = 0$  before and after :

$$\begin{aligned} n &\rightarrow p + e^- + \bar{\nu}_e \\ B : +1 &\rightarrow (+1) + 0 + 0 \\ L_e : 0 &\rightarrow 0 + (+1) + (-1) \end{aligned}$$

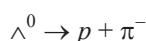
This is the only way in which the neutron can decay and still conserve both energy and baryon number  $B$ . Similarly, pion decay illustrates the conservation of  $L_\mu$  :

$$\begin{aligned} \pi^- &\rightarrow \mu^- + \bar{\nu}_\mu \\ L_\mu : 0 &\rightarrow (+1) + (-1) \end{aligned}$$

(c) **Strangeness number.** The properties of  $K$ -mesons and hyperons appeared at first instance to be paradoxical in that they were produced by high energy reactions but always in pairs *i.e.*, if one particle of some kind is produced then simultaneously another different particle is also emitted. This is termed as the *associated production* of kaons and hyperons which are known as *strange particles*. Furthermore as they are produced in strong interaction, one should expect them to decay also through strong interactions in a very short time but this is not observed. Instead they decay slowly. Because of this strange behaviour they were called as strange particles. Their production is typical of strong interaction such as :



and their decay



$$K^0 \rightarrow \pi^+ + \pi^-$$

is through weak interaction.

This and other considerations led to the notion of *strangeness number*  $S$ .  $S = 0$  for the nucleons, and  $S \neq 0$  for kaons and hyperons (Table 24.3). It is found that  $S$  is conserved in all processes mediated by the strong and electromagnetic interactions. Kaons and hyperons, which have  $S \neq 0$ , are created in high-energy collisions that involve the strong interaction. Their multiple appearance results from the necessity to conserve  $S$ . An example is the proton-proton collision :

$$\begin{aligned} p^+ + p^+ &\rightarrow \Lambda^0 + K^0 + p^+ + \pi^+ \\ S : 0 + 0 &\rightarrow (-1) + 1 + 0 + 0 \end{aligned}$$

On the other hand,  $S$  can change in an event governed by the weak interaction. The decays of kaons and hyperons proceed via the weak interaction and accordingly are extremely slow. Even the weak interaction, however, is unable to permit  $S$  to change by more than  $\pm 1$  in a decay.

(d) **Hypercharge.** A quantity called *hypercharge* ( $Y$ ) is conserved in strong interaction. Hypercharge is equal to the sum of the strangeness and baryon numbers of the particle families :  $Y = S + B$ . The various hypercharge assignments are listed in Table 24.3.

Table 24.3.

Class	Name	Symbol	Spin	$B$	$Le$	$L_\mu$	$S$	$Y$	$I$
LEPTON	$e$ -neutrino	$\nu_e$	$\frac{1}{2}$	0	+ 1	0			
	$\mu$ -neutrino	$\nu_\mu$	$\frac{1}{2}$	0	0	+ 1			
	Electron	$e^-$	$\frac{1}{2}$	0	+ 1	0			
	Muon	$\mu^-$	$\frac{1}{2}$	0	0	+ 1			
MESON	Pion	$\pi^\pm$ $\pi^0$	0	0	0	0	0	0	1
	Kaon	$K^+$ $K^0$	0	0	0	0	+ 1	+ 1	$\frac{1}{2}$
	$\eta$ meson	$\eta^0$	0	0	0	0	0	0	0
BARYON	Nucleon (Proton)	$p$	$\frac{1}{2}$	+ 1	0	0	0	+ 1	$\frac{1}{2}$
	(Neutron)	$n$							
	$\Lambda$ hyperon	$\Lambda^0$	$\frac{1}{2}$	+ 1	0	0	- 1	0	0
	$\Sigma$ hyperon	$\Sigma^+$ $\Sigma^0$ $\Sigma^-$	$\frac{1}{2}$	+ 1	0	0	- 1	0	1
	$\Xi$ hyperon	$\Xi^0$ $\Xi^-$	$\frac{1}{2}$	+ 1	0	0	- 2	- 1	$\frac{1}{2}$
	$\Omega$ hyperon	$\Omega^-$	$\frac{3}{2}$	+ 1	0	0	- 3	- 2	0

(e) **Isospin and Isospin quantum number.** As far as strong interactions are concerned, the neutron and the proton are two states of equal mass of a nucleon doublet. From Table 24.1, it is found that particles occur in multiplets. For example, singlets  $\eta^0, \Omega^-, \Lambda^0$ . Doublet :  $p, n$ , triplet  $\pi^0, \pi^-, \pi^+$ . It is natural to think of the members of a multiplet as representing different charge states of a single fundamental entity. It has proved useful to categorize each multiplet according to the number of charge states it exhibits by a number  $I$  such that the multiplicity of the state is given by  $2I + 1$ . Thus the nucleon multiplet is assigned  $I = 1/2$ , and its  $2 \times 1/2 + 1 = 2$  states are the neutron and the proton. The pion multiplet has  $I = 1$ , and its  $2 \times 1 + 1 = 3$  states are  $\pi^+, \pi^0$  and  $\pi^-$ .

Isospin can be represented by a vector  $I$  in “isospin space” whose component in any specified direction is governed by a quantum number customarily denoted  $I_3$ . The possible values of  $I_3$  are restricted to  $I, I - 1, I - 2, \dots, 0, \dots, -(I - 1), -I$ . Hence  $I_3$  is half-integral, if  $I$  is half-integral and integral or zero if  $I$  is integral. For the nucleon,  $I = 1/2$  and  $I_3 = 1/2$  corresponds to the proton,  $I_3 = -1/2$  for the neutron. Similarly, for the pion triplet  $I = 1$  giving  $I_3 = +1$  for  $\pi^+$ , 0 for  $\pi^0$  and  $-1$  for  $\pi^-$ . The charge of a meson or baryon is related to its baryon number  $B$ , its strangeness number  $S$ , and the component  $I_3$  of its isospin by the formula

$$q = e\left(I_3 + \frac{B+S}{2}\right) = e\left(I_3 + \frac{Y}{2}\right)$$

## 24.6 CONSERVATION LAWS AND SYMMETRY

A very important set of conservation laws is related to symmetries involving parity ( $P$ ), charge conjugation ( $C$ ), and time reversal ( $T$ ).

**Conservation of parity.** Parity relates to the symmetry of the wave function that represents the system. If the wave function is unchanged, when the coordinates  $(x, y, z)$  are replaced by  $(-x, -y, -z)$ , then the system has a parity of +1. If the wave function has its sign changed, when the coordinates are reversed, then the system has a parity of -1. If we write  $\psi(x, y, z) = P\psi(-x, -y, -z)$ , we can regard  $P$  as a quantum number characterizing  $\psi$  whose possible values are +1 and -1. During a reaction in which parity is conserved, the total parity number does not change.

Changing the coordinates  $(x, y, z)$  into  $(-x, -y, -z)$  converts a right-handed coordinate system into a left-handed coordinate system. In terms of symmetry, the meaning of conservation of parity is that in any situation where parity is conserved, the description of the reaction will not be changed if the word “left” is changed to the word “right” and vice versa. This means that such reactions can provide no clue that will distinguish between the directions right and left. Prior to 1956 it was believed that all reactions in nature obeyed the law of conservation of parity. However, Yang and Lee pointed out that in reactions involving the weak interaction, parity was not conserved, and that experiments could be devised that would absolutely distinguish between right and left. Indeed parity conservation is found to hold true only in the strong and electromagnetic interactions.

**Charge conjugation symmetry.** Charge conjugation is the act of symmetry operation in which every particle in a system is replaced by its antiparticle. If the anti-system, or antimatter counterpart exhibits the same physical phenomena, then charge parity ( $C$ ) is conserved. For example, if in a hydrogen atom, the proton is replaced by an antiproton and the electron is replaced by a positron, then this antimatter atom will behave exactly like an ordinary atom, if observed by people also made of antimatter. In fact  $C$  is not conserved in the weak interaction.

**Time reversal symmetry.** Time parity  $T$  describes the behaviour of a wave function when  $t$  is replaced by  $-t$ . The symmetry operation that corresponds to the conservation of time parity is time reversal. Time reversal symmetry implies that the direction of time is not significant, so that the reverse of any process that can occur is also a process that can occur. In other words, if symmetry under time reversal holds, it is impossible to establish by viewing it whether a motion picture of an

event is being run forward or backward. Prior to 1964, time parity  $T$  was considered to be conserved in every interaction. It was discovered in 1964 that one form of the  $K^0$  kaon can decay into  $\pi^+ + \pi^-$ , which violates the conservation of  $T$ . The symmetry of phenomena under time reversal thus does not seem to be universal.

**Combined inversion of CPT.** The combined symmetry operation in which the antimatter mirror-image of a system is run in reverse allows a test of *CPT* invariance. All the evidence supports the conservation of *CPT*. The conservation of *CPT* means that for every process there is an antimatter mirror-image counterpart that takes place in reverse. This particular symmetry seems to hold for all interactions, even though its component symmetries sometimes fail individually.

## 24.7 THE QUARK MODEL

Murray Gell-Mann and G. Zweig proposed the quark model in 1964. This theory is based on the idea that the hadrons are built up from a limited number of “fundamental” units, which have acquired the name quarks. The original three quarks were labeled  $u$  (for “up”),  $d$  (for “down”) and  $s$  (for “strange”).

$u$  quark has electric charge  $+\frac{2}{3}e$  and strangeness 0.

$d$  quark has electric charge  $-\frac{1}{3}e$  and strangeness 0.

$s$  quark has electric charge  $-\frac{1}{3}e$  and strangeness -1.

Each quark has a baryon number of  $B = 1/3$ .

Each quark has an antiquark associated with it ( $\bar{u}$ ,  $\bar{d}$ , and  $\bar{s}$ ). The magnitude of each of the quantum numbers for the antiquarks has the same magnitude as those for the quarks, but the sign is changed.

### Compositions of hadrons according to the quark model

Hadrons may be *baryons* or *mesons*. A baryon is made up of three quarks. For example, the proton is made up of two  $u$  quarks and a  $d$  quark ( $uud$ ). For these quarks, the electric charges are  $+2/3$ ,  $+2/3$ , and  $-1/3$ , for a total value of  $+1$ . The baryon numbers are  $+1/3$ ,  $+1/3$  and  $+1/3$ , for a total of  $+1$ . The strangeness numbers are 0, 0 and 0 for a total strangeness of 0. All are in agreement with the quantum numbers for the proton. Fig. 24.1 shows quark models of the proton, antiproton, neutron and antineutron. Electric charges are given in units of  $e$ .

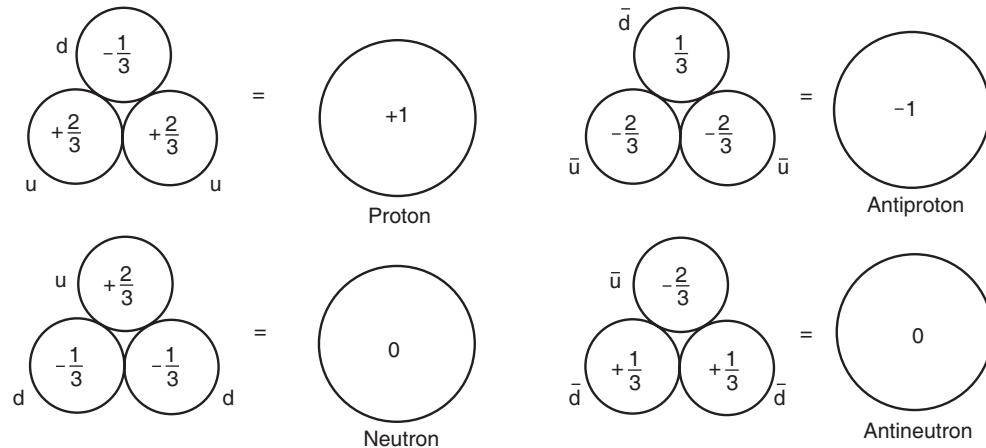


Fig. 24.1

A meson is made up of one quark and one antiquark. For example, the  $\pi^+$  **meson** is the combination of a  $u$  quark and a  $d$  antiquark ( $u\bar{d}$ ). Electric charges of these quarks are  $+2/3$  and  $+1/3$  for a total of  $+1$ . The baryon numbers are  $+1/3$  and  $-1/3$  for a total baryon number of  $0$ . The strangeness numbers are  $0$  and  $0$  for a total of  $0$ . All of these are in agreement with the quantum numbers for the pi-meson. Quarks all have spins of  $1/2$ , which accounts for the observed half-integral spins of baryons and the  $0$  or  $1$  spins of mesons.

All known hadrons can be explained in terms of the various quarks and their antiquarks. Table 24.4 shows the quark contents of five hadrons and how they account for the observed charges, spins, and strangeness numbers of these particles.

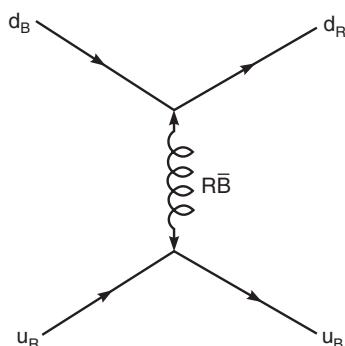
**Table 24.4. Compositions of some hadrons according to the Quark Model**

Hadron	Quark content	Baryon number	Charge, $e$	Spin	Strangeness
$\pi^+$	$u\bar{d}$	$\frac{1}{3} - \frac{1}{3} = 0$	$+\frac{2}{3} + \frac{1}{3} = +1$	$\uparrow\downarrow = 0$	$0 + 0 = 0$
$K^+$	$u\bar{s}$	$\frac{1}{3} - \frac{1}{3} = 0$	$+\frac{2}{3} + \frac{1}{3} = +1$	$\uparrow\downarrow = 0$	$0 + 1 = +1$
$P^+$	$uud$	$\frac{1}{3} + \frac{1}{3} + \frac{1}{3} = +1$	$+\frac{2}{3} + \frac{2}{3} - \frac{1}{3} = +1$	$\uparrow\uparrow\downarrow = \frac{1}{2}$	$0 + 0 + 0 = 0$
$n^0$	$ddu$	$\frac{1}{3} + \frac{1}{3} + \frac{1}{3} = +1$	$-\frac{1}{3} - \frac{1}{3} + \frac{2}{3} = 0$	$\downarrow\uparrow\uparrow = \frac{1}{2}$	$0 + 0 + 0 = 0$
$\Omega^-$	$sss$	$\frac{1}{3} + \frac{1}{3} + \frac{1}{3} = +1$	$-\frac{1}{3} - \frac{1}{3} - \frac{1}{3} = -1$	$\uparrow\uparrow\uparrow = \frac{3}{2}$	$-1 - 1 - 1 = -3$

**Coloured quarks and gluons :** There were problems with the quark model, one of them being  $\Omega^-$  hyperon. It was believed to contain three identical  $s$  quarks ( $sss$ ). This violates the Pauli exclusion principle, that prohibits two or more fermions from occupying identical quantum states. The proton, neutron, and others with two identical quarks would violate this principle also. We can resolve this difficulty by assigning a new property to the quarks. We can regard this new property as an additional quantum number that can be used to label the three otherwise identical quarks in the  $\Omega^-$ . If this additional quantum number can take any one of three possible values, we can restore the Pauli principle by giving each quark a different value of this new quantum number, which is known as *colour*. The three colours are labeled red ( $R$ ) blue ( $B$ ) and green ( $G$ ). The  $\Omega^-$  for example, would then  $s_R s_B s_G$ . The antiquark colours are antired ( $\bar{R}$ ) antiblue ( $\bar{B}$ ) and antigreen ( $\bar{G}$ ).

An essential component of the quark model with colour is that *all observed meson and baryon states are “colourless”*, i.e., either colour anticolour combinations in the case of mesons, or equal mixtures of  $R$ ,  $B$  and  $G$  in the case of baryons.

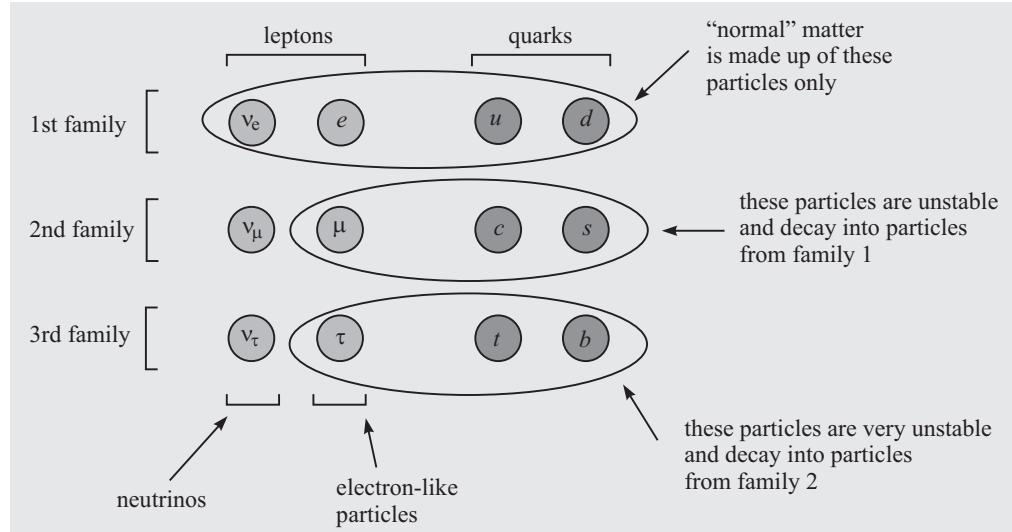
Since hadrons seem to be composed of quarks, the strong interaction between hadrons should ultimately be traceable to an interaction between quarks. The force between quarks can be modeled as an exchange force, mediated by the exchange of massless spin – 1 particles called gluons. Eight gluons have been postulated. The field that binds the quarks is a *colourfield*. Colour is to the strong interaction between quarks as electric charge is to the electromagnetic interaction between electrons. It is the fundamental strong “charge” and is carried by the gluons. The gluons must therefore be represented as combinations of a colour and a possibly different anticolour. The gluons are massless and carry their colour-anticolour properties just as other particles may



**Fig. 24.2**

carry electric charge. For example, Fig. 24.2 shows a gluon  $R\bar{B}$  being exchanged by red and blue quarks. In effect the red quark emits its redness into a gluon and acquires blueness by also emitting antiblueness. The blue quark, on the other hand, absorbs the  $R\bar{B}$  gluon, canceling its blueness and acquiring a red colour in the process.

**Charm, Bottom, and Top.** In 1970, Glashow, Iliopoulos, and Maiani proposed the existence of a fourth quark, called  $c$  or *charmed* quark. The charmed quark was suggested to explain the suppression of certain decay processes that are not observed. With only three quarks, the processes would proceed at measurable rates and should have been observed. The charm quark has a charge of  $\frac{2}{3}e$ , strangeness 0 and a charm quantum number of +1. Other quarks have 0 charm.



In 1977, a new particle was discovered at Fermi Lab that provided evidence for yet another quark. This particle, called the upsilon-meson, was thought to be made up of the new quark called  $b$  (for bottom or beauty) along with the associated antiquark  $\bar{b}$ .  $b$  quark has electric charge  $-1/3e$ .

Because quarks seem to come in pairs, it is expected that there is a partner to the  $b$  quark, called  $t$  (for top, if  $b$  = bottom, or truth, if  $b$  = beauty). It has a charge of  $+2/3e$ .

**Three generations of quarks and leptons.** Both leptons and quarks appear to come in three generations of doublets, with all particles having spin 1/2. Table 24.5 shows the properties of the three generations of quarks and leptons. The first generation contains two leptons, the electron and the electron neutrino, and two quarks, up and down. All the properties of ordinary matter can be understood on the basis of these particles. The second generation includes the muon and muon-neutrino and the charm and strange quarks. These particles are responsible for most of the unstable particles and resonances created in high energy collisions. The third generation includes the tau and the tau-neutrino and the top and bottom quarks.

Generation	Lepton	Symbol	Charge, $e$
1	Electron	$e^-$	-1
	e-Neutrino	$\nu_e$	0
2	Muon	$\mu^-$	-1
	$\mu$ -Neutrino	$\nu_\mu$	0
3	Tau	$\tau^-$	-1
	$\tau$ -Neutrino	$\nu_\tau$	0

**Table 24.5. Properties of the three generations of Quarks and Leptons**

<i>Generation</i>	<i>Quark</i>	<i>Symbol</i>	<i>Charge, e</i>	<i>Strangeness</i>	<i>Charm</i>
1	Up	<i>u</i>	$+\frac{2}{3}$	0	0
	Down	<i>d</i>	$-\frac{1}{3}$	0	0
2	Charm	<i>c</i>	$+\frac{2}{3}$	0	+ 1
	Strange	<i>s</i>	$-\frac{1}{3}$	- 1	0
3	Top	<i>t</i>	$+\frac{2}{3}$	0	0
	Bottom	<i>b</i>	$-\frac{1}{3}$	0	0

**EXERCISE**

1. Hyperons are particles, which are  
 (a) heavier than proton  
 (b) of the same mass as proton  
 (c) lighter than proton  
 (d) none of the above

(H.P.U. 1999)

2. Value of electric charge and strangeness of *d* quark is

(a)  $+\frac{2}{3}e, 0$       (b)  $-\frac{2}{3}e, 0$       (c)  $-\frac{1}{3}e, 0$

(d)  $-\frac{1}{3}e, -1$   
(B.U. 2006)

3. Charge of the ‘*u*’ quark is

(a)  $\frac{1}{3}e$

(b)  $-\frac{2}{3}e$

(c)  $-\frac{1}{3}e$

(d)  $\frac{2}{3}e$   
(B.U. 2013)

4. A baryon is made up of \_\_\_\_\_ quarks.

- (a) three quarks  
 (b) two quarks  
 (c) one quark and one antiquark  
 (d) two antiquarks

(B.U. M.Sc. 2015)

5. The combination of one ‘*u*’ quark and two ‘*d*’ quarks is called

- (a) Positron      (b) Electron      (c) Proton

- (d) Neutron

(B.U. 2011)



**P A R T - VI**

# **SOLID STATE PHYSICS**

**[Condensed Matter Physics]**

**[Material Science]**

---

**25. Crystallography**

---

**26. Lattice Vibrations and Phonons**

---

**27. Free Electron Theory and Band Theory**

---

**28. Superconductivity**

---

**29. Dielectrics and Ferroelectrics**

---

**30. Magnetism**

---

## CRYSTALLOGRAPHY

**After reading this chapter, you should be able to**

- ◆ Explain the concepts of space lattice, unit cell, Bravais lattices and symmetry elements
- ◆ List the characteristics of the unit cells of simple cubic, BCC, FCC, HCP and diamond cube structures
- ◆ Discuss different methods used in the X-ray analysis of crystal structure
- ◆ Deduce the reciprocal lattices corresponding to some simple direct lattices such as face centred and body centred lattices
- ◆ Explain the various types of bonding in crystals, with emphasis on physical concepts.
- ◆ Analyze the various types of imperfections that can occur in crystals
- ◆ Apply the principles of Crystallography in research studies

## CRYSTAL STRUCTURE

### 25.1 TYPES OF SOLIDS

- Solids are of two types: crystalline and amorphous.

#### What is a Crystal?

A crystal is a solid composed of a periodic array of atoms, i.e., a representative unit is repeated at regular intervals along any and all directions in the crystal.

- A substance is said to be crystalline when the arrangement of the units of matter inside it is regular and periodic; the units of matter may be ions, atoms or molecules.
- Crystals are usually bounded by plane faces.
- Though two crystals of the same substance may look different in external appearance, the angles between corresponding faces are always the same.
- Crystals can be distinguished from amorphous substances by their *anisotropy*, i.e., the dependence of quantities like refractive index on the direction in which they are measured in the crystal.
- A crystal is the regular polyhedral form, bound by smooth surfaces, which is assumed by a chemical compound under the action of its interatomic forces when passing from the state of a liquid or gas to that of a solid under suitable conditions.
- A crystal is a homogeneous, anisotropic body having natural shape of a polyhedron.
- The study of the geometrical form and other physical properties of crystalline solids by using X-rays, electron beams and neutron beams, etc., constitute the science of crystallography.

**Amorphous solids.** In amorphous solids, there is no order in the arrangement of their constituent particles.

#### 25.1.1. Distinction between Crystalline and Amorphous Solids

Sl. No.	Crystalline solids	Amorphous solids
1.	Crystalline solids have a regular arrangement of particles.	Amorphous solids have a completely random arrangement of particles.
2.	The presence of long-range order is the defining property of a crystal.	Amorphous solids exhibit short-range order in their structures.
3.	Crystalline solids have different physical properties (thermal conductivity, electrical conductivity, refractive index, etc.) in different directions. Their properties vary with direction. Crystals are therefore anisotropic.	Amorphous solids are homogeneous and isotropic because there is no long-range order or periodicity in their internal atomic arrangement.
4.	The cooling curve for crystalline solids has breaks. The middle of the curve corresponds to the process of crystallization.	The cooling curve for amorphous solids is smooth.

Sl. No.	Crystalline solids	Amorphous solids
5.	All the bonds in crystalline solids are equally strong due to their symmetry. When a crystalline solid is heated, all the bonds break simultaneously at a unique temperature. So, the melting point is very sharp.	All the bonds in amorphous solid are not equally strong. Therefore when it is heated, the weakest bonds break first while the stronger ones later at a higher temperature. So, the amorphous solid does not have a sharp melting point.
6.	<b>Examples:</b> Sodium chloride, NaCl, Diamond, C Zinc-blende, ZnS	<b>Examples:</b> Glass, pitch, and many plastics.

### 25.1.2. Different Features of the Crystals

(1) **Faces.** The crystals are bound by a number of perfectly flat surfaces. These surfaces are called faces. In some crystals, all the faces are alike. In some crystals, all the faces are not alike *i.e.*, they have unlike faces. So, the faces may be like or unlike.

The examples of crystals having like faces are: fluorspar, which is generally obtained in cubes, alum, which is generally obtained in regular octahedron and galena which exhibits a combination of the cube and an octahedron. (Fig. 25.1)

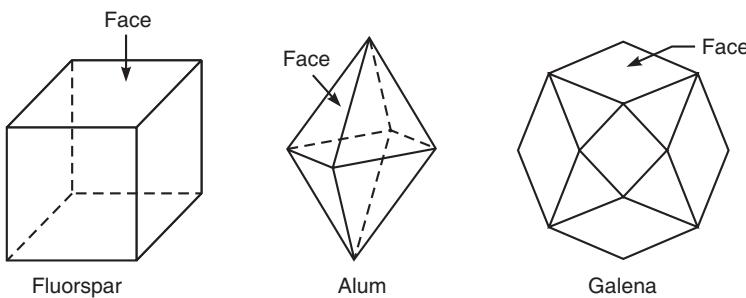


Fig. 25.1

(2) **Form.** All the faces corresponding to a crystal are said to constitute a form. The crystal that consists of all like faces is termed to have a *simple form*, while the crystal having two or more simple forms is termed to have a *combination form*.

(3) **Edges and interfacial angles.** Observations show that crystals are bounded by optically plane faces, sharp straight edges and interfacial angles. A relationship among these elements can be expressed by the formula

$$f + c = e + 2 \quad \dots(1)$$

Here  $f$  is the number of faces,  $c$  is the number of angles and  $e$  is the number of edges. Out of the three elements in Eq. (1), if two are known then the third can be easily obtained.

**EXAMPLE.** Determine the number of edges in a quartz crystal, if there are 18 faces and 14 angles in it.

**SOL.** Given  $f = 18$ ,  $c = 14$  (also refer Fig. 25.2).

Using Eq. (1), number of edges

$$e = 18 + 14 - 2 = 30.$$

This can be verified from Fig. 25.2.

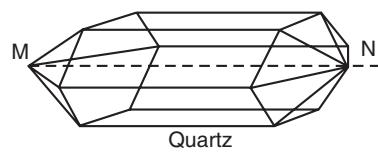


Fig. 25.2. External forms of the crystal found in nature

Steno (1669) gave a law about the interfacial angles which is as follows: "under the same physical conditions, (same temperature, same chemical composition), the angle between the corresponding faces on various crystals of the same substance are constant."

There is also a law about the constancy of symmetry which is as follows:

"All crystals of one and of the same substance have the same symmetry."

**Periodic Arrays of Atoms.** An ideal crystal is constructed by the infinite regular repetition in space of identical structural units in the form of parallelepipeds. In the simplest crystals such as copper, silver, gold and the alkali metals the structural unit contains a single atom. More generally the structural unit contains several atoms or molecules. A crystal may be composed of more than one chemical element (as in  $NaCl$ ) or it may contain associated groups of identical atoms (as in  $H_2$ ). We describe the structure of all crystals in terms of a single periodic *lattice*, but with a group of atoms attached to each lattice point. This group of atoms is called the *basis*; the basis is repeated in space to form the crystal.

**Crystal Translation vectors and Lattices.** An ideal crystal is composed of atoms arranged on a lattice defined by three *fundamental translation vectors*  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  such that the atomic arrangement looks the same in every respect when viewed from any point  $\mathbf{r}$  as when viewed from the point

$$\mathbf{r}' = \mathbf{r} + n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}, \quad \dots(1)$$

where  $n_1, n_2, n_3$  are arbitrary integers. The set of points  $\mathbf{r}'$  specified by (1) for all values of the integers  $n_1, n_2, n_3$  defines a *lattice*. The lattice and the translation vectors  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  are said to be *primitive* if any two points  $\mathbf{r}, \mathbf{r}'$  from which the atomic arrangement looks the same always satisfy (1) with a suitable choice of the integers  $n_1, n_2, n_3$ .

We often use the primitive translation vectors to define the *crystal axes*  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  although nonprimitive crystal axes may be used when they are more convenient or simpler. The crystal axes  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  form three adjacent edges of a parallelepiped. If there are lattice points only at the corners of the parallelepiped, then it is a *primitive parallelepiped*.

A *lattice translation operation* or *crystal translation operation* is defined as the displacement of a crystal parallel to itself by a crystal translation vector

$$\mathbf{T} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c} \quad \dots(2)$$

A vector  $\mathbf{T}$  connects any two lattice points.

**Illustration.** Fig. 25.3 gives the diagram of a two dimensional lattice with lattice vectors  $\mathbf{a}$  and  $\mathbf{b}$ . The atomic arrangement in the crystal looks exactly the same to an observer at  $\mathbf{r}'$  as to an observer at  $\mathbf{r}$ , provided that the vector  $\mathbf{T}$  which connects  $\mathbf{r}'$  and  $\mathbf{r}$  may be expressed as an integral multiple of the vectors  $\mathbf{a}$  and  $\mathbf{b}$ . In this illustration  $\mathbf{T} = -\mathbf{a} + 3\mathbf{b}$ . The vectors  $\mathbf{a}$  and  $\mathbf{b}$  are primitive translation vectors of the two-dimensional lattice.

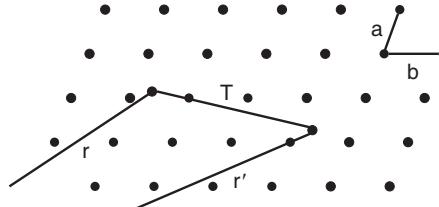


Fig. 25.3.

**The crystal lattice.** A crystal is constructed by the infinite repetition in space of identical structural units (atoms, molecules or ions). One can replace each unit by a geometrical point. The result is a pattern of points having the same geometrical properties as the crystal. This geometrical pattern is the *crystal lattice* or simply the *lattice*. The points are called lattice points.

**Definition.** The regular pattern of points which describe the three dimensional arrangement of particles (atoms, molecules or ions) in a crystal structure is called the *crystal lattice* or *space lattice*.

Fig. 25.4 (a) represents an array of points in two dimensions. Note that environments about any two points is the same. Hence, Fig. 25.4 (a) represents a lattice.

**Basis.** A crystal structure is formed by associating with every lattice point a unit assembly of atoms or molecules identical in composition. This unit assembly is called the basis. Fig. 25.4 (b) shows a particular arrangement of two different ions which represents the basis. The number of atoms in the basis may be as low as one, as for many metals but there are structures for which the basis exceeds 1000 atoms.

**The crystal structure.** A crystal structure is formed by the addition of a basis to every lattice point i.e.,  $Lattice + Basis = Crystal\ structure$ . The crystal structure is real whereas the lattice is imaginary. Fig. 25.4 clearly illustrates the difference between the crystal lattice, the basis, and the crystal structure. The crystal structure is formed by the addition of the basis (b) to every lattice point of the lattice (a). By looking at (c) we can recognise the basis and then we can abstract the space lattice.

**Unit cell.** The atoms in a crystal are arranged in a periodic array. It is therefore possible to isolate a representative *unit cell* in each variety of crystal. This unit cell may be a group of ions, atoms or molecules. Now, we can construct the crystal by repeatedly translating the unit cell in three dimensions.

**Definition.** *The smallest portion of a space lattice which can generate the complete crystal by repeating its own dimensions in various directions is called unit cell.*

A unit cell (Fig. 25.5) is defined by the length of its edges and by the angles between them.  $OA = a$ ,  $OB = b$  and  $OC = c$  are the dimensions of the unit cell. The angles between  $a$ ,  $b$ ,  $c$  and  $c$ ,  $a$  are denoted by  $\gamma$ ,  $\alpha$  and  $\beta$  respectively. They are called interfacial angles.

The vectors  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  define the axes of the crystal.

**Primitive Lattice Cell.** The parallelepiped (Fig. 25.6) defined by primitive axes  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  is called a *primitive cell*. A primitive cell is a type of *unit cell*. A unit cell will fill all space under the action of suitable crystal translation operations; a primitive cell is a minimum-volume unit cell. There is a density of one lattice point per primitive cell. There are lattice points at the eight corners of the parallelepiped, but each corner point is shared among the eight cells which touch there. The volume  $V_c$  of a primitive cell defined by primitive axes  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  is

$$V_c = |(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}|$$

Another way of choosing a cell of equal volume  $V_c$  is shown in Fig. 25.7.

(i) Draw lines to connect a given lattice point to all nearby lattice points.

(ii) At the midpoint and normal to these lines, draw new lines or planes. The smallest volume enclosed in this way is the *Wigner-Seitz primitive cell*. The basis associated with a lattice point of a primitive cell may be called a primitive basis. No basis contains fewer atoms than a primitive basis contains.

Sometimes, it is more convenient to choose a non-primitive unit cell depending upon the symmetry of the lattice. These cells contain more than one lattice points per unit cell (Fig. 25.8).

•	•	•	•
•	•	•	•
•	•	•	•
•	•	•	•

a) Space lattice  
0°

b) Basis containing  
two different ions  
0° 0° 0° 0°  
0° 0° 0° 0°

c) Crystal structure

Fig. 25.4.

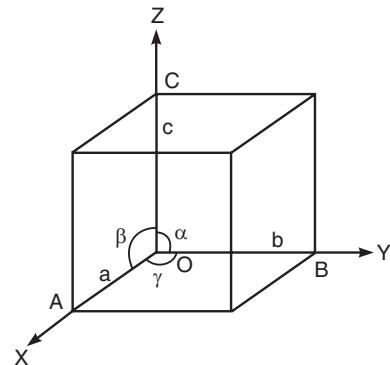


Fig. 25.5.

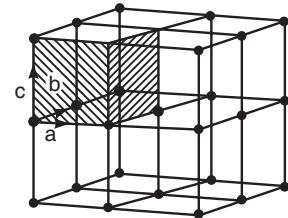


Fig. 25.6.

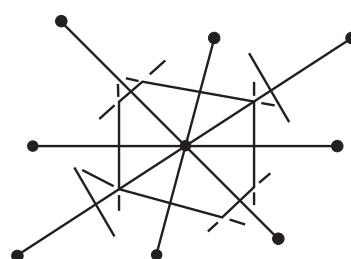


Fig. 25.7.

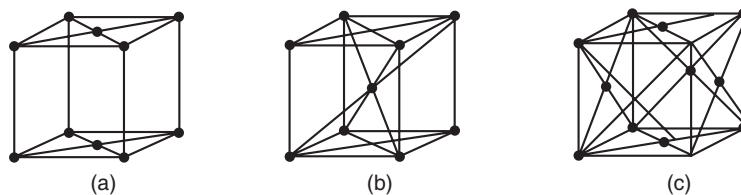


Fig. 25.8. Some non-primitive cells

## 25.2 BRAVAIS LATTICES

Bravais showed that only 14 types of space lattice are possible. Fig. 25.9 shows the 14 *Bravais lattices*. From the figure it is clear that seven sets of axes are sufficient to construct the 14 Bravais lattices. This leads to the classification of all crystals into seven crystal systems:

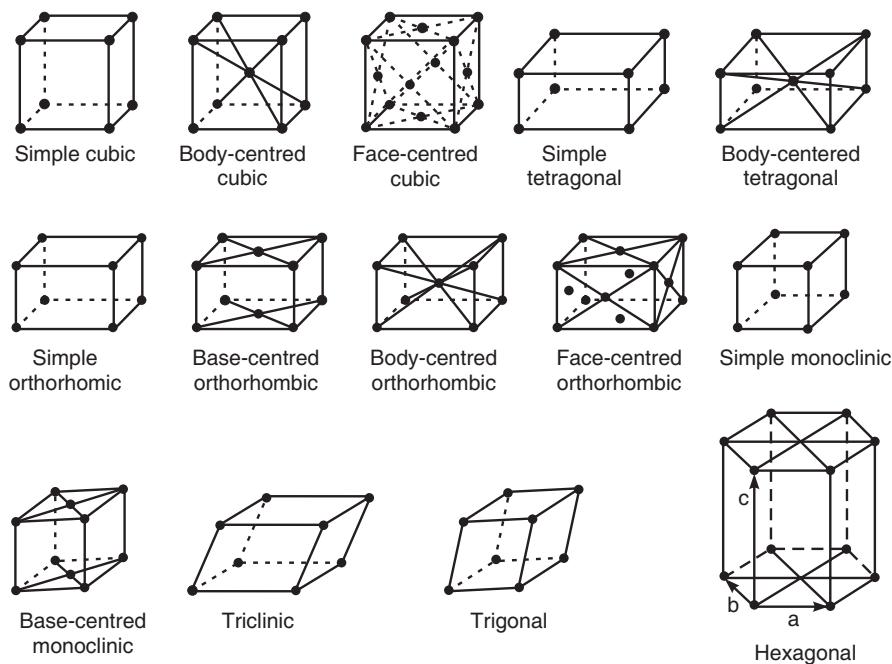


Fig. 25.9

(1) Cubic (2) Tetragonal (3) Ortho-rhombic (4) Monoclinic (5) Triclinic (6) Trigonal (sometimes called rhombohedral) and (7) hexagonal.

Table 25.1 gives the seven crystal systems, divided into fourteen Bravais lattices. Note that a simple lattice has points only at the corners, a body-centred lattice has one additional point at the centre of the cell, and a face-centred lattice has six additional points, one on each face.

**Table 25.1**  
The seven crystal systems, divided into fourteen Bravais lattices

System	Bravais lattice	Unit cell characteristics	Characteristic symmetry elements
Cubic	Simple Body-centred Face-centred	$a = b = c$ $\alpha = \beta = \gamma = 90^\circ$	Four 3-fold rotation axes (along cube diagonal)

Tetragonal	Simple Body-centred	$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	One 4-fold rotation axis
Orthorhombic	Simple	$a \neq b \neq c$	Three mutually orthogonal 2-fold rotation axes
	Base-centred	$\alpha = \beta = \gamma = 90^\circ$	
	Body-centred		
	Face-centred		
Monoclinic	Simple Base-centred	$a \neq b \neq c$ $\alpha = \beta = 90^\circ \neq \gamma$	One 2-fold rotation axis
Triclinic	Simple	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma \neq 90^\circ$	None
Trigonal (rhombohedral)	Simple	$a = b = c$ $\alpha = \beta = \gamma \neq 90^\circ$	One 3-fold rotation axis
Hexagonal	Simple	$a = b \neq c$ $\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$	One 3-fold rotation axis

There is an interesting point about cubic lattices. This is called the coordination number. *It is defined as the number of particles ‘immediately’ adjacent to each particle in the crystal lattice.* It is characteristic of a given space lattice and is determined by an inspection of the model. In a simple cubic lattice, each particle is adjoined by six other particles and so the coordination number is six. The coordination number of body-centred and face-centred cubic lattice are 8 and 12 respectively.

#### Definition of “Bravais lattice”.

We give two equivalent definitions of a Bravais lattice:

(a) A Bravais lattice is an infinite array of discrete points with an arrangement and orientation that appears *exactly* the same, from whichever of the points the array is viewed.

(b) A (three-dimensional) Bravais lattice consists of all points with position vectors  $\mathbf{R}$  of the form

$$\mathbf{R} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3, \quad \dots(1)$$

Here,  $\mathbf{a}_1$ ,  $\mathbf{a}_2$ , and  $\mathbf{a}_3$  are any three vectors not all in the same plane.  $n_1$ ,  $n_2$ , and  $n_3$  range through all integral values.

Thus the point  $\sum n_i \mathbf{a}_i$  is reached by moving  $n_i$  steps of length  $\mathbf{a}_i$  in the direction of  $\mathbf{a}_i$  for  $i = 1, 2$  and 3.

The vectors  $\mathbf{a}_i$  are called *primitive vectors* and are said to *generate* or *span* the lattice.

**Example.** Fig. 25.9(a) shows a simple cubic three-dimensional Bravais lattice. The three primitive vectors can be taken to be mutually perpendicular, and with a common magnitude.

#### 25.2.1. The Seven Crystal Systems

(i) **Cubic crystal.** In this crystal (Fig. 25.10), all the three lengths of the unit cell are equal and are at right angles to each other.

$$a = b = c$$

$$\alpha = \beta = \gamma = 90^\circ.$$

The examples are NaCl, CaF<sub>2</sub>.

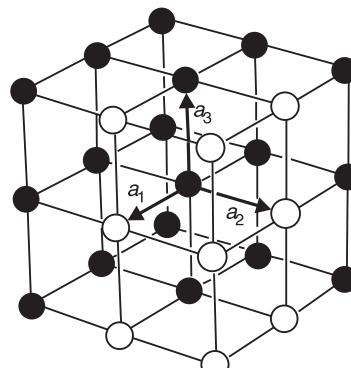
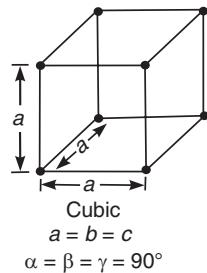


Fig. 25.9(a)

**Fig. 25.10**

**(ii) Tetragonal crystal.** In this crystal (Fig. 25.11), the two lengths of the unit cell are equal while the third one is longer.

$$a = b \neq c.$$

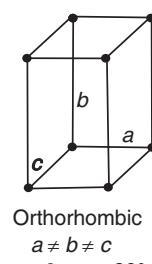
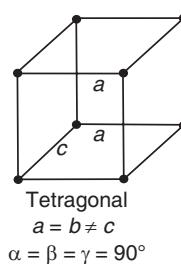
- The three axes are perpendicular,  
 $\alpha = \beta = \gamma = 90^\circ$ .
- The examples are  $\text{NiSO}_4$ ,  $\text{SnO}_2$ , etc.

**(iii) Orthorhombic crystal.** In this crystal (Fig. 25.12), the lengths of unit cell are different but three axes are perpendicular to each other.

$$a \neq b \neq c$$

$$\alpha = \beta = \gamma = 90^\circ.$$

- The examples are  $\text{KNO}_3$ ,  $\text{BaSO}_4$ , etc.

**Fig. 25.11****Fig. 25.12**

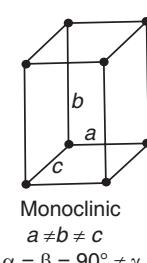
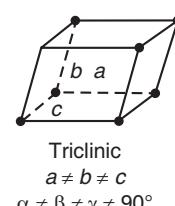
**(iv) Monoclinic crystal.** In this crystal (Fig. 25.13), the lengths of unit cell are different.

$$a \neq b \neq c.$$

- Two axes are right angles and the third is obliquely inclined.

$$\alpha = \beta = 90^\circ \neq \gamma.$$

- The examples are  $\text{Na}_2\text{SO}_4$ ,  $\text{FeSO}_4$  etc.

**Fig. 25.13****Fig. 25.14**

(v) **Triclinic crystal.** In this crystal (Fig. 25.14), the lengths of unit cell are different and oblique to each other.

$$\begin{aligned}a &\neq b \neq c \\ \alpha &\neq \beta \neq \gamma \neq 90^\circ\end{aligned}$$

The examples are  $\text{CuSO}_4$ ,  $\text{K}_2\text{Cr}_2\text{O}_7$ , etc.

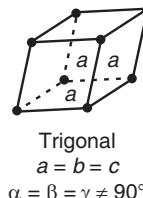
(vi) **Trigonal (Rhombohedral) crystal.** In this crystal (Fig. 25.15), all the three lengths of the unit cell are equal in length. Moreover, they are equally inclined to each other at an angle other than  $90^\circ$ .

$$a = b = c \text{ and } \alpha = \beta = \gamma \neq 90^\circ$$

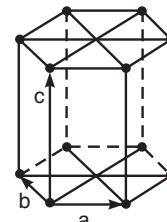
The examples are  $\text{CaSO}_4$ , Calcite, As, Sb, Bi, etc.

(vii) **Hexagonal crystal.** In this crystal (Fig. 25.16), to this plane. i.e.,  $a = b \neq c$  and  $\alpha = \beta = 90^\circ$ .  $\gamma = 120^\circ$ .

The examples are  $\text{SiO}_2$ ,  $\text{AgI}$ , Quartz, Zn, Cd, etc.



**Fig. 25.15**



Hexagonal  
 $a = b \neq c$   
 $\alpha = \beta = 90^\circ$   
 $\gamma = 120^\circ$

**Fig. 25.16**

### 25.3 MILLER INDICES

Miller introduced a system to designate a plane in a crystal. He introduced a set of three numbers to specify a plane in a crystal. This set of three numbers is known as Miller indices of the concerned plane.

**Procedure for finding Miller indices.** (1) Determine the intercepts of the plane along the axes  $X$ ,  $Y$  and  $Z$  in terms of the lattice constants  $a$ ,  $b$  and  $c$ .

(2) Determine the reciprocals of these numbers.

(3) Find the least common denominator (*lcd*) and multiply each by this *lcd*.

The result is written in the form  $(hkl)$  and is called the Miller indices of the plane.

**EXAMPLE.** Let a plane have intercepts 4, 1 and 2 on the three axes. The reciprocals are  $1/4$ ,  $1$  and  $1/2$ . *lcd* is 4. Multiplying each by 4, we get 1, 4 and 2. Hence  $(142)$  are the Miller indices of the plane.

#### Notes

- (1) For an intercept at infinity, the corresponding index is zero.
  - (2) If a plane cuts an axis on the negative side of the origin, corresponding index is negative.
  - (3) The indices  $(hkl)$  do not define a particular plane but a set of parallel planes. Thus the planes whose intercepts are  $1, 1, 1$ ;  $2, 2, 2$ ;  $-3, -3, -3$ , etc., are all represented by the same set of Miller indices.
- Miller indices of some important planes in a cubic crystal are shown in Fig. 25.17.

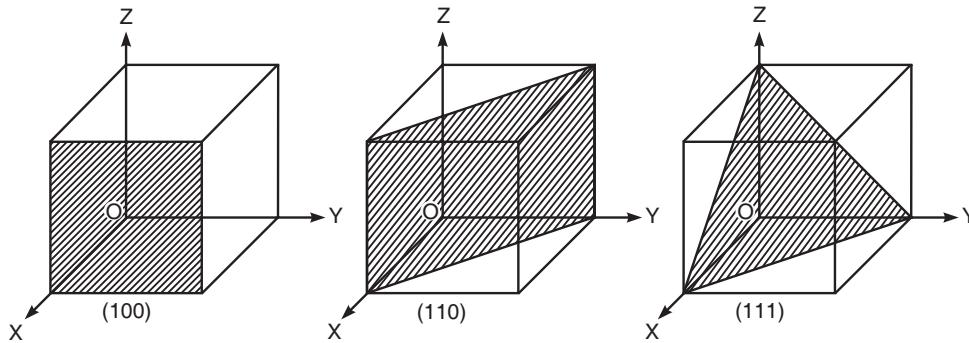


Fig. 25.17

4. It is only the ratio of the indices which is important in this notation. The (622) planes are the same as (311) planes.

**EXAMPLE.** In a crystal, a lattice plane cuts intercepts of  $2a$ ,  $3b$  and  $6c$  along the axes, where  $a$ ,  $b$  and  $c$  are primitive vectors of the unit cell. Determine the Miller indices of the given plane.

**SOL.** If the given plane cuts intercepts  $ra$ ,  $sb$ ,  $tc$  along the three axes, we have

$$ra : sb : tc = 2a : 3b : 6c$$

where  $a$ ,  $b$  and  $c$  are primitive vectors of the unit cell.

$$r : s : t = 2 : 3 : 6$$

$$\frac{1}{r} : \frac{1}{s} : \frac{1}{t} = \frac{1}{2} : \frac{1}{3} : \frac{1}{6} = 3 : 2 : 1$$

∴ Miller indices of the plane are (321).

## 25.4 SPACING BETWEEN THREE DIMENSIONAL LATTICE PLANES

Consider  $O$  as the origin and  $OX$ ,  $OY$ ,  $OZ$  as three rectangular axes (Fig. 25.18). Let a reference plane pass through  $O$ , and consider a set of parallel planes defined by Miller indices  $(hkl)$ . Beside reference plane through  $O$ , if next plane passes through  $A$ ,  $B$  and  $C$  then the respective intercepts are  $a/h$ ,  $b/k$  and  $c/l$ . If  $ON$  is the normal between this plane and the reference plane, then  $ON = d$  is the interplanar spacing.

Let us now find the value of  $d$  in terms of  $a$ ,  $b$  and  $c$ . Let the normal make angles  $\alpha$ ,  $\beta$ ,  $\gamma$  with crystal axes such that  $\angle NOX = \alpha$ ,  $\angle NOY = \beta$ , and  $\angle NOZ = \gamma$ .

We have from Fig. 25.18,

$$d = \frac{a}{h} \cos \alpha = \frac{b}{k} \cos \beta = \frac{c}{l} \cos \gamma$$

$$\text{or } \cos \alpha = \frac{d}{(a/h)}, \cos \beta = \frac{d}{(b/k)}, \cos \gamma = \frac{d}{(c/l)}.$$

$$\text{But } \cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1$$

$$\therefore \left[ \frac{d}{(a/h)} \right]^2 + \left[ \frac{d}{(b/k)} \right]^2 + \left[ \frac{d}{(c/l)} \right]^2 = 1$$

$$\text{or } d^2 \left[ \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right] = 1$$

$$\text{or } d = \left[ \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right]^{-\frac{1}{2}}$$

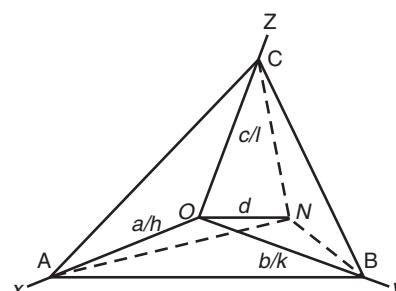


Fig. 25.18

For a simple cubic lattice,  $a = b = c$ .

Hence for such a lattice, inter-planar separation is given by

$$d_{kkl} = \frac{a}{\sqrt{(h^2 + k^2 + l^2)}}$$

$$d_{100} = \frac{a}{\sqrt{(1^2 + 0^2 + 0^2)}} = a; \quad d_{110} = \frac{a}{\sqrt{(1^2 + 1^2 + 0^2)}} = \frac{a}{\sqrt{2}}$$

$$d_{111} = \frac{a}{\sqrt{(1^2 + 1^2 + 1^2)}} = \frac{a}{\sqrt{3}}$$

Thus the separation between successive (100), (110) and (111) planes are  $a$ ,  $a/\sqrt{2}$  and  $a/\sqrt{3}$  respectively. Hence the ratio of their separation is

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

Similarly, for *fcc lattice*,

$$d_{100} = a/2; d_{110} = a/2\sqrt{2} \text{ and } d_{111} = a/\sqrt{3}$$

For *bcc lattice*

$$d_{100} = a/2; d_{110} = a/\sqrt{2} \text{ and } d_{111} = a/2\sqrt{3}.$$

**EXAMPLE 1.** Calculate the interplanar spacing for a (321) plane in a simple cubic lattice whose lattice constant is  $4.2 \times 10^{-10}\text{m}$ . (Rohilkhand 93)

**SOL.**  $a = b = c = 4.2 \times 10^{-10}\text{m}$

and

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

For the plane (321),  $h = 3$ ,  $k = 2$  and  $l = 1$

$$\therefore d_{321} = \frac{4.2 \times 10^{-10}}{\sqrt{(3^2 + 2^2 + 1^2)}} \text{ m} = 1.1 \times 10^{-10} \text{ m}$$

**EXAMPLE 2.** In a tetragonal lattice,  $a = b = 0.25 \text{ nm}$  and  $c = 0.18 \text{ nm}$ . Deduce the lattice spacings between (111) planes. (Agra 1974; Rohilkhand 81 S)

**SOL.** The lattice spacings for a given set of parallel planes ( $hkl$ ) is given by

$$d = \left[ \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right]^{-1/2}$$

Here  $h = 1$ ,  $k = 1$  and  $l = 1$

Also  $a = b = 0.25 \text{ nm}$  and  $c = 0.18 \text{ nm}$

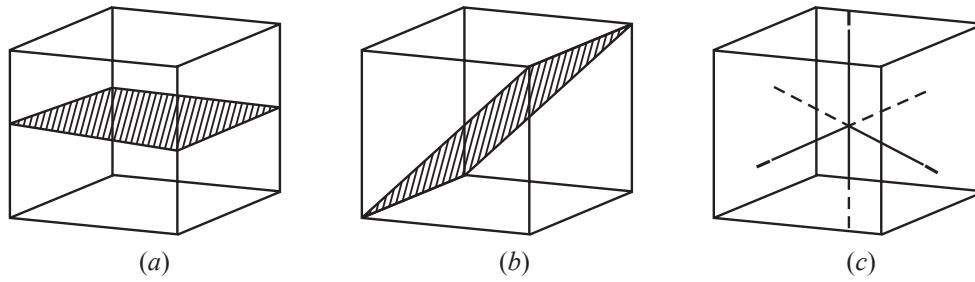
$$\therefore d_{111} = \left[ \frac{1^2}{(0.25)^2} + \frac{1^2}{(0.25)^2} + \frac{1^2}{(0.18)^2} \right]^{-1/2} = 0.126 \text{ nm}$$

## 25.5 ELEMENTS OF SYMMETRY

What is meant by a symmetry? Formally, we might say that a symmetry of a particular kind exists, when a certain operation leaves something unchanged. A candle is symmetric about a vertical axis because it can be rotated about that axis without changing in appearance or any other feature.

Crystals possess different symmetries. The symmetry operations of a crystal carry the crystal structure into itself. Symmetry operations performed about a point or line are called point-group symmetry. The point-group symmetry elements possessed by a crystal are :

(1) **Plane of symmetry.** A crystal is said to have a *plane of symmetry*, when it is divided by an imaginary plane into two halves, such that one is the mirror image of the other. For example, a cube has nine planes of symmetry. Fig. 25.19 (a) represents a plane of symmetry parallel to the faces of a cube. There are three such planes. Fig. 25.19 (b) represents one of the six diagonal planes of symmetry.



**Fig. 25.19**

(2) **Axis of symmetry.** This is an axis such that, if the crystal is rotated around it through some angle, the crystal remains invariant. The axis is called *n*-fold if the angle of rotation is  $360^\circ/n$ .

**Example.** If equivalent configuration occurs after rotation of  $180^\circ$ ,  $120^\circ$  and  $90^\circ$ , the axes of rotation are known as two-fold, three-fold and four-fold, axes of symmetry respectively. For example, Fig. 25.19 (c) shows the three tetrad (four-fold) axes of a cube.

(3) **Centre of symmetry.** It is a point such that any line drawn through it will meet the surface of the crystal at equal distances on either side. It is important to mention here that a crystal may possess a number of planes or axes of symmetry but it can have only one centre of symmetry. *Example.* A cube has a single centre of symmetry at the centre of the cube.

**Point Group.** A lattice point group is defined as the collection of the symmetry operations which, when applied about a lattice point, leave the lattice invariant. 32 such groups are possible. Every known crystal can be assigned to one of these 32 crystal classes.

**Space Groups.** When identical objects are placed at the lattice points of a space lattice a regular special array is obtained, which has, perhaps, symmetry elements over and above those of the point group of the original lattice. Thus, because of some symmetry property of the identical objects (*i.e.*, of the unit cells in a crystal), there may exist *glide* (= reflection + translation) planes and *screw* (= rotation + translation) axes which define operations turning the structure into itself. It has been shown that there are only 230 different types of symmetry possible for such a system. Thus 230 three dimensional space groups or "structures" exist in total.

#### Symmetry Operations

A symmetry operation is one which leaves the crystal invariant. Symmetry operations may be grouped into three classes:

1. Translation operations
2. Point operations
3. Hybrid operations

**1. Translation Operation.** A translation operation is defined as the displacement of a crystal parallel to itself by a crystal translation vector defined by  $\mathbf{T} = n_1\mathbf{a} + n_2\mathbf{b}$ .

#### 2. Point Operations

**(i) The mirror reflection:** In this operation, the reflection of a structure at a mirror plane  $m$  passing through a lattice point leaves the crystal unchanged. Fig. 25.20 shows two mirror planes in a two-dimensional crystal.

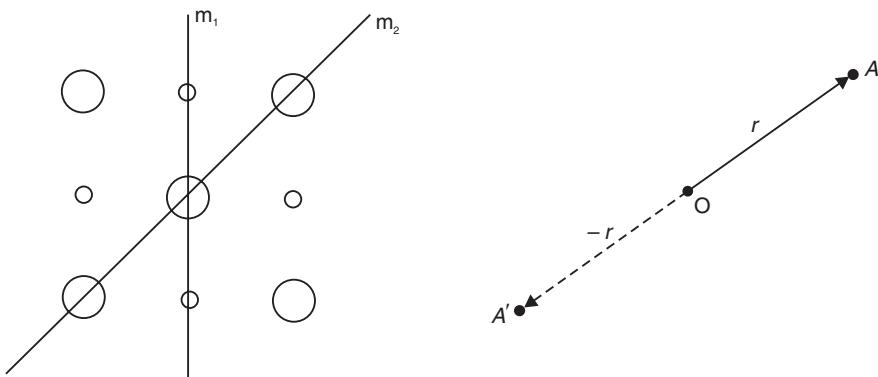


Fig. 25.20

Fig. 25.21

(ii) **Inversion:** A crystal structure has an inversion symmetry if for every lattice point of position vector  $r$ , there is a corresponding lattice point at the position  $-r$  [Fig. 25.21]. The origin about which the position vector  $r$  is chosen is called the *centre of inversion* and is represented by  $i$ .

(iii) **Rotation:** A structure is said to possess rotational symmetry about an axis if rotation of the structure about this axis by an angle  $\phi$  gives an unchanged configuration of the original structure. The angle  $\phi$  which satisfies this condition is given by

$$\phi = \frac{2\pi}{n}; n = 1, 2, 3, \dots$$

where the rotation axis is  $n$ -fold. In a crystal lattice only one, two, three, four, and six-fold rotation axes of symmetry are possible. Fig. 25.22 shows a 6-fold axis of rotation.

### 3. Hybrid Operations

(i) **Rotoreflection:** It is the combination of an  $n$ -fold rotation followed by a reflection in a plane perpendicular to the rotation axis. It is denoted by  $n/m$ , where  $n = 1, 2, 3, 4, 6$ .

(ii) **Rotoinversion:** This is the combination of an  $n$ -fold rotation followed by an inversion and is denoted by  $\bar{n}$  with  $n = 1, 2, 3, 4, 6$ .

The axis of rotation in rotoreflection and rotoinversion is referred to as improper axis.

(iii) **Screw translation:** In this the  $n$ -fold rotation axis is coupled with the translation parallel to rotation axis. The new symmetry element is called a *screw axis* and is denoted by  $n_m$ .

(iv) **Glide reflection:** We can combine a reflection with a translation parallel to the reflection plane to achieve congruence in some crystals. This combination is called a glide reflection. The translation component of a glide reflection is expressed as the combination of translations by fractions of repeat distances in a crystal. For example, the translation in a diagonal glide may be written as

$$t = \frac{1}{2}\mathbf{a} + \frac{1}{2}\mathbf{b}, \frac{1}{2}\mathbf{b} + \frac{1}{2}\mathbf{c} \text{ or } \frac{1}{2}\mathbf{c} + \frac{1}{2}\mathbf{a}.$$

The list of symmetry operations is not without a very special symmetry operation, called the *identity operation*. This operation means doing no operation on the crystal or keeping the crystal undisturbed. Thus every crystal has the identity operation. It is generally denoted by the symbol  $I$ .

#### Absence of Five-fold Symmetry Axis

**Allowed rotation :** An axis of *rotational symmetry* is an axis such that the lattice takes up an identical position on rotation through a suitable angle about it.

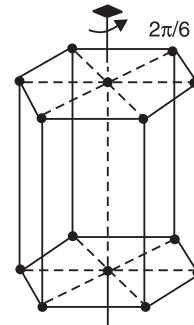


Fig. 25.22

The angle of rotation must be of the form  $2\pi/n$ .

Here  $n$  is an integer and such an axis is called an  $n$ -fold axis.

*A crystal can have only one, two, three, four, and six-fold axes of symmetry.*

**Proof.** Let  $A$  (Fig. 25.23) be one of the points of a crystal lattice through which an axis of symmetry passes (perpendicular to the plane of the diagram).  $a$  is the lattice parameter.  $B$  is another point of the lattice separated by one lattice vector from  $A$ . Another identical axis of symmetry must pass through  $B$ . Now, rotate the crystal through an angle  $\theta = 2\pi/n$ , about the axis through  $A$ . Then, the point  $B$ , as well as the axis through  $B$ , takes the position  $B'$ . Similarly, a rotation about  $B$  takes  $A$  to  $A'$ .

By assumption  $A$  and  $B$  are lattice points of the same crystal and hence can replace each other by a parallel displacement. So it is apparent that  $A' B'$  must also be a translational period of the lattice. Thus, if  $a$  is the shortest period in the given direction, then the distance  $A' B' = pa$ .

Here,  $p$  is an integer.

$$\begin{aligned} A' B' &= A' C + CD + DB' \\ &= a \sin\left(\theta - \frac{\pi}{2}\right) + a + a \sin\left(\theta - \frac{\pi}{2}\right) \\ &= a - 2a \cos\theta = pa \end{aligned}$$

i.e.,

$$\cos\theta = (1-p)/2.$$

Since  $|\cos\theta| \leq 1$ ,  $p$  can only have the values  $p = 0, \pm 1, 2, 3$ .

Now we have to find the values of  $\theta$  and  $n$ .

$p$	$\cos\theta$	$\theta$	$n = \frac{360}{\theta}$
0	$\frac{1}{2}$	60	6
+1	0	90	4
-1	+1	360 or 0	1
2	$-\frac{1}{2}$	120	3
3	-1	180	2

This proves that a crystal can have only 1, 2, 3, 4, and 6-fold axes of rotation and it cannot have a five fold rotation axis.

This is illustrated in Fig. 25.24. We observe that an array of pentagons cannot cover all space.

The array has point symmetry, i.e., 5-fold rotational axis but not the translational symmetry.

**CALCULATION OF NUMBER OF ATOMS PER UNIT CELL, ATOMIC RADIUS, COORDINATION NUMBER AND PACKING FACTOR FOR SIMPLE CUBIC, BCC, FCC, HCP AND DIAMOND STRUCTURES.**

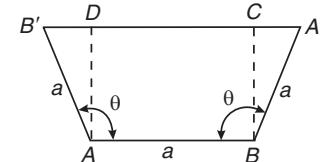


Fig. 25.23

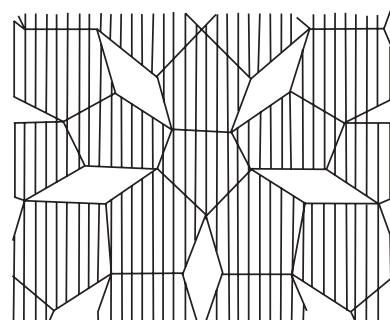


Fig. 25.24

## 25.6 DEFINITIONS

(1) **Atomic radius  $r$**  : Atomic radius is defined as half the distance between two nearest neighbouring atoms in a crystal.

(2) **Nearest Neighbour Distance ( $2r$ )**: The distance between the centres of two nearest neighbouring atoms is called *nearest neighbour distance*. If  $r$  is the radius of the atom, nearest neighbour distance is  $2r$ .

(3) **Co-ordination number** : *Coordination number is the number of nearest neighbour atoms that an atom has in the given crystal structure.*

**Explanation.** Each atom in a unit cell has some nearest neighbouring atoms. The number of these nearest neighbours to a given atom in crystal lattice is called the coordination number ( $CN$ ) in crystal lattice. Thus the number of atoms in the immediate neighbourhood of a given atom in a crystal is called the co-ordination number ( $CN$ ).

The coordination number is calculated by taking the nearest neighbouring atoms at a distance,  $2r$ .

It gives an indication of the closeness of the packing of the atoms.

(4) **Packing factor** : *Atomic packing factor is the ratio of the volume of atoms in the unit cell to the volume of the unit cell.*

**Explanation.** The ratio of volume occupied by atoms in a unit cell to the total volume of the unit cell is called *packing factor*.

$$\text{Packing Factor, } \text{P.F.} = v/V.$$

Here,  $v$  = volume occupied by the atoms in the unit cell and

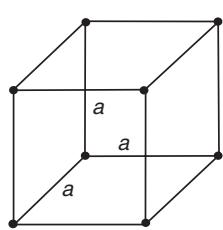
$V$  = total volume of the unit cell.

A high value of P.F. indicates that the atoms are very closely packed in the unit cell and there is very small unoccupied space. A low value of P.F. indicates that the atoms are very loosely packed in the unit cell and there is enormous unoccupied space.

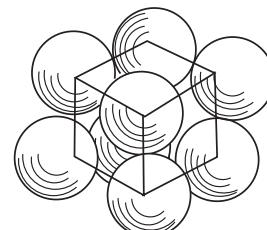
We assume that the atoms are hard spheres and that the nearest neighbours touch each other.

## 25.7 SIMPLE CUBIC (SC) STRUCTURE

The cubic unit cell for simple cubic lattice is shown in Fig. 25.25.



(a) atomic-site model



(b) hard sphere model

Fig. 25.25

### (i) Number of atoms per unit cell

There are 8 atoms at the 8 corners of the cell.

Each corner atom is shared by 8 unit cells that adjoin at each corner.

The share of each corner atom to a unit cell is  $1/8$  of an atom.

Total number of atoms per unit cell is  $8 \times \frac{1}{8} = 1$ .

**(ii) Atomic radius ,  $r$** 

In a SC cell the atoms are in contact along the edges of the cube (Fig. 25.26).

$a$  is length of side of cubic cell.

The atoms touch along cube edges.

Nearest neighbour distance,

$$2r = a$$

$$\therefore r = a/2.$$

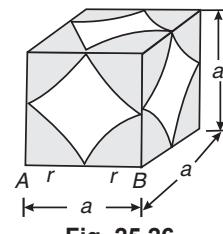


Fig. 25.26

**(iii) Coordination number, CN**

Each corner atom touches four atoms in its horizontal plane, and two atoms, one vertically above it and one vertically below it. So there are six equally spaced nearest neighbour atoms each at a distance ' $a$ ' from that atom. Therefore, the coordination number is six.

$$CN = 6.$$

**(iv) Packing Factor**

$$\begin{aligned} \text{Packing factor} &= \frac{\text{Volume of all atoms in the unit cell}}{\text{Volume of the unit cell}} \\ &= \frac{\text{No. of atoms in unit cell} \times \text{Volume of one atom}}{\text{Volume of the unit cell}} \\ &= \frac{1 \times \frac{4}{3}\pi r^3}{(a)^3} = \frac{4\pi r^3}{3(2r)^3} \quad (\because a = 2r) \\ &= \frac{\pi}{6} = 0.52. \end{aligned}$$

$\therefore$  Packing factor = 52%

Simple cubic structure is a loosely packed structure.

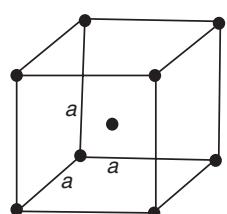
Polonium is the only element which exhibits this structure.

## 25.8 BODY CENTRED CUBIC (BCC) STRUCTURE

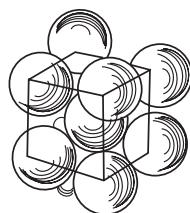
Fig. 25.27 (a) shows the arrangement of lattice points in a BCC cell.

Fig. 25.27 (b) shows the arrangement of atoms represented as hard spheres.

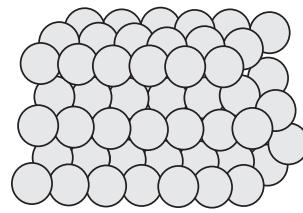
Fig. 25.27 (c) shows the BCC repeating structure.



(a) atomic site model



(b) hard sphere model



(c) repeating structure in a crystal

Fig. 25.27

Let us now compute the characteristics of a BCC cell.

**(i) Number of atoms per unit cell**

There are 8 atoms at the 8 corners, each shared by 8 unit cells and one atom at the body centre. The atom at the centre of the body of the cell wholly belongs to the unit cell.

Thus, the total number of atoms per unit cell is  $8 \times \frac{1}{8} + 1 = 2$ .

**(ii) Atomic radius,  $r$**

In this case, atoms touch each other along the diagonal of the cube (Fig. 25.28).

The length of the body diagonal =  $4r$ .

$$AG^2 = AC^2 + CG^2 = (AB^2 + BC^2) + CG^2$$

$$(4r)^2 = 3a^2$$

$$\text{Atomic radius } r = \frac{\sqrt{3}a}{4}$$

$$\text{Lattice constant } a = \frac{4}{\sqrt{3}}r.$$

**(iii) Coordination number,  $CN$**

The unit cell contains one atom at each of the eight corners of the unit cell and one atom at the body centre. The nearest atom to the corner atom is the body centred atom and not the other corner atoms. The number of nearest neighbours is eight (Fig. 25.29).

The body centre atom is in contact with all the eight corner atoms. Therefore, the coordination number is eight.

$$CN = 8.$$

**(iv) Packing Factor**

$$\begin{aligned} \text{Packing factor} &= \frac{\text{Volume of all atoms in the unit cell}}{\text{Volume of the unit cell}} \\ &= \frac{\text{No. of atoms in unit cell} \times \text{Volume of one atom}}{\text{Volume of the unit cell}} \\ &= \frac{2 \times (4/3)\pi r^3}{a^3} = \frac{8\pi r^3}{3a^3} \\ &= \frac{8\pi \left(\frac{\sqrt{3}a}{4}\right)^3}{3a^3} \quad \left(\because r = \frac{\sqrt{3}a}{4}\right) \\ &= \frac{\sqrt{3}\pi}{8} = 0.68. \end{aligned}$$

$\therefore$  Packing factor = 68%.

Thus the packing factor is equal to 68%. So it is not a closely packed structure.

Tungsten, sodium, iron and chromium have this type of structure.

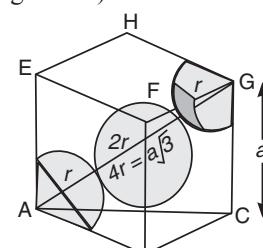


Fig. 25.28

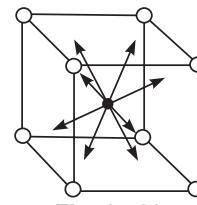


Fig. 25.29

## 25.9 FACE CENTRED CUBIC (FCC) STRUCTURE

The unit cell of this structure is shown in Fig. 25.30.

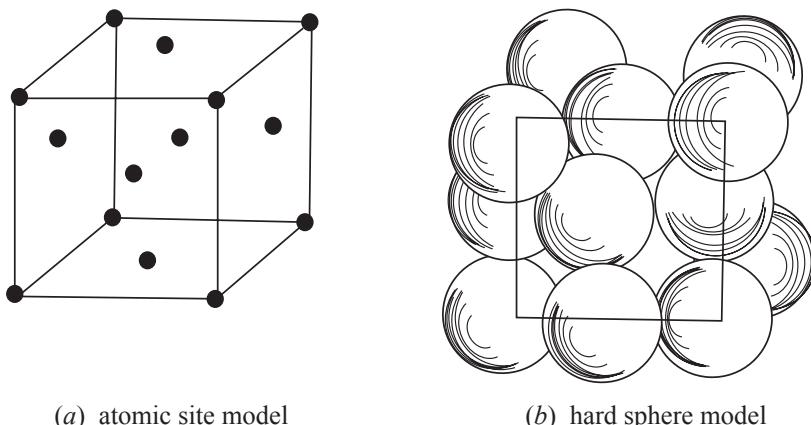


Fig. 25.30

The face centred cubic unit cell is a nonprimitive cell.

It has six atoms at the centres of six faces and eight atoms at the eight corners of the cube.

We shall now determine the characteristics of the *FCC* cell.

#### (i) Number of atoms per unit cell

There are 8 atoms, each shared by 8 unit cells at 8 corners and 6 face-centred atoms, each shared by 2 cells.

The total number of atoms per unit cell = contribution due to 8 corner atoms + contribution due to 6 face centred atoms

$$= \left(\frac{1}{8} \times 8\right) + \left(\frac{1}{2} \times 6\right) = 1 + 3 = 4$$

Thus the total number of atoms per unit cell is 4.

#### (ii) Atomic radius, $r$

In a *FCC* cell, atoms are in contact along the face diagonal of the cube (Fig. 25.31).

Length of the face diagonal,  $AF = 4r$ .

$$\begin{aligned} AF^2 &= AB^2 + BF^2 \\ (4r)^2 &= a^2 + a^2 = 2a^2 \\ 4r &= \sqrt{2}a \end{aligned}$$

$$\therefore \text{Atomic radius, } r = \frac{a}{2\sqrt{2}}$$

$$\text{Lattice constant } a = 2\sqrt{2}r.$$

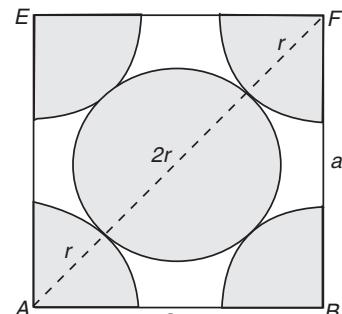


Fig. 25.31

#### (iii) Coordination number CN

In this case, the nearest neighbours of any corner atom are the face centred atoms of the surrounding unit cells. Any corner atom has four such atoms in its own plane, four in a plane above it and four in a plane below it.

Therefore, the coordination number  $CN = 12$ .

Theoretically, a sphere can touch simultaneously 12 identical spheres at the maximum in a three dimensional structure. Thus, an *FCC* cell has the maximum value for the coordination number.

## (iv) Packing Factor

$$\begin{aligned}
 \text{Packing Factor} &= \frac{\text{Volume of all atoms in the unit cell}}{\text{Volume of the unit cell}} \\
 &= \frac{\text{No. of atoms in unit cell} \times \text{Volume of one atom}}{\text{Volume of unit cell}} \\
 &= \frac{4 \times \frac{4}{3} \pi r^3}{a^3} \\
 &= \frac{16}{3} \cdot \frac{\pi r^3}{(2\sqrt{2}r)^3} \quad (\because a = 2\sqrt{2}r) \\
 &= \frac{\pi}{3\sqrt{2}} = 0.74
 \end{aligned}$$

$$\therefore \text{Packing factor} = 74\%$$

When compared with SC and BCC, this has the highest packing factor.

Most of the metals like copper, aluminium, lead and silver have this structure.

## 25.10 | CALCULATION OF LATTICE CONSTANT

Consider a cubic lattice of lattice constant  $a$ .

Let  $\rho$  be the density of the crystal.

Volume of the unit cell =  $a^3$

$$\therefore \text{Mass in each unit cell} = \rho a^3 \quad \dots(1)$$

Let  $n$  = No. of molecules (lattice points) per unit cell,

$M$  = molecular weight and

$N$  = Avogadro number (*i.e.*, number of molecules per kg mole of the substance).

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

$$\text{Mass of each unit cell} = n \times \frac{M}{N} \quad \dots(2)$$

From Eqs. (1) and (2), we have

$$\begin{aligned}
 \rho a^3 &= n \cdot \frac{M}{N} \text{ or } a^3 = \frac{nM}{\rho N} \\
 \therefore a &= \left( \frac{nM}{\rho N} \right)^{1/3}. \quad \dots(3)
 \end{aligned}$$

**EXAMPLE 1.** A substance with face centred cubic lattice has density  $6250 \text{ kg/m}^3$  and molecular weight 60.2. Calculate the lattice constant  $a$ . Given: Avogadro number =  $6.02 \times 10^{26} (\text{kg.mole})^{-1}$ .

$$\text{SOL.} \quad a = \left( \frac{nM}{\rho N} \right)^{1/3}$$

Here  $n$  = number of molecules per unit cell in FCC lattice = 4

$$M = 60.2, N = 6.02 \times 10^{26} \text{ and } \rho = 6250 \text{ kg/m}^3.$$

$$a = \left[ \frac{4 \times 60.2}{(6.02 \times 10^{26}) \times 6250} \right]^{\frac{1}{3}} = 0.4 \times 10^{-9} \text{ m} = 0.4 \text{ nm}.$$

**EXAMPLE 2.** *NaCl crystals have FCC structure. The density of sodium chloride is 2180 kg m<sup>-3</sup>. Calculate the distance between two adjacent atoms. Atomic weight of sodium = 23 and that of chlorine = 35.5.*

**SOL.**  $a = \left( \frac{nM}{N\rho} \right)^{\frac{1}{3}}$

Here  $n = 4$ ,  $M = 23 + 35.5 = 58.5$ ,  $N = 6.02 \times 10^{26}$  (kg mole)<sup>-1</sup> and  $\rho = 2180 \text{ kg m}^{-3}$ .

$$\therefore a = \left[ \frac{4 \times 58.5}{(6.02 \times 10^{26}) \times 2180} \right]^{\frac{1}{3}} = 0.5657 \times 10^{-9} \text{ m}.$$

The distance between two adjacent atoms is half of this value, i.e.,

$$d = \frac{a}{2} = \frac{0.5657 \times 10^{-9}}{2} \text{ m} = 0.2828 \times 10^{-9} \text{ m} = 0.2828 \text{ nm.}$$

**EXAMPLE 3.** *Copper has FCC structure and the atomic radius is 1.278 Å. Calculate its density.*

**SOL.**  $\rho = \frac{nM}{Na^3}$

Here  $n = 4$  atoms,  $M = 63.54$ ,  $N = 6.023 \times 10^{26}$  (kg mole)<sup>-1</sup>,  $r = 1.278 \text{ \AA} = 1.278 \times 10^{-10} \text{ m}$ .

$$a = 2\sqrt{2}r = 2 \times 1.414 \times (1.278 \times 10^{-10}) \text{ m} = 3.614 \times 10^{-10} \text{ m.}$$

$$\rho = \frac{4 \times 63.54}{(6.023 \times 10^{26}) \times (3.614 \times 10^{-10})^3} = 8940 \text{ kg m}^{-3}.$$

**EXAMPLE 4.** *Lead has a face centred cubic structure with an atomic radius of 1.746 Å. Find the spacing of (i) 200 planes and (ii) 220 planes.*

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

The lattice constant of unit cell of lead (FCC) structure is related to its atomic radius by the relation

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

(i) For (200) plane,

$$d_{200} = \frac{4.93}{\sqrt{(2)^2 + (0)^2 + (0)^2}} = \frac{4.93}{2} = 2.465 \text{ \AA}.$$

(ii) For (220) planes

$$d_{220} = \frac{4.93}{\sqrt{(2)^2 + (2)^2 + (0)^2}} = \frac{4.93}{2\sqrt{2}} = 1.74 \text{ \AA}.$$

**EXAMPLE 5.** Copper has an FCC structure with lattice constant  $a = 3.61 \text{ \AA}$ . Calculate the radius of copper-atom.

$$\text{SOL. } r = \frac{\sqrt{2}a}{4} = \frac{\sqrt{2} \times 3.61}{4} \text{ \AA} = 1.28 \text{ \AA}.$$

**EXAMPLE 6.** Metallic iron changes from BCC to FCC form at  $910^\circ\text{C}$ . At this temperature, the atomic radii of the iron atoms in the two structures are  $1.258 \text{ \AA}$  and  $1.292 \text{ \AA}$  respectively. Calculate the percent volume change during this structural change.

**SOL.**

$$(i) \text{ In BCC, } a = \frac{4r}{\sqrt{3}} = \frac{4 \times 1.258 \text{ \AA}}{\sqrt{3}} = 2.905 \text{ \AA} = 2.905 \times 10^{-10} \text{ m.}$$

$$\text{Volume of unit cell} = a^3 = (2.905 \times 10^{-10})^3 = 24.515 \times 10^{-30} \text{ m}^3.$$

$$\text{Number of atoms per unit cell} = 2$$

$$\therefore \text{Volume occupied by one atom} = \frac{24.515 \times 10^{-30}}{2} = 12.26 \times 10^{-30} \text{ m}^3.$$

$$(ii) \text{ In FCC, } a = 2\sqrt{2}r = 2\sqrt{2} \times 1.292 \text{ \AA}$$

$$= 2\sqrt{2} \times (1.292 \times 10^{-10}) \text{ m} = 3.654 \times 10^{-10} \text{ m.}$$

$$\text{Volume of unit cell} = a^3 = (3.654 \times 10^{-10})^3 = 48.787 \times 10^{-30} \text{ m}^3.$$

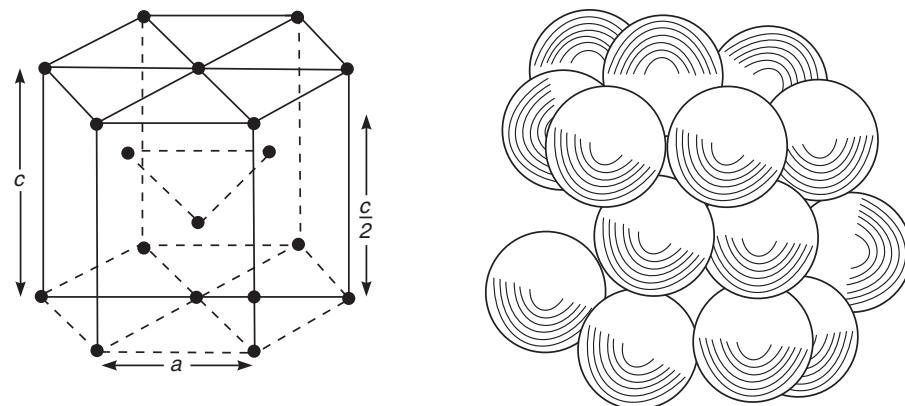
$$\text{Number of atoms per unit cell} = 4$$

$$\text{Volume occupied by one atom} = \frac{48.787 \times 10^{-30}}{4} = 12.2 \times 10^{-30} \text{ m}^3$$

$$\therefore \text{Volume change in \%} = \frac{12.26 - 12.2}{12.26} \times 100 = 0.489\%.$$

## 25.11 HEXAGONAL CLOSE PACKED (HCP) STRUCTURE

The conventional unit cell for hexagonal close packed structure is shown in Fig. 25.32.



(a) atomic site model

(b) hard sphere model

Fig. 25.32

There are three layers of atoms in the unit cell.

- (i) At the bottom layer, central atom has six nearest neighbour atoms in the same plane.
- (ii) Top layer has the same atomic arrangement as the bottom layer.
- (iii) At  $c/2$  distance from the bottom layer there is a middle layer containing three atoms.

Let us now compute the characteristics of HCP cell.

#### (i) Number of atoms per unit cell

The top layer contains 6 atoms at the corners and one atom at the centre.

The corner atoms are shared between 6 unit cells. Hence 6 atoms contribute only 1 atom to the unit cell.

The central atom is shared between 2 unit cells and therefore contributes  $\frac{1}{2}$  atom to unit cell.

Hence the top layer contributes  $1 + \frac{1}{2} = \frac{3}{2}$  atoms to the unit cell.

Similarly, the bottom layer contributes  $\frac{3}{2}$  atoms to the unit cell.

The three atoms of the middle layer lie within the volume of the unit cell and hence fully ( 3 atoms ) contribute to the cell.

Thus, the total number of atoms/unit cell.

$$= \frac{3}{2} + \frac{3}{2} + 3 = 6.$$

#### (ii) Atomic radius, $r$

The atoms are in contact along the edges of the hexagon (Fig. 25.33)

$$\begin{aligned} \therefore \quad & 2r = a \\ \text{or} \quad & r = a/2. \end{aligned}$$

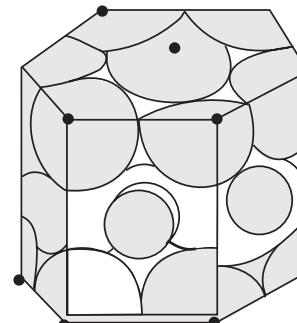


Fig. 25.33

#### (iii) Coordination number, CN

Consider the central atom at the top layer of HCP structure. It touches 6 atoms in its plane, 3 atoms in the plane just below it and 3 atoms in the plane just above it. Hence the coordination number is 12.

#### (iv) Volume of the unit cell

The volume of the unit cell is determined by computing the area of the base of the unit cell and then by multiplying it by the cell height. The area of the base of the unit cell is the area of the hexagon ABDEFG (Fig. 25.34). It is equal to the sum of the areas of the six equal-sized triangles.

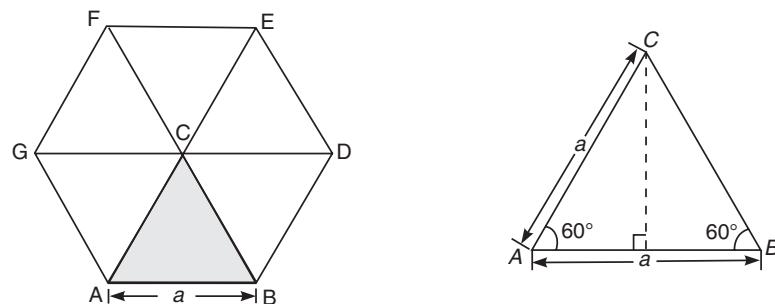


Fig. 25.34

$\therefore$  Area of hexagon  $ABDEFG = 6$  (area of  $\Delta ABC$ )

$$\begin{aligned} &= 6 \times \left[ \frac{1}{2} \times a \times (a \sin 60^\circ) \right] \\ &= 3a^2 \left( \frac{\sqrt{3}}{2} \right) = \frac{3\sqrt{3}}{2} a^2. \end{aligned}$$

$\therefore$  Volume of the unit cell,  $V = (\text{area of the base}) \times (\text{Height of the cell})$

$$= \left( \frac{3\sqrt{3}}{2} a^2 \right) c.$$

#### Calculation of c/a ratio

Let  $c$  be the height of the unit cell and  $a$  be its edge. Also  $a = 2r$ .

The three body atoms lie in a horizontal plane at a height  $c/2$  from the orthocentres of alternate equilateral triangles in the base or at top of the hexagonal cell.

From Fig. 25.35,

$$(2r)^2 = x^2 + \left( \frac{c}{2} \right)^2$$

But,

$$x = \frac{2}{3} AN = \frac{2}{3} \left( a \frac{\sqrt{3}}{2} \right) = \frac{a}{\sqrt{3}}$$

Also

$$2r = a$$

$\therefore$

$$a^2 = \left( \frac{a}{\sqrt{3}} \right)^2 + \left( \frac{c}{2} \right)^2 = \frac{a^2}{3} + \frac{c^2}{4}$$

$$\frac{c^2}{4} = a^2 - \frac{a^2}{3} = \frac{2a^2}{3}$$

or

$$\frac{c^2}{a^2} = \frac{8}{3}$$

$\therefore$

$$\frac{c}{a} = \sqrt{\frac{8}{3}} = 1.633.$$

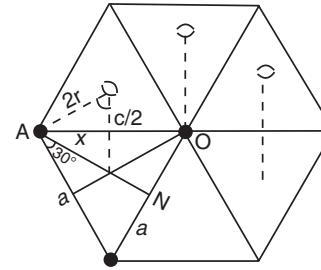


Fig. 25.35

#### (v) Calculation of packing factor

$$\text{Packing factor} = \frac{\text{Volume of all atoms in the unit cell}}{\text{Volume of the unit cell}}$$

$$= \frac{\text{No. of atoms in unit cell} \times \text{Volume of one atom}}{\text{Volume of the unit cell}}$$

$$= \frac{6 \times \frac{4}{3} \pi r^3}{\frac{3\sqrt{3}}{2} a^2 c} = \frac{16\pi r^3}{3\sqrt{3} a^2 c}$$

$$= \frac{16\pi (a/2)^3}{3\sqrt{3} a^2 c} \quad \left( \because r = \frac{a}{2} \right)$$

$$= \frac{2\pi a^3}{3\sqrt{3} a^2 c} = \frac{2\pi}{3\sqrt{3}} \frac{a}{c} = \frac{2\pi}{3\sqrt{3}} \sqrt{\frac{3}{8}} \quad \left( \because \frac{c}{a} = \sqrt{\frac{8}{3}} \right)$$

$$= \frac{\pi}{3\sqrt{2}} = 0.74$$

$\therefore$  Packing factor = 74%

Since the density of packing is 74%, it is a close packed structure.

Magnesium, zinc and cadmium crystallize in this structure.

**EXAMPLE 1.** Zinc has HCP structure. The height of the unit cell is 0.494 nm. The nearest neighbour distance is 0.27 nm. The atomic weight of zinc is 65.37. Calculate the volume of the unit cell and density of zinc.

**SOL.** Here,  $a = 2r = 0.27 \times 10^{-9}$  m;  $c = 0.494 \times 10^{-9}$  m.

Volume of the unit cell.

$$V = \frac{3\sqrt{3}a^2c}{2} = \frac{3\sqrt{3} \times (0.27 \times 10^{-9})^2 \times (0.494 \times 10^{-9})}{2}$$

$$= 9.356 \times 10^{-29} \text{ m}^3.$$

Let

$\rho$  = density of the crystal.

Mass of the unit cell =  $\rho V = (9.356 \times 10^{-29})\rho$  kg.

The number of atoms belonging to a unit cell of HCP = 6.

If  $M$  is the atomic weight and  $N$  the Avogadro number, then

$$\text{mass per unit cell} = \frac{6M}{N} = \frac{6 \times 65.37}{6.02 \times 10^{26}}$$

$$(9.356 \times 10^{-29}) \rho = \frac{6 \times 65.37}{6.02 \times 10^{26}}$$

$$\rho = \frac{6 \times 65.37}{6.02 \times 10^{26} \times (9.356 \times 10^{-29})}$$

$$\rho = 6964 \text{ kg m}^{-3}.$$

## 25.12 DIAMOND CUBIC STRUCTURE

The diamond structure is a combination of two inter-penetrating face-centred cubic (FCC) sublattices.

The unit cell is sketched in Fig. 25.36 (a).

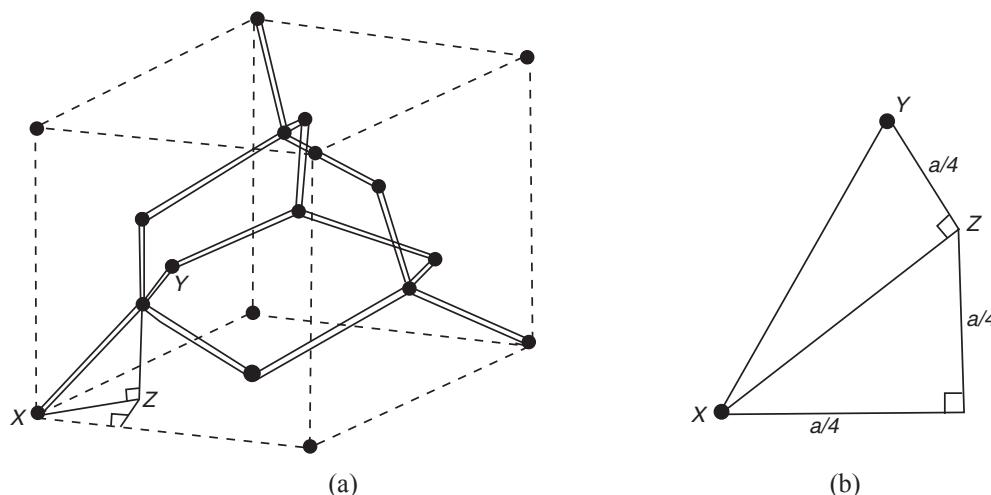


Fig. 25.36

One sublattice, say  $X$ , has its origin at  $(0, 0, 0)$ . The other sublattice,  $Y$ , has its origin quarter of the way along the body diagonal i.e., at the point  $(a/4, a/4, a/4)$ .

The diamond cubic structure is loosely packed, since each atom has only four nearest neighbours.

#### (i) Number of atoms per unit cell

In the unit cell, in addition to the eight corner atoms, there are six face centred atoms and four more atoms are located inside the unit cell. Each corner atom is shared by eight adjacent unit cells and each face centred atom is shared by two unit cells. Hence the total number of atoms per unit cell is,

$$\left(\frac{1}{8} \times 8\right) + \left(\frac{1}{2} \times 6\right) + 4 = 8.$$

#### (ii) Atomic radius, $r$

From Fig. 25.36 (b),

$$XZ^2 = \left(\frac{a}{4}\right)^2 + \left(\frac{a}{4}\right)^2 = \frac{a^2}{8}$$

$$\text{Also } XY^2 = XZ^2 + YZ^2 = \frac{a^2}{8} + \frac{a^2}{16} = \frac{3a^2}{16}.$$

But

$$XY = 2r$$

$$\therefore (2r)^2 = \frac{3a^2}{16}$$

$$r = \frac{\sqrt{3}}{8}a$$

$$\text{Lattice constant } a = \frac{8r}{\sqrt{3}}.$$

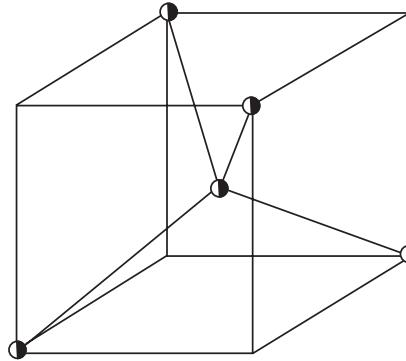


Fig. 25.37

#### (iii) Coordination Number

In a diamond crystal, the carbon atoms are linked by directional covalent bonds. Each carbon atom forms covalent bonds with four other carbon atoms that occupy four corners of a cube in a tetrahedral structure (Fig. 25.37).

In the diamond lattice, each atom has four nearest neighbours with which it forms covalent bonds. Thus the coordination number of diamond crystal is 4.

#### (iv) Packing Factor

$$\text{Packing factor} = \frac{\text{Volume of all atoms in the unit cell}}{\text{Volume of the unit cell}}$$

$$= \frac{\text{No. of atoms in unit cell} \times \text{Volume of one atom}}{\text{Volume of the unit cell}}$$

$$= \frac{8 \times \frac{4}{3}\pi r^3}{a^3}$$

$$= \frac{32\pi r^3}{3\left(\frac{8r}{\sqrt{3}}\right)^3} \quad \left(\because a = \frac{8r}{\sqrt{3}}\right)$$

$$= \frac{\pi\sqrt{3}}{16} = 0.34$$

$\therefore$  Packing factor = 34 %

Thus it is a loosely packed structure.

Carbon, silicon, germanium and grey tin crystallize in diamond structure.

**EXAMPLE 1.** Silicon has the same structure as diamond and a bond length of 0.2351 nm. Estimate its density.

**SOLUTION.** In the case of diamond cubic structure,

$$\text{Nearest neighbour distance } 2r = \frac{\sqrt{3}}{4} a .$$

Number of atoms per unit cell = 8

$$a = \frac{4}{\sqrt{3}}(2r) = \frac{4}{\sqrt{3}} \times 0.2351 \text{ nm} = 0.543 \text{ nm} = 0.543 \times 10^{-9} \text{ m}$$

$$\rho = \frac{nM}{Na^3} = \frac{8 \times 28.09}{(6.02 \times 10^{26})(0.543 \times 10^{-9})^3} = 2331 \text{ kg m}^{-3} .$$

### Summary

The characteristics of the unit cells of simple cubic, BCC, FCC, HCP and diamond cube structures are summarized in Table 25.2.

**Table 25.2**

Sl. No.	Characteristic	(SC)	(BCC)	(FCC)	(HCP)	(DC)
1.	Number of atoms per unit cell	1	2	4	6	8
2.	Atomic radius, $r$	$\frac{a}{2}$	$\frac{a\sqrt{3}}{4}$	$\frac{a}{2\sqrt{2}}$	$\frac{a}{2}$	$\frac{a\sqrt{3}}{8}$
3.	Nearest neighbour distance, $(2r)$	$a$	$\frac{a\sqrt{3}}{2}$	$\frac{a}{\sqrt{2}}$	$a$	$\frac{a\sqrt{3}}{4}$
4.	Coordination number	6	8	12	12	4
5.	Unit cell volume	$a^3$	$a^3$	$a^3$	$\frac{3\sqrt{3}}{2}a^2c$	$a^3$
6.	Packing factor	0.52	0.68	0.74	0.74	0.34
7.	Examples	Polonium	Iron, Barium, Chromium	Aluminium, Copper, Lead, Gold	Magnesium, Zinc, Cadmium	Germanium, Silicon, Carbon, diamond

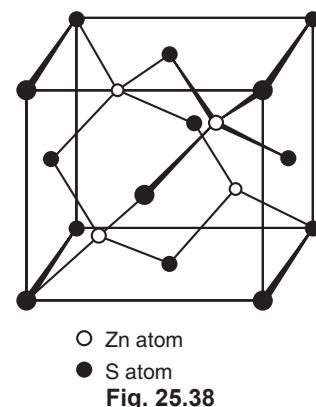
### 25.13 SIMPLE CRYSTAL STRUCTURES

(1) **Structure of zinc blende (ZnS).** The zinc blende structure consists of two *fcc* sub-lattices. Zn atoms occupy the lattice points of one sub-lattice and S atoms of the other (Fig. 25.38).

Thus each Zn atom has four S neighbours at the corners of a regular tetrahedron and *vice versa*. In this case, there are double bonds between atoms which are partially covalent and partially ionic.

Many semiconductor compounds crystallize in the Zinc-blende structure.

- Examples of the cubic zinc sulphide structure are CdS, AgI, ZnS, ZnSe and SiC.
- The conventional cell is a cube. There are four molecules of ZnS per conventional cell. About each atom there are four equally distant atoms of the opposite kind arranged at the corners of a regular tetrahedron.

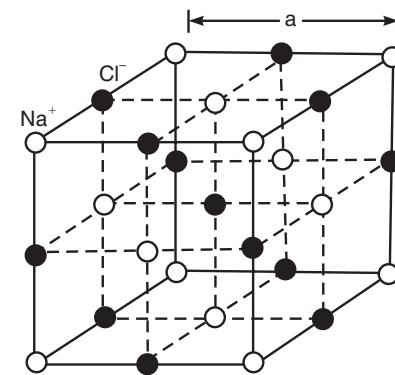


The diamond structure allows a center-of-inversion symmetry operation at the midpoint of every line between nearest-neighbour atoms. The cubic ZnS structure does not have inversion symmetry.

**(2) NaCl structure.** The NaCl crystal has a fcc structure with  $\text{Na}^+$  and  $\text{Cl}^-$  ions as the basis. This is an ionic crystal.

Fig. 25.39 represents a unit cell of NaCl lattice. The  $\text{Na}^+$  ions are situated at the corners as well as at the centres of the faces of the cube. The  $\text{Cl}^-$  ions are relatively displaced half the edge of the unit cell along each axis. Thus NaCl crystal can be thought of as composed of superposed fcc  $\text{Na}^+$  and  $\text{Cl}^-$  sub-lattices.

Each cell has 8 corners and 8 cells meet at each corner. Thus an ion at a corner of the cell is shared by 8 cells. So only  $1/8$  ion belongs to any one cell. Similarly, an ion at the centre of a face of the cell is shared by 2 cells. So only  $1/2$  ion belongs to any one cell. A cell has 8 corners and 6 faces. So it has  $\left(\frac{1}{8} \times 8\right) + \left(\frac{1}{2} \times 6\right) = 4$  ions of one kind, and similarly 4 ions of the other kind. Thus *there are 4  $\text{Na}^+$  —  $\text{Cl}^-$  ion pairs per unit cell*.



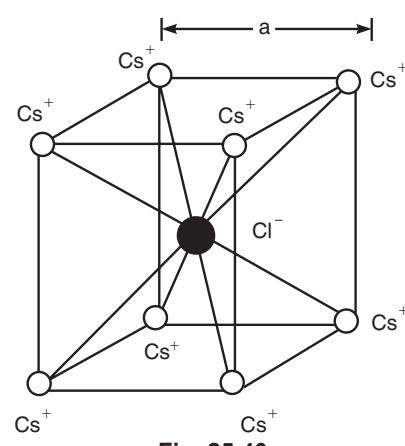
There are four molecules of NaCl in each unit cube, with ions in the positions :

$$\text{Na} : 000; \frac{1}{2} \frac{1}{2} 0; \frac{1}{2} 0 \frac{1}{2}; 0 \frac{1}{2} \frac{1}{2}$$

$$\text{Cl} : \frac{1}{2} \frac{1}{2} \frac{1}{2}; 0 0 \frac{1}{2}; 0 \frac{1}{2} 0; \frac{1}{2} 0 0$$

Each  $\text{Na}^+$  ion has 6  $\text{Cl}^-$  ions as nearest neighbours. Similarly each  $\text{Cl}^-$  ion has 6  $\text{Na}^+$  ions as nearest neighbours. Hence the coordination number of NaCl for opposite kind of ions is 6 and the nearest neighbour distance is  $a/2$ . For same kind of ions coordination number is 12 and nearest neighbour distance is  $a/\sqrt{2}$ .

**3. The Cesium Chloride Structure (CsCl).** The structure of CsCl is shown in Fig. 25.40. The CsCl crystal has simple cubic structure with  $\text{Cs}^+$  and  $\text{Cl}^-$  ions as the basis.



The  $\text{Cs}^+$  ions are situated at the corners of the cubic cell and  $\text{Cl}^-$  ion at the body centre or *vice versa*. Thus  $\text{CsCl}$  crystal may be considered to be a combination of two simple cubic sub-lattices, one of  $\text{Cs}^+$  ion and the other of  $\text{Cl}^-$  ion. There is one molecule per primitive cell, with ions at the corners 0 0 0 and body-centred positions  $\frac{1}{2} \frac{1}{2} \frac{1}{2}$  of the simple cubic space lattice. The coordination number for same kind of ions is 6 and the nearest neighbour distance is  $a$ . The coordination number for opposite kind of ions is 8 and the nearest neighbour distance is  $\frac{a\sqrt{3}}{2}$ .

**4. Determination of crystal structure by Bragg's law.** To find the structure of a crystal, the *lattice constant* is determined by using different planes of the crystal as reflecting surface for the same known wavelength of X-rays. The lattice constants are expressed as ratios with one of them taken as unity. This ratio comes to be different for different crystals. The experimentally observed ratio is then compared with the calculated ratio and thus a particular structure may be identified.

**Structure of  $KCl$  crystal.** Consider a crystal of simple cubic system in which an atom lies at each corner of the cube (Fig. 25.41). The structure is repeated in all directions constituting a *space lattice*. In this case, there are three sets of planes which are rich in atoms.

(i) The first set of planes such as  $ABFE$ ,  $CDHG$ ,  $ADHE$ ,  $BCGF$ ,  $ABCD$  and  $EFGH$  are all alike. They are called (100) planes. Let the distance between the consecutive planes be  $d_1$ .

(ii) The second set of planes consist of parallel planes like  $ADGF$  or  $ABGH$  (110 planes) inclined at  $45^\circ$  to the cubic faces. If  $d_2$  is the spacing between these planes, then  $d_2 = d_1/\sqrt{2}$ .

(iii) The third set of planes consist of planes like  $AFH$  (111 planes). The separation between these planes ( $d_3$ ) can also be calculated in terms of  $d_1$ . ( $d_3 = d_1/\sqrt{3}$ )

$$\therefore \text{For a simple cubic system, } d_1 : d_2 : d_3 = 1 : \frac{1}{\sqrt{2}} : \frac{1}{\sqrt{3}}.$$

The first order spectrum from the (100), (110) and (111) planes of  $KCl$  was observed at the glancing angles  $5.22^\circ$ ,  $7.30^\circ$  and  $9.05^\circ$  respectively.

Then since  $d \propto 1/\sin \theta$ ,

$$d_1 : d_2 : d_3 = \frac{1}{\sin 5.22^\circ} : \frac{1}{\sin 7.30^\circ} : \frac{1}{\sin 9.05^\circ} = 1 : 1/\sqrt{2} : 1/\sqrt{3}.$$

This is in agreement with the theoretical values. Hence  $KCl$  belongs to the simple cubic system. But, from these studies, no information is obtained as to the nature of the constituents at the lattice corners, whether they are molecules of  $KCl$  or ions of  $K$  and  $Cl$  alternately. The answer to this question is obtained from observations made on  $\text{NaCl}$  crystal.

**Structure of  $\text{NaCl}$  crystal.** The results obtained with  $\text{NaCl}$  crystal for the same three planes are identical with the previous case. However, a weak reflection at the (111) face at about  $5^\circ$  is also obtained. Taking this into account, we find that  $d_1 : d_2 : d_3 = 1 : 1/\sqrt{2} : 2/\sqrt{3}$ , which agrees with the calculated values for a face-centred cube. Hence  $\text{NaCl}$  appears to belong to the face-centred cubic system.

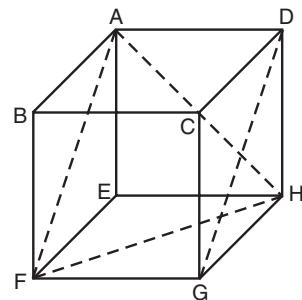


Fig. 25.41

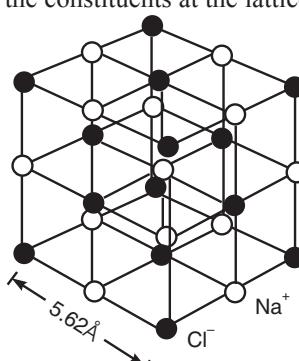


Fig. 25.42

The difference between the  $KCl$  and  $NaCl$  crystals is explained by the fact that “the number of electrons (atomic number) in the atom determines the amount by which an atom scatters X-rays”. The atomic numbers of potassium (19) and chlorine (17) are not very much different. If we assume that all the atoms are identical, it means that the face-centred arrangement of  $NaCl$  (Fig. 25.42) becomes a simple cubic arrangement for  $KCl$ . But, in the case of  $NaCl$  crystal, there is difference between the atomic numbers of  $Na$  (11) and  $Cl$  (17). It means that their scattering powers are different and hence give rise to a face-centred cubic lattice. From the above discussion, we can say that “both  $NaCl$  and  $KCl$  have the same structure”.

**5. Diamond Structure.** The diamond structure is a fcc structure with a basis of two carbon atoms, one located at  $(0, 0, 0)$  and other at  $(a/4, a/4, a/4)$  associated with each lattice point. Fig. 25.43 shows the positions of atoms in the cubic cell of the diamond structure projected on a cubic face. The fractions denote height above the base in units of a cube edge. The points at 0 and  $\frac{1}{2}$  are on the fcc lattice while those at  $1/4$  and  $3/4$  are on a similar lattice displaced along the body diagonal by one-fourth of its length. Thus the diamond lattice is composed of two inter-penetrating fcc sub-lattices, one of which is shifted relative to the other by one-fourth of a body diagonal.

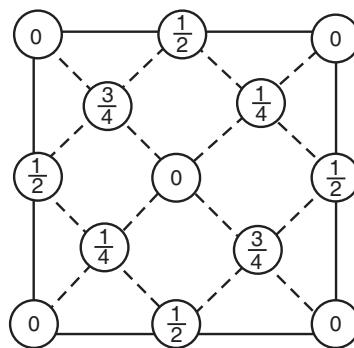


Fig. 25.43

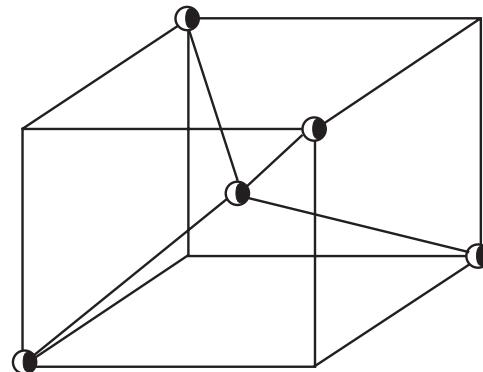


Fig. 25.44

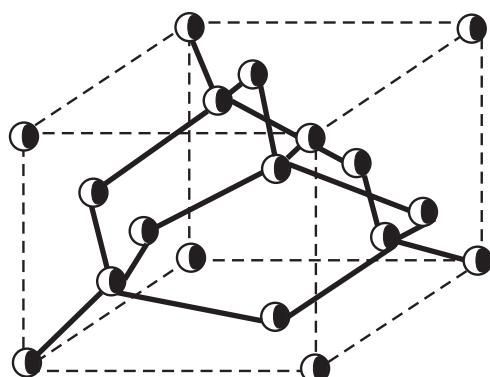


Fig. 25.45

In a diamond crystal, the carbon atoms are linked by directional covalent bonds. Each carbon atom forms covalent bonds with four other carbon atoms that occupy four corners of a cube in a tetrahedral structure (Fig. 25.44). The length of each bond is 0.154 nm and the angle between the bonds is  $109.5^\circ$ . The entire diamond lattice is constructed of such tetrahedral units (Fig. 25.45).

In the diamond lattice, each atom has four nearest neighbours with which it forms covalent bonds. Thus the coordination number of diamond crystal is 4. The number of atoms per unit cell is 8. The diamond structure is loosely packed since each atom has only 4 nearest neighbours. Carbon, silicon, germanium and grey tin crystallize in diamond structure.

## 25.14 GLASS

Generally a *glass* is an amorphous, brittle and transparent solid. It is not metallic. It does not exhibit ferromagnetism. A glass has the elastic properties of an isotropic solid. A glass has the random structure of the liquid from which it is derived by cooling below the freezing point, without crystallization. Fig. 25.46 illustrates the difference between (a) a crystal and (b) a glass.

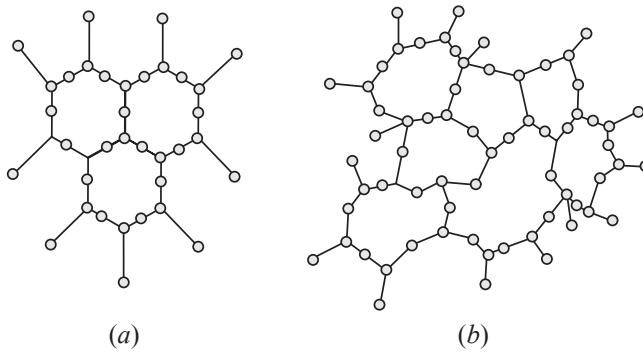


Fig. 25.46

In certain cases, the solid state may correspond to a supercooled liquid in which the molecular arrangement of the liquid state is frozen. Because of rapid cooling and a high viscosity of the liquid, crystals may not have had time to grow and a glassy material results [Fig. 25.46(b)]. Upon annealing, such glassy substances may crystallize (devitrify).

The following Table illustrates the characteristics of glasses and compares these to those of metals.

<i>Property</i>	<i>Glass</i>	<i>Metal</i>
<b>Structure</b>	amorphous	crystalline
<b>Bonding</b>	covalent	metallic
<b>Yield-stress</b>	almost ideal	non-ideal
<b>Workability</b>	poor, brittle	good, ductile
<b>Hardness</b>	very high	low to high
<b>Optical transmission</b>	transparent	opaque
<b>Thermal conductivity</b>	poor	very good
<b>Resistance</b>	high	very low
<b>Corrosion resistance</b>	very good	poor to good
<b>Magnetic properties</b>	non-existent	various

### Glass Transition Temperature

*The glass transition temperature is the temperature at which the liquid like atomic configuration can be frozen into a solid.*

A liquid on being cooled becomes a glass when the viscosity equals  $10^{13}$  poise. This defines the *glass transition temperature*  $T_g$ . At temperatures above  $T_g$  we have a liquid. At temperatures below  $T_g$  we have a glass. The transition is not a thermodynamic phase transition, only a transition for practical purposes.

### Metallic Glasses

Metallic glasses are metal alloys which have no long range atomic order. To achieve this structure, a suitable alloy must be quenched so rapidly that the random liquid like structure is retained in the solid.

### Properties

- (i) Metallic glasses share the properties of both metals and glasses. They are amorphous alloys with an atomic configuration similar to that of the molten liquid, *i.e.*, there is no translational symmetry.
- (ii) Metallic glasses are all strong, ductile, malleable, opaque and brittle.
- (iii) A metallic glass has conductivity similar in magnitude to that of solid and liquid metals and optical behaviour typical of metals.

(iv) A metallic glass is homogeneous, non-crystalline, but disordered structure. It is densely packed as in a compact crystalline structure.

### Applications

(i) Even though metallic glasses are non-crystalline, they are ferromagnetic. They possess low magnetic losses, high permeability and saturation magnetisation with low coercivity. So they are used in tape-recorder heads, cores of high power transformers and magnetic shields.

**EXAMPLE.**  $\text{Fe}_{75}\text{P}_{15}\text{C}_{10}$ ,  $\text{Fe}_{24}\text{Zr}_{76}$ ,  $\text{Ni}_{60}\text{Nb}_{40}$ .

(ii) Metallic glasses possess high tensile strength and hardness. For example, the tensile strength of  $\text{Fe}_{80}\text{B}_{20}$  is 3.6 GPa. For  $\text{Fe}_{40}\text{Mo}_{40}\text{B}_{20}$ , the hardness (HV) is  $1950 \text{ kg/mm}^2$ . So they are used as reinforcing elements in concrete, plastic or rubber.

(iii) The electrical resistance of metallic glasses is high and does not change appreciably with temperature. So they are used to make accurate standard resistances, computer memories and magneto resistance sensors.

In metallic glasses, there is a sharp variation in resistance at very low temperatures. So they are used as low temperature thermometers having resistivity of about  $60 \mu\text{V/K}$ .

(iv) Many metallic glasses based on titanium, niobium, lanthanum and molybdenum are superconductors. So they are used in producing high magnetic fields for fusion reactors and for magnetic levitation of trains.

(v) Chromium and Phosphorus based metallic glasses have high corrosion resistance. So they are used in marine cables, chemical filters, inner surfaces of reactor vessels, orthopaedical implants and surgical clips.

### 25.15 QUASICRYSTALS

We have learned that a necessary condition for having a crystalline phase is lattice periodicity and a crystal cannot have a five-fold rotational symmetry. But there are physical systems with many properties of the usual crystalline state but without three-dimensional lattice periodicity like  $\text{Cu}_{0.2}\text{Al}_{0.65}\text{Fe}_{0.15}$ . These systems do not have a strict crystal-like periodicity, but have only some subtle type of quasi/translational periodicity and a long-range order, but with five-fold symmetry axes. Such quasiperiodic crystals having many properties similar to usual familiar crystals are termed as *quasicrystals*. The full point symmetry group of the experimentally observed diffraction patterns of these five-fold symmetry crystals correspond to a platonic solid called an *icosahedron* (Fig. 25.47). The icosahedron has six five-fold symmetry axes, 10 three-fold and 15 two-fold symmetry axes.

Quasiperiodicity was first observed in grains of size  $2 \mu\text{m}$  in the alloy of Al with 14 atomic per cent Mn. The smaller Mn atoms are each surrounded by 12 Al atoms arranged at the corners of an icosahedron. The structure is made up of parallel icosahedra attached at their edges. As a result of tiling in two dimensions, we get the structure known as *Penrose patterns* (Fig. 25.48). There is no long-range translationally

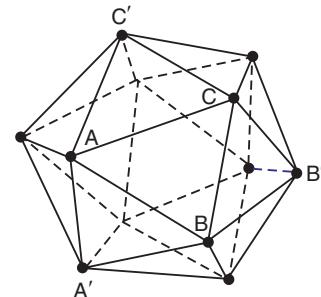


Fig. 25.47

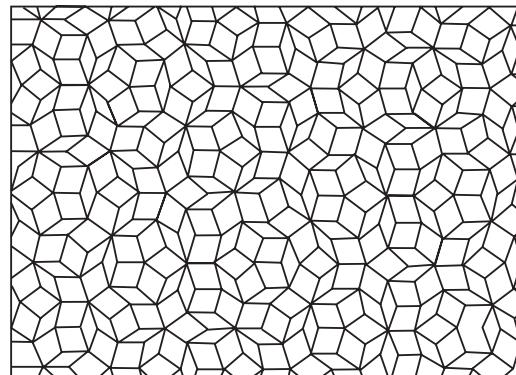


Fig. 25.48

periodic order, but there is a long-range orientational order: all of the space cannot be filled by repeating the basic unit, but a crystal can be constructed by nucleation at a centre cell, followed by an outward growth from there; thus overall, the crystal could not exhibit the five-fold symmetry of an icosahedron but the X-ray diffraction pattern could give a five-fold symmetry.

### EXPERIMENTAL METHODS IN X-RAY DIFFRACTION

#### 25.16 ROTATING CRYSTAL METHOD

This method is widely used for the determination of the size of the unit cell of a crystal.

**Apparatus.** Fig. 25.49 shows the experimental arrangement.

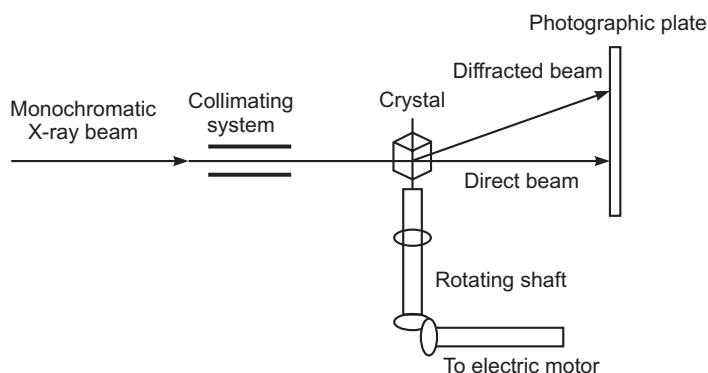


Fig. 25.49

- X-rays are generated in the X-ray tube. The beam is made monochromatic by using a filter. The beam is then passed through the collimating system which gives a fine pencil of parallel X-rays.
- A small single crystal (of  $\sim 1$  mm dimensions) is mounted on a shaft arranged perpendicular to the incident beam. The shaft is rotated by a small electric motor.
- The crystal is positioned at the centre of a cylindrical holder concentric with the rotating shaft [Fig. 25.50]
- A photographic film is attached at the inner circular surface of the cylinder.

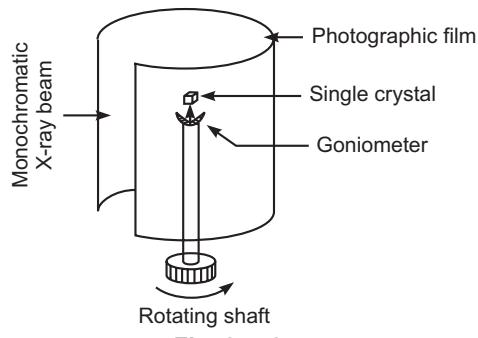


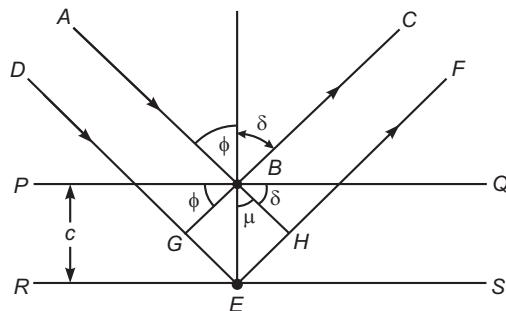
Fig. 25.50

#### Experimental procedure

When the crystal is set into rotation about a fixed axis, the sets of atomic planes come successively into their reflecting positions. The diffracted beams from all planes parallel to the vertical rotation axis lie in the horizontal plane. The diffracted beams from planes with other orientations will lie in layers above and below the horizontal plane. The diffraction pattern is registered on a film of cylindrical camera, the axis of which coincides with the axis of rotation of the crystal.

#### Theory of rotation photograph

Assume that the crystal is mounted in such a way that one of the crystallographic axes coincides with the axis of rotation of the crystal, say  $z$ -axis.  $B$  and  $E$  are two neighbouring atoms at the lattice points along the  $z$ -axis (Fig. 25.51).



**Fig. 25.51.** Diffraction of X-rays by lattice elements along the rotation axis.

- The distance between  $B$  and  $E$  is  $c$ , the primitive translation.
- The angle of incidence and the angle of reflection are  $\phi$  and  $\delta$  respectively.
- $G$  is the foot of perpendicular drawn from  $B$  on the incident ray  $DE$ .
- $H$  is the foot of perpendicular drawn from  $B$  on the reflected ray  $EF$ .
- $\angle EBH = (90 - \delta) = \mu$ .

The path difference between the two interfering rays is

$$GE + EH = BE \sin (90 - \phi) + BE \sin (90 - \delta) = c (\cos \phi + \cos \delta)$$

The condition of diffraction is given by

$$c (\cos \delta + \cos \phi) = n \lambda \quad \dots(1)$$

If the rotation axis is perpendicular to the beam, then  $\phi = 90^\circ$  and Eq. (1) reduces to

$$c \cos \delta = c \cos (90 - \mu) = c \sin \mu = n \lambda \quad \dots(2)$$

For various values of  $n = 0, 1, 2, 3, \dots$ , we get the series of equations:

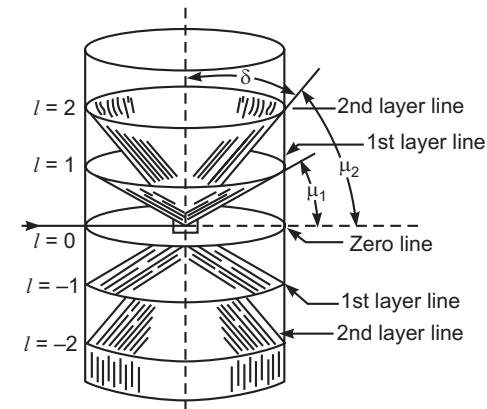
$$\begin{aligned} \cos \delta_0 &= \sin \mu_0 = 0; \quad \cos \delta_1 = \sin \mu_1 = \frac{\lambda}{c}; \\ \cos \delta_2 &= \sin \mu_2 = \frac{2\lambda}{c}; \quad \dots \text{etc.} \\ \cos \delta_n &= \sin \mu_n = \frac{n\lambda}{c} \end{aligned} \quad \dots(3)$$

#### Formation of layer lines on a cylindrical film

When  $\lambda$  is constant, such equations give the loci of all possible diffracted rays as the crystal turns. These loci are elements of a series of cones, of which the half apex angles are  $\delta_0, \delta_1, \delta_2, \delta_3, \dots$  etc. (Fig. 25.52). Any element of each of the cones makes an angle  $\mu_0, \mu_1, \mu_2, \text{etc.}$ , respectively with the horizontal plane.

The central horizontal plane contains all diffracted beams of order of zero. The Miller indices of all planes giving diffracted beams in horizontal plane must be represented by  $(h, k, 0)$ . Similarly, X-rays diffracted by planes of indices,  $(h, k, l)$  lie on the first order cone defined by

$$\mu_l = \sin^{-1} \left( \frac{\lambda}{c} \right) \quad \dots(4)$$



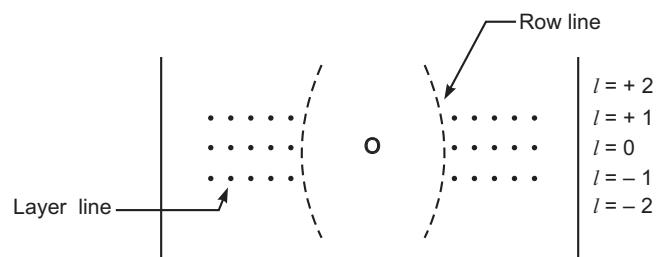
**Fig. 25.52**

In general, X-rays diffracted by planes of indices  $(hkn)$  lie on the  $n$ th order cone defined by

$$\mu_n = \sin^{-1} \left( \frac{n\lambda}{c} \right) \quad \dots(5)$$

Fig. 25.52 shows that the cones intersect the cylindrical film in a series of circles lying in planes perpendicular to the axis of rotation which becomes a series of parallel straight lines, known as **layer lines**, when the film is unrolled.

Fig. 25.53 shows *layer lines produced after flattening the photographic film*.



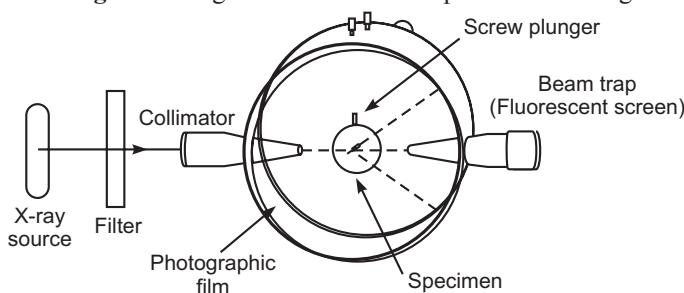
**Fig. 25.53**

- The cone for  $n = 0$  is a plane perpendicular to the axis and containing the direction of incidence and the intersection of this direction with the film is known as *equatorial or zero layer line*. The succeeding lines are termed first, second, third, etc., layer lines.
- Diffraction images from planes with the same values of  $h$  and  $k$  but different values of  $l$  all lie on one of a series of curves called *row lines*, which are transverse to the layer lines.

The dimensions of the unit cell of the structure are determined in the following way: The spacing of layer lines give lattice translations. If the distance of the film from the crystal is known, the distance of the layer lines from the equatorial line gives the values of angle  $\delta$  and as  $\lambda$  is known, the value of  $c$  can be calculated. Now the rotational photographs are taken separately about all three axes. In this way, lattice translations  $a$ ,  $b$ ,  $c$  are calculated which give the dimensions of the unit cell of the crystal.

### 25.17 POWDER METHOD (DEBYE-SCHERRER METHOD)

**Experimental arrangement:** Fig. 25.54 shows the experimental arrangement.



**Fig. 25.54**

- The finely powdered sample is filled in a thin capillary tube.
- The specimen is mounted on the holder at the centre of the cylindrical camera. The specimen is aligned using a screw plunger so that the axis of rotation of the specimen coincides with the axis of the camera. This ensures that the specimen is always within the X-ray beam profile during rotation.

- A collimated beam of monochromatic x-rays is produced by passing the x-rays through a filter and a collimator. The x-rays enter the camera through a fine hole in the collimator and fall on the polycrystalline rotating specimen. The direct beam being intense, enters a beam trap after passing through the specimen.
- The diffracted rays go out from individual crystallites which happen to be oriented with planes making a glancing angle  $\theta$  with the beam satisfying Bragg's equation  $2d \sin \theta = n\lambda$ . The diffracted rays due to any one  $hkl$  plane leave the specimen along the surface of a cone concentric with the original beam with a semi-apex angle  $2\theta$  because of the spherical symmetry of the specimen (Fig. 25.55). In a similar way, other lattice planes generate their diffraction cones.
- These cones intercept the film strip, kept cylindrically along the inner wall of the camera, in a series of concentric rings (a pair of arcs).
- When the Bragg angle is  $45^\circ$ , the corresponding cone opens out into a circular disc which gives rise to a straight line at the midpoint of the film strip. When the Bragg angle is more than  $45^\circ$ , back reflections are produced. Thus, by using the cylindrical film, the Bragg angle from zero up to a maximum of  $90^\circ$  can be obtained.

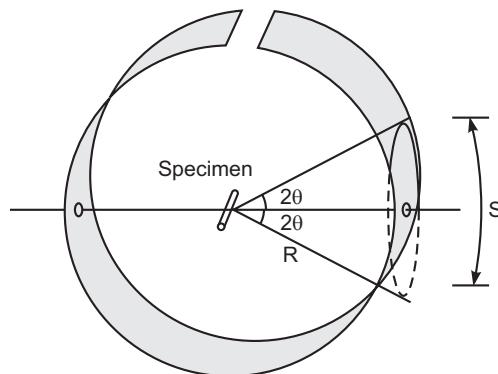


Fig. 25.55

### Experimental procedure

- The sample is prepared, mounted on the sample holder and centred.
- The camera with the sample is taken to the dark room. Two holes are punched on the film. The film is loaded without disturbing the sample. The camera is closed, brought out and exposed to  $X$ -rays. During exposure, the sample is continuously rotated by connecting to a motor attachment. On completion of the exposure time (say 2 hrs for NaCl), the camera is taken to the dark room, the film is taken out, processed (developed and fixed) and dried for indexing.

Fig. 25.56 shows flattened photographic film after developing and indexing of diffraction lines.

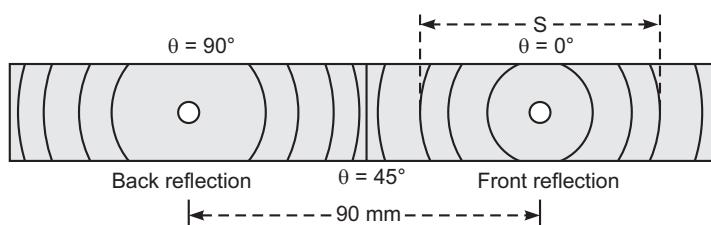


Fig. 25.56

### Interpretation of the lines

- It is necessary to know the positions of low and high angle ( $\theta = 0$  and  $\theta = 90^\circ$ ) diffraction lines on the film before any measurements are made. This can be done by observing the background intensity on the film. The background intensity is maximum near  $\theta = 0$ . This arises due to the scattering of the  $X$ -rays by air molecules present inside the camera. Thus one expects more blackening of the film near the low angle side, i.e., around the collimator.

- The film is placed over a comparator (film reader). The telescope is focussed on the lines. The linear distance ( $S$ ) between different pairs of arcs is determined accurately. The radius ( $R$ ) of the camera is known.

The angle  $\theta$  corresponding to a particular pair of arcs is related to the distance  $S$  between the arcs as

$$4\theta \text{ (radians)} = S/R \quad \dots(1)$$

If  $\theta$  is measured in degrees, Eq. (1) is modified as

$$4\theta \text{ (degrees)} = \frac{57.3 S}{R} \quad \dots(2)$$

or  $\theta = \frac{57.3 S}{4R}$  degrees  $\dots(3)$

The various diffraction angles ( $\theta$ ) can be calculated using Eq. (3).

Knowing the Bragg angle  $\theta$  for different lines, the interplanar spacings ( $d$  values) can be calculated using Bragg equation, i.e.,  $2d \sin \theta = n\lambda$ .

These  $d$  values are used to determine the space lattice of the crystal structure.

### 25.18 X-RAY DIFFRACTION: LAUE EQUATIONS

Let us examine X-ray scattering by two identical scattering centres in the crystal separated by a distance  $\mathbf{r}$ . Let  $\hat{\mathbf{n}}_0$  be defined as unit vector in the direction of the incident beam and  $\hat{\mathbf{n}}_1$  a unit vector in an arbitrary scattered direction (Fig. 25.57). The incident radiation is assumed to be a parallel beam.

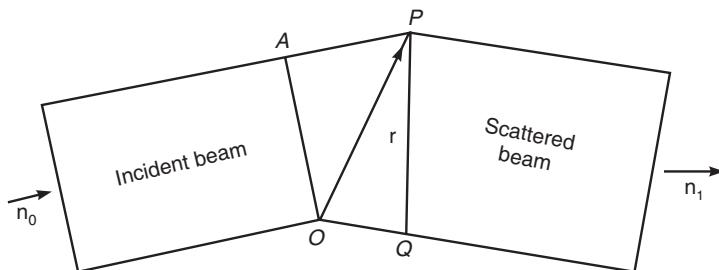


Fig. 25.57

The scattered beam is assumed to be detected at a very distant observation point.

The path difference between the waves scattered from the points  $P$  and  $O$  is

$$PA - OQ = \mathbf{r} \cdot \hat{\mathbf{n}}_0 - \mathbf{r} \cdot \hat{\mathbf{n}}_1 = \mathbf{r} \cdot (\hat{\mathbf{n}}_0 - \hat{\mathbf{n}}_1) = \mathbf{r} \cdot \mathbf{N} \quad \dots(1)$$

Here,  $\mathbf{N} = \hat{\mathbf{n}}_0 - \hat{\mathbf{n}}_1$ . The vector  $\mathbf{N}$  is the normal to Bragg reflecting plane whose magnitude is given by

$$|\mathbf{N}| = 2 \sin \theta \quad (\text{From Fig. 25.58}) \quad \dots(2)$$

The phase difference  $\phi_r$  between the radiation scattered at the two points is

$$\begin{aligned} \phi_r &= \frac{2\pi}{\lambda} (\text{Path difference}) \\ &= \frac{2\pi}{\lambda} (\mathbf{r} \cdot \mathbf{N}) \end{aligned} \quad \dots(3)$$

The radiations scattered by atoms  $P$  and  $O$  will interfere constructively only if the phase difference becomes an integral multiple of  $2\pi$ . Due to periodicity of the crystal, the other atoms placed in the same direction would also scatter the radiations in phase with those scattered from  $P$  and  $O$ . In a three-dimensional crystal,  $\mathbf{r}$  may coincide with any of the three crystallographic axes  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$ . Thus for the occurrence of a diffraction maximum, the following three conditions must be satisfied simultaneously:

$$\phi_a = \frac{2\pi}{\lambda}(\mathbf{a} \cdot \mathbf{N}) = 2\pi h' = 2\pi nh$$

$$\phi_b = \frac{2\pi}{\lambda}(\mathbf{b} \cdot \mathbf{N}) = 2\pi k' = 2\pi nk$$

$$\phi_c = \frac{2\pi}{\lambda}(\mathbf{c} \cdot \mathbf{N}) = 2\pi l' = 2\pi nl$$

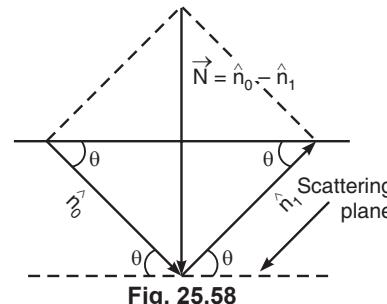


Fig. 25.58

...(4)

Here,  $h'$ ,  $k'$ ,  $l'$  are any integers.  $h' = nh$ ,  $k' = nk$ ,  $l' = nl$ , where  $h$ ,  $k$ , and  $l$  are three integers.

Let  $\alpha$ ,  $\beta$ ,  $\gamma$  be the angles between the scattering normal  $\mathbf{N}$  and the crystallographic axes  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$  respectively. Therefore, Eqs. (4) become

$$\mathbf{a} \cdot \mathbf{N} = 2a \sin \theta \cos \alpha = h' \lambda = nh \lambda$$

$$\mathbf{b} \cdot \mathbf{N} = 2b \sin \theta \cos \beta = k' \lambda = nk \lambda$$

$$\mathbf{c} \cdot \mathbf{N} = 2c \sin \theta \cos \gamma = l' \lambda = nl \lambda \quad ... (5)$$

These equations are called *Laue equations*.

- The conditions in Eq. (5) are called the *van Laue's equations*. When three cones each representing simultaneous scattering from each array intersect, it results in only one allowed direction for diffracted beam.

The three equations of (5) serve to determine a unique value for  $\theta$  and  $\mathbf{N}$ , thus defining a scattering direction. The direction cosines of the scattering normal  $\mathbf{N}$  are seen from Eq. (5) to be proportional to  $\frac{h}{a}$ ,  $\frac{k}{b}$  and  $\frac{l}{c}$ . Neighbouring planes whose Miller indices are  $(hkl)$  intersect the  $a$ ,  $b$ ,  $c$  axes at intervals  $\frac{a}{h}$ ,  $\frac{b}{k}$ ,  $\frac{c}{l}$ . The direction cosines of the normal to the  $(h, k, l)$  family of planes are

therefore also proportional to  $\frac{h}{a}$ ,  $\frac{k}{b}$  and  $\frac{l}{c}$ . The scattering normal  $\mathbf{N}$  is thus identical to the normal to the  $(h, k, l)$  planes and the  $(h, k, l)$  planes may be regarded as the reflecting planes of the Bragg picture.

If we put

$$h = \frac{a \cos \alpha}{d}, \quad k = \frac{b \cos \beta}{d}, \quad l = \frac{c \cos \gamma}{d}$$

each of these Laue equations reduces to Bragg's equation

$$2d \sin \theta = n\lambda.$$

Here  $d$  is the distance between adjacent planes of the  $(hkl)$  system.

- Thus we find that the three Laue conditions are directly equivalent to the Bragg reflection condition.

### The Laue Method

**Principle.** A single crystal is held stationary in a beam of X-rays of continuous wavelength. The crystal selects and diffracts the discrete values of  $\lambda$  for which planes exist of spacing  $d$  and incidence angle  $\theta$  satisfying the Bragg law.

**Experimental Arrangement.** Fig. 25.59 shows the experimental arrangement. A source is used that produces a beam of X-rays over a wide range of wavelengths preferably from 0.02 nm to 0.2 nm. A pinhole arrangement produces a well-collimated beam. The dimensions of the crystal are usually less than 1 mm. Flat films are placed in front of and behind the specimen.

**Working.** Since  $\lambda$  covers a continuous range, the crystal selects that particular wavelength which satisfies Bragg's law at the present orientation. A diffracted beam emerges at the corresponding angle. The diffracted beam is then recorded as a spot on the film. But since the wavelength corresponding to a spot is not measured, one cannot determine the actual values of the interplanar spacings—only their ratios. Therefore one can determine the shape but not the absolute size of the unit cell.

The pattern will show the symmetry of the crystal : if a crystal has a fourfold axis of symmetry parallel to the beam, the Laue pattern will show fourfold symmetry.

### Advantages

- (1) The Laue method is widely used to orient crystals for solid state experiments.
- (2) The shape of the unit cell can be established from the symmetry of the pattern.

### Disadvantage

This method is, however, not suitable for determining the crystal structure. This is because, out of a continuous range of wavelengths, several wavelengths are reflected in different orders from a single plane, so that different orders of reflection may overlap on a single spot. This makes difficult the measurement of the reflection intensity of individual spots and hence of the missing reflections. The missing reflections would have been an important step in the determination of the crystal structure.

## RECIPROCAL LATTICE

### 25.19 RECIPROCAL LATTICE

In a crystal lattice there exist many sets of planes, with different orientations and spacings, which can cause diffraction. Fig. 25.60 shows three sets of lattice planes with interplanar spacings  $d_1$ ,  $d_2$  and  $d_3$ . Let us draw, from a common origin, normals to all sets of planes, the length of each normal being proportional to the reciprocal of the interplanar spacing of the corresponding set. Then, the end-points of the normals form a lattice which is called the '*reciprocal lattice*'.

Each point in the reciprocal lattice preserves the characteristics of the set of planes which it represents. Its direction with respect to the origin represents the orientation of the planes, and its distance from the origin represents the interplanar spacing of the planes.

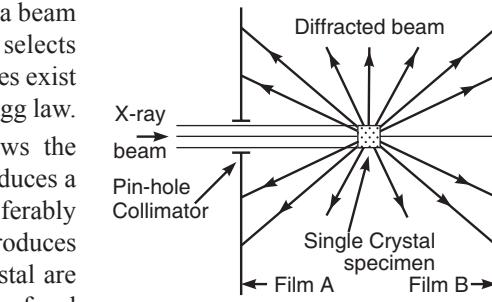


Fig. 25.59

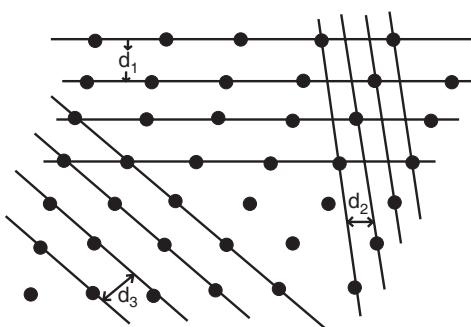


Fig. 25.60

**Definition**

1. The reciprocal space lattice is a set of imaginary points constructed in such a way that the direction of a vector from one point to another coincides with the direction of a normal to the space planes and the separation of those points (absolute value of the vector) is equal to the reciprocal of the real interplanar distance.

2. Consider a set of points  $\mathbf{R}$  constituting a Bravais lattice, and a plane wave,  $e^{i\mathbf{k}\cdot\mathbf{r}}$ . For general  $\mathbf{k}$ , such a plane wave will not have the periodicity of the Bravais lattice, but for certain special choices of wave vector it will. The set of all wave vectors  $\mathbf{K}$  that yield plane waves with the periodicity of a given Bravais lattice is known as its reciprocal lattice. Analytically,  $\mathbf{K}$  belongs to the reciprocal lattice of a Bravais lattice of points  $\mathbf{R}$ , provided that the relation

$$e^{i\mathbf{K}\cdot(\mathbf{r} + \mathbf{R})} = e^{i\mathbf{K}\cdot\mathbf{r}} \quad \dots(1)$$

holds for any  $\mathbf{r}$ , and for all  $\mathbf{R}$  in the Bravais lattice. Factoring out  $e^{i\mathbf{K}\cdot\mathbf{r}}$ , we can characterize the reciprocal lattice as the set of wave vectors  $\mathbf{K}$  satisfying

$$e^{i\mathbf{K}\cdot\mathbf{R}} = 1 \quad \dots(2)$$

for all  $\mathbf{R}$  in the Bravais lattice.

Note that a reciprocal lattice is defined with reference to a particular Bravais lattice. The Bravais lattice that determines a given reciprocal lattice is often referred to as the *direct lattice*, when viewed in relation to its reciprocal. Note also that although one could define a set of vectors  $\mathbf{K}$  satisfying (2) for an arbitrary set of vectors  $\mathbf{R}$ , such a set of  $\mathbf{K}$  is called a reciprocal lattice only if the set of vectors  $\mathbf{R}$  is a Bravais lattice.

**25.19.1. Reciprocal Lattice Vectors**

Let  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$  be the primitive vectors of the (direct) crystal lattice. Then, primitive vectors  $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$  of the reciprocal lattice are defined by

$$\mathbf{b}_1 = 2\pi \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3}; \mathbf{b}_2 = 2\pi \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3}; \mathbf{b}_3 = 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3}. \quad \dots(1)$$

Each vector defined by Eq. (1) is orthogonal to two axis vectors of the crystal lattice. Thus  $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$  have the property

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij} \quad \dots(2)$$

Here,  $\delta_{ij} = 1$  if  $i = j$  and  $\delta_{ij} = 0$  if  $i \neq j$ .

Points in the reciprocal lattice are mapped by the set of vectors

$$\mathbf{G} = v_1 \mathbf{b}_1 + v_2 \mathbf{b}_2 + v_3 \mathbf{b}_3. \quad \dots(3)$$

Here,  $v_1, v_2, v_3$  are integers. A vector  $\mathbf{G}$  of this form is a *reciprocal lattice vector*.

It connects any two points of the reciprocal lattice.

**25.19.2. Properties of Reciprocal Lattice**

- Every crystal structure has two lattices associated with it, the crystal lattice and the reciprocal lattice. A diffraction pattern of a crystal is a map of the reciprocal lattice of the crystal. A microscope image is a map of the crystal structure in real space. Thus when we rotate a crystal in a holder, we rotate both the direct lattice and the reciprocal lattice.
- Vectors in the direct lattice have the dimensions of [length]. Vectors in the reciprocal lattice have the dimensions of length<sup>-1</sup>.
- A crystal lattice is a lattice in real space. The reciprocal lattice is a lattice in the Fourier space associated with the crystal.

4. Each point in a reciprocal lattice corresponds to particular set of parallel planes of the direct lattice.
5. The distance of a reciprocal lattice point from an arbitrarily fixed origin is inversely proportional to the interplanar spacing of the corresponding parallel planes of the direct lattice.
6. The volume of a unit cell of the reciprocal lattice is inversely proportional to the volume of a unit cell of the direct lattice.
7. The unit cell of the reciprocal lattice need not be a parallelopiped. In fact, we almost always deal with the Wigner-Seitz cell of the reciprocal lattice. This is called a *Brillouin zone*.
8. The direct lattice is the reciprocal of its own reciprocal lattice.
9. Simple cubic lattice is its own reciprocal. bcc and fcc lattices are reciprocal to each other.

**Importance.** A diffraction pattern of a crystal is a map of the reciprocal lattice of the crystal. The Bragg's diffraction law can be expressed in terms of the reciprocal lattice vector  $\mathbf{G}$ , as

$$2 \mathbf{k} \cdot \mathbf{G} + G^2 = 0.$$

Here,  $\mathbf{k}$  is the wave vector of the incident wave. This equation is very important in wave propagation and elastic scattering in crystal lattices. A similar result arises in the theory of the electron energy band structure of crystals. Thus the concept of reciprocal lattice is very important in the study of lattice dynamics of solids.

### 25.19.3. Bragg's Law for Reciprocal Lattice

The Ewald construction for diffraction of X-rays is shown in Fig. 25.61. Draw a vector  $\mathbf{AO}$  in the incident direction of length  $1/\lambda$  terminating at the origin  $O$ .

$$AO = AB = \frac{1}{\lambda} \quad AE = \text{Reflecting plane}$$

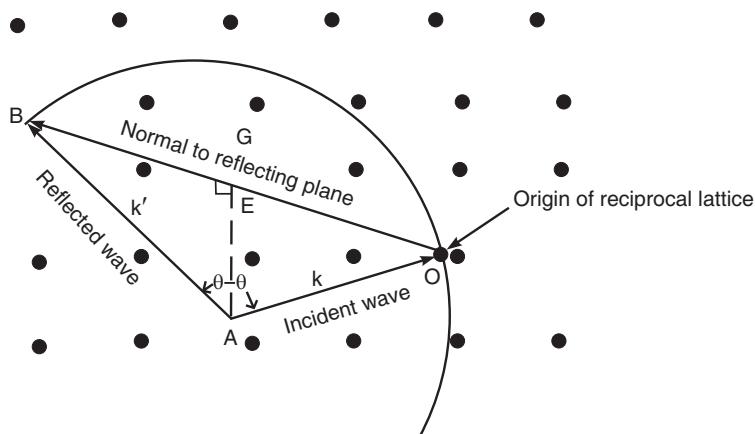


Fig. 25.61

Construct a circle of radius  $1/\lambda$  (a sphere, called *reflex sphere*, of radius  $1/\lambda$  in three dimensions) with centre at  $A$ . The circle passes through any point  $B$  of the reciprocal lattice. Join  $A$  and  $O$  to  $B$ . Thus,  $\mathbf{OB}$  is a reciprocal lattice vector,  $\mathbf{G}$ .  $\mathbf{OB}$  is normal to the lattice plane  $AE$ .

Hence,  $OB = |\mathbf{G}| = 1/d$ , where  $d$  is the interplanar spacing. The wave vectors of the incoming and outgoing beams are  $\mathbf{k} = \mathbf{AO}$  and  $\mathbf{k}' = \mathbf{AB}$ . Thus,

$$\mathbf{k}' = \mathbf{k} + \mathbf{G} \quad \dots(1)$$

For diffraction, it is necessary that the vector  $\mathbf{k}'$  be equal in magnitude to the vector  $\mathbf{k}$ :

$$\begin{aligned} \mathbf{k}'^2 &= (\mathbf{k} + \mathbf{G})^2 = \mathbf{k}^2 \\ \text{or } 2\mathbf{k} \cdot \mathbf{G} + \mathbf{G}^2 &= 0 \end{aligned} \quad \dots(2)$$

This is the Bragg's law in vector form.

## 25.20 BRILLOUIN ZONES

A Brillouin zone is defined as a Wigner-Seitz primitive cell in the 'reciprocal' lattice.

The first Brillouin zone of an oblique lattice in two dimensions is constructed in Fig. 25.62. We first draw an adequate number of vectors from  $O$  to nearby points in the reciprocal lattice. Next we draw lines perpendicular to these vectors at their midpoints. The smallest enclosed area is the first Brillouin zone. Any wave whose wave vector drawn from the origin terminates on the boundary of the zone will be diffracted by the crystal. The first Brillouin zone is thus the smallest volume entirely enclosed by planes which are the perpendicular bisectors of the reciprocal lattice vectors drawn from the origin.

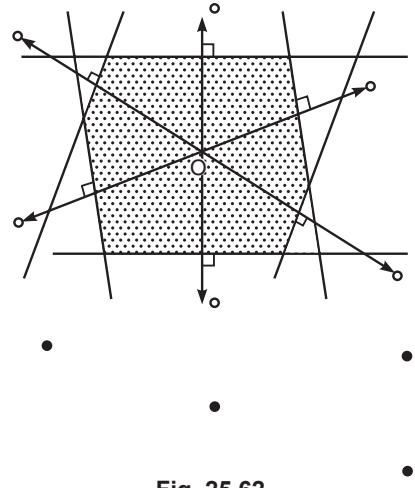


Fig. 25.62

## 25.21 RECIPROCAL LATTICE TO SIMPLE CUBIC (SC) LATTICE

The primitive translation vectors of a simple cubic lattice may be written as

$$\mathbf{a}_1 = a\hat{\mathbf{x}}; \quad \mathbf{a}_2 = a\hat{\mathbf{y}}; \quad \mathbf{a}_3 = a\hat{\mathbf{z}} \quad \dots(1)$$

Here  $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$  are orthogonal vectors of unit length.

The volume of the unit cell

$$= \mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3 = a^3. \quad \dots(2)$$

The primitive translation vectors of the reciprocal lattice are

$$\mathbf{b}_1 = (2\pi/a)\hat{\mathbf{x}}; \quad \mathbf{b}_2 = (2\pi/a)\hat{\mathbf{y}}; \quad \mathbf{b}_3 = (2\pi/a)\hat{\mathbf{z}} \quad \dots(3)$$

Eqs. (3) indicates that all the three reciprocal lattice vectors are equal in magnitude. Thus the reciprocal lattice is itself a simple cubic lattice, now of lattice constant  $2\pi/a$  (Fig. 25.63).

The boundaries of the first Brillouin zones are the planes normal to the six reciprocal lattice vectors  $\pm \mathbf{b}_1, \pm \mathbf{b}_2, \pm \mathbf{b}_3$ , at their midpoints.

$$\pm \frac{1}{2}\mathbf{b}_1 = \pm(\pi/a)\hat{\mathbf{x}}; \quad \pm \frac{1}{2}\mathbf{b}_2 = \pm(\pi/a)\hat{\mathbf{y}}; \quad \pm \frac{1}{2}\mathbf{b}_3 = \pm(\pi/a)\hat{\mathbf{z}}.$$

The six planes bound a cube of edge  $2\pi/a$  and of volume  $(2\pi/a)^3$ . This cube is the first Brillouin zone of the sc crystal lattice.

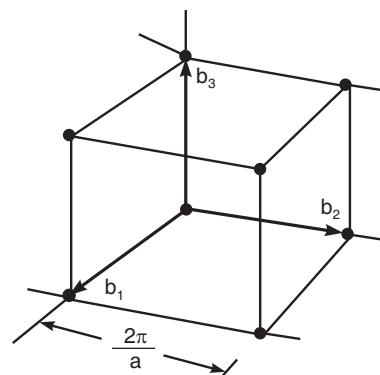


Fig. 25.63

## 25.22 RECIPROCAL LATTICE TO BCC LATTICE

The primitive translation vectors of the bcc lattice (Fig. 25.64) are

$$\mathbf{a}_1 = \frac{1}{2}a(-\hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}}); \quad \mathbf{a}_2 = \frac{1}{2}a(\hat{\mathbf{x}} - \hat{\mathbf{y}} + \hat{\mathbf{z}}); \quad \mathbf{a}_3 = \frac{1}{2}a(\hat{\mathbf{x}} + \hat{\mathbf{y}} - \hat{\mathbf{z}}) \quad \dots(1)$$

Here,  $a$  is the side of the cube edge.

$\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$  are orthogonal unit vectors parallel to the cube edges.

The volume of the primitive cell is

$$V = |\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3| = \frac{1}{2}a^3 \quad \dots(2)$$

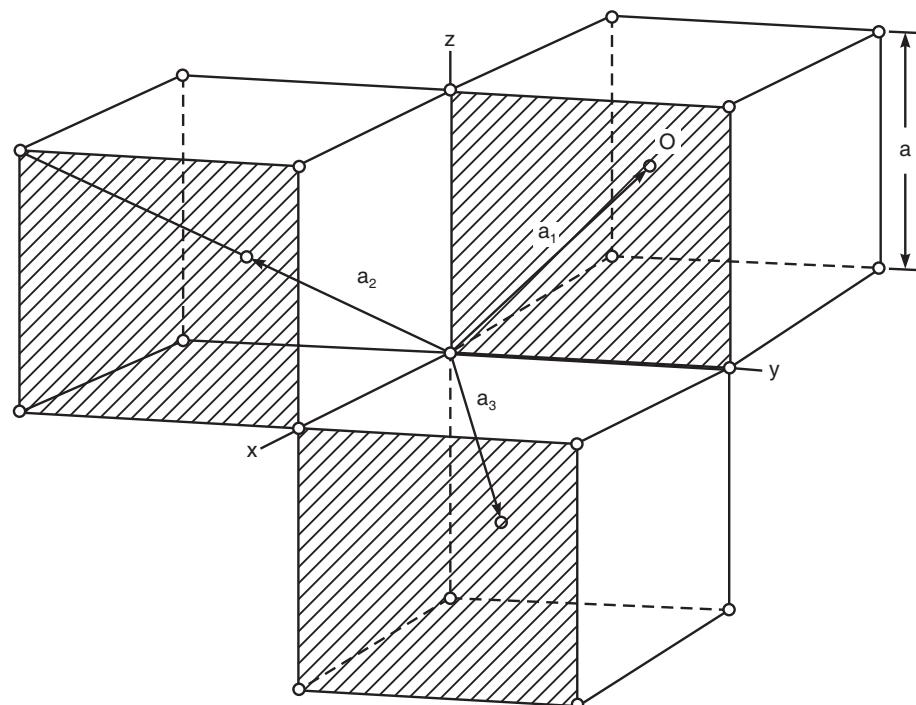


Fig. 25.64

The primitive translation vectors of the reciprocal lattice are

$$\mathbf{b}_1 = (2\pi/a)(\hat{\mathbf{y}} + \hat{\mathbf{z}}); \quad \mathbf{b}_2 = (2\pi/a)(\hat{\mathbf{x}} + \hat{\mathbf{z}}); \quad \mathbf{b}_3 = (2\pi/a)(\hat{\mathbf{x}} + \hat{\mathbf{y}}).$$

These are the primitive translation vectors of an fcc lattice.

Thus the reciprocal lattice to a bcc lattice is an fcc lattice.

Let  $v_1, v_2, v_3$ , be integers.

The general reciprocal lattice vector is,

$$\mathbf{G} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + v_3\mathbf{b}_3 = (2\pi/a)[(v_2 + v_3)\hat{\mathbf{x}} + (v_1 + v_3)\hat{\mathbf{y}} + (v_1 + v_2)\hat{\mathbf{z}}] \quad \dots(4)$$

The shortest  $\mathbf{G}$ 's are the following 12 vectors, where all choices of sign are independent:

$$(2\pi/a)(\pm \hat{\mathbf{y}} \pm \hat{\mathbf{z}}); \quad (2\pi/a)(\pm \hat{\mathbf{x}} \pm \hat{\mathbf{z}}); \quad (2\pi/a)(\pm \hat{\mathbf{x}} \pm \hat{\mathbf{y}}). \quad \dots(5)$$

The first Brillouin zone boundaries are determined by the twelve planes normal to these vectors at their midpoints.

The zone thus formed is a regular 12-faced solid, a rhombic dodecahedron (Fig. 25.65).

The vectors from the origin to the centre of each face of the first *BZ* are

$$\begin{aligned} &(\pi/a)(\pm\hat{\mathbf{y}}\pm\hat{\mathbf{z}}); \quad (\pi/a)(\pm\hat{\mathbf{x}}\pm\hat{\mathbf{z}}); \\ &(\pi/a)(\pm\hat{\mathbf{x}}\pm\hat{\mathbf{y}}). \end{aligned} \quad \dots(6)$$

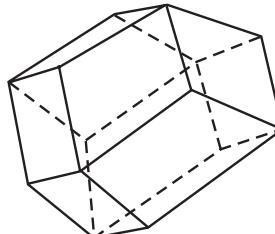


Fig. 25.65

### 25.23 RECIPROCAL LATTICE TO FCC LATTICE

The primitive translation vectors of the fcc lattice (Fig. 25.66) are

$$\begin{aligned} \mathbf{a}_1 &= \frac{1}{2}a(\hat{\mathbf{y}}+\hat{\mathbf{z}}); \quad \mathbf{a}_2 = \frac{1}{2}a(\hat{\mathbf{x}}+\hat{\mathbf{z}}); \\ \mathbf{a}_3 &= \frac{1}{2}a(\hat{\mathbf{x}}+\hat{\mathbf{y}}). \end{aligned}$$

The volume of the primitive cell is

$$V = |\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3| = \frac{1}{4}a^3$$

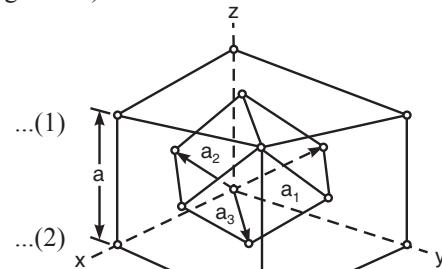


Fig. 25.66

The primitive translation vectors of the reciprocal lattice are

$$\begin{aligned} \mathbf{b}_1 &= (2\pi/a)(-\hat{\mathbf{x}}+\hat{\mathbf{y}}+\hat{\mathbf{z}}); \\ \mathbf{b}_2 &= (2\pi/a)(\hat{\mathbf{x}}-\hat{\mathbf{y}}+\hat{\mathbf{z}}); \\ \mathbf{b}_3 &= (2\pi/a)(\hat{\mathbf{x}}+\hat{\mathbf{y}}-\hat{\mathbf{z}}). \end{aligned} \quad \dots(3)$$

These are primitive translation vectors of a bcc lattice having length of the cube edge as  $(2\pi/a)$ . Thus the reciprocal lattice to an fcc lattice is a bcc lattice.

The shortest  $\mathbf{G}$ 's are the following eight vectors

$$(2\pi/a)(\pm\hat{\mathbf{x}}\pm\hat{\mathbf{y}}\pm\hat{\mathbf{z}}) \quad \dots(4)$$

The boundaries of the first Brillouin zone are determined mostly by the normal bisector planes to the above eight vectors. But the corners of the octahedron thus formed are cut by the planes that are the perpendicular bisectors of six other reciprocal lattice vectors;

$$\begin{aligned} &(2\pi/a)(\pm 2\hat{\mathbf{x}}); \quad (2\pi/a)(\pm 2\hat{\mathbf{y}}); \\ &(2\pi/a)(\pm 2\hat{\mathbf{z}}). \end{aligned} \quad \dots(5)$$

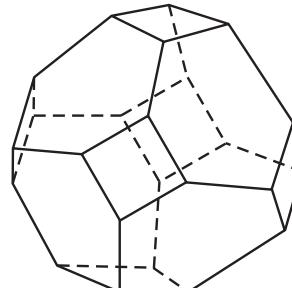


Fig. 25.67

The first Brillouin zone has the shape of the truncated octahedron (Fig. 25.67).

### 25.24 ATOMIC SCATTERING FACTOR

The atomic scattering factor,  $f$ , describes the scattering power of a single atom in relation to the scattering power of a single electron.

$$f = \frac{\text{amplitude of the radiation scattered from the atom}}{\text{amplitude of the radiation scattered from an electron}}$$

This factor ' $f$ ' is called the atomic scattering factor or form factor.

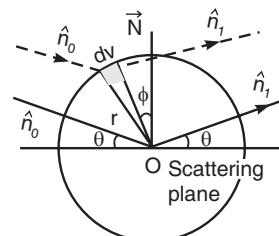


Fig. 25.68

### Calculation of $f$

Consider an atom containing electrons arranged in a spherically symmetric configuration around its centre which is taken as origin  $O$  (Fig. 25.68). Let  $r$  be the radius of the atom. Let  $\rho(\mathbf{r})$  be the charge density at a point  $\mathbf{r}$ . Consider a small volume element  $dV$  at  $\mathbf{r}$ .

Charge located at  $\mathbf{r} = \rho(\mathbf{r}) dV$ .

Let  $\hat{\mathbf{n}}_0$  and  $\hat{\mathbf{n}}_1$  represent the directions of the incident and scattered beam respectively. We first consider the scattering from the charge  $\rho(\mathbf{r}) dV$  and an electron located at the origin.

The phase difference between the radiation scattered from the charge  $\rho(\mathbf{r}) dV$  and that scattered from the electron is,

$$\phi_r = \frac{2\pi}{\lambda} (\mathbf{r} \cdot \mathbf{N}) \quad \dots(1)$$

Here,  $\mathbf{N}$  is the scattering normal.

Now, consider that the scattering amplitude from the point electron in the direction  $\hat{\mathbf{n}}_1$  is represented by  $Ae^{i(kx - \omega t)}$ . Here,  $x$  is the distance coordinate along  $\hat{\mathbf{n}}_1$  and  $k$  is the wave number. Then the scattering amplitude from the charge  $\rho(\mathbf{r}) dV$  in the same direction will be proportional to the magnitude of the charge and will contain the phase factor  $e^{i\phi_r}$ . It is represented by

$$Ae^{i(kx - \omega t) + i\phi_r} \rho(\mathbf{r}) dV$$

Therefore, the ratio of the amplitude of the radiation scattered by the charge element  $dV$  to that scattered by a point electron at the origin is,

$$\frac{df}{dV} = \frac{Ae^{i(kx - \omega t) + i\phi_r} \rho(\mathbf{r}) dV}{Ae^{i(kx - \omega t)}} = e^{i\phi_r} \rho(\mathbf{r}) dV$$

Therefore, ' $f$ ' the ratio of the scattered amplitude from the whole atom to that from a point electron is

$$f = \int_V \rho(\mathbf{r}) e^{i\phi_r} dV \quad \dots(2)$$

Here,  $V$  is the volume of the atom.

Electrons in the atom have spherically symmetric charge distribution. So  $\rho(\mathbf{r})$  is a function of  $r$  only. The volume element  $dV$  in the spherical coordinate system is,

$$dV = 2\pi r^2 \sin \phi d\phi dr \quad \dots(3)$$

We have,

$$|\mathbf{N}| = 2 \sin \theta$$

Therefore, Eq. (1) becomes

$$\phi_r = (2\pi/\lambda) r N \cos \phi = (4\pi/\lambda) r \sin \theta \cos \phi = \mu r \cos \phi \quad \dots(4)$$

Here,

$$\mu = (4\pi/\lambda) \sin \theta \quad \dots(5)$$

Substitution of Eqs. (3) and (4) in Eq. (2) gives,

$$f = \int_{r=0}^{\infty} \int_{\phi=0}^{\pi} \rho(\mathbf{r}) e^{i\mu r \cos \phi} 2\pi r^2 \sin \phi d\phi dr$$

$$\text{Now } \int_0^{\pi} e^{i\mu r \cos \phi} \sin \phi d\phi = \frac{2(\sin \mu r)}{\mu r}$$

$$\therefore f = \int_0^{\infty} 4\pi r^2 \rho(\mathbf{r}) \frac{\sin \mu r}{\mu r} dr \quad \dots(6)$$

This is the general expression for atomic scattering factor.

As  $\theta \rightarrow 0$ ,  $\mu \rightarrow 0$  and  $(\sin \mu r) / \mu r \rightarrow 1$ , we have

$$\lim_{\mu \rightarrow 0} f(\mu) = \int_0^\infty 4\pi r^2 \rho(r) dr = Z \quad \dots(7)$$

## 25.25 THE GEOMETRICAL STRUCTURE FACTOR

The intensity of an x-ray beam diffracted by a crystal depends upon the atomic scattering factors of the various atoms involved. It depends also on the contents of the unit cell, i.e., on the number, type and distribution of atoms in the unit cell. So if we wish to discuss the characteristics of the intensity of a particular order reflection and the relative intensities of the various reflections, we must take into account the contributions made by all the atoms in the unit cell to the scattered amplitude in a given direction.

Let us consider the  $(h' k' l')$  reflection. The total scattering amplitude  $F(h' k' l')$  for the reflection  $(h' k' l')$  is defined as the ratio of the amplitude of radiation scattered by the entire unit cell to the amplitude of radiation scattered by a single point electron placed at origin for the same wavelength. It is given by

$$F(h' k' l') = \sum_j f_j e^{i\phi_j} = \sum_j f_j e^{i(2\pi/\lambda)(\mathbf{r}_j \cdot \mathbf{N})} \quad \dots(1)$$

The summation extends over all the atoms belonging to a unit cell.

Here,

$f_j$  = the atomic scattering factor for the  $j$ th atom,

$\phi_j$  = phase difference between the radiation scattered from the  $j$ th atom of the unit cell and that scattered from the electron placed at the origin,

$\mathbf{r}_j$  = vector position of the  $j$ th atom,

$\mathbf{N}$  = scattering normal.

Let  $(u_j, v_j, w_j)$  represent the coordinates of  $j$ th atom. Then,

$$\mathbf{r}_j = u_j \mathbf{a} + v_j \mathbf{b} + w_j \mathbf{c} \quad \dots(2)$$

$$\therefore \mathbf{r}_j \cdot \mathbf{N} = \lambda (u_j h' + v_j k' + w_j l') \quad \dots(3)$$

Substituting Eq. (3) in Eq. (1), we get

$$F(h' k' l') = \sum_j f_j e^{i2\pi(u_j h' + v_j k' + w_j l')} \quad \dots(4)$$

For identical atoms, all the  $f_j$ 's have the same value  $f$ . Therefore, Eq. (4) becomes

$$F(h' k' l') = f \mathbf{S} \quad \dots(5)$$

Here,

$$\mathbf{S} = \sum_j e^{i2\pi(u_j h' + v_j k' + w_j l')} \quad \dots(6)$$

$\mathbf{S}$  is called the *geometrical structure factor*.  $\mathbf{S}$  depends upon the geometrical arrangement of atoms within the unit cell.

The intensity of the diffracted beam is proportional to the square of the amplitude  $F$ .  
 $I = F^2 = F^* F$

$$= \left[ \sum_j f_j \cos 2\pi(u_j h' + v_j k' + w_j l') \right]^2 + \left[ \sum_j f_j \sin 2\pi(u_j h' + v_j k' + w_j l') \right]^2 \quad \dots(7)$$

Here  $F^*$  is the complex conjugate of  $F$ .

The scattered intensity vanishes in all directions except those in which the structure factor  $S$  is nonvanishing. These latter directions are therefore the directions of diffraction. When the Bragg condition is satisfied, then the incident beam is diffracted into a single beam, which is recorded at the detector as a single spot on a film. This spot represents the whole set of reflecting planes  $(h' k' l')$ . When the crystal is rotated so that a new set of planes again satisfies the Bragg condition, then this new set appears as a new spot on the film at the detector. Therefore each spot on the film represents a whole set of crystalline planes. From the arrangement of these spots one can determine the structure of the crystal.

Consider the bcc lattice. The unit cell has two atoms whose coordinates are

$$(u_j, v_j, w_j) = (0, 0, 0) \text{ and } \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right).$$

$$\text{From Eq. (4), } F(h' k' l') = f[1 + e^{i\pi(h' + k' + l')}] \quad \dots(8)$$

This expression can take only two values. When  $(h' + k' + l')$  is even,  $F(h' k' l') = 2f$ .

When  $(h' + k' + l')$  is odd,  $F(h' k' l') = 0$ . Thus for the bcc lattice, the diffraction is absent for all those planes in which the sum  $(h' + k' + l')$  is odd, and is present for the planes in which  $(h' + k' + l')$  is even.

Metallic sodium has a bcc structure. The diffraction pattern does not contain lines such as (100), (300), (111), or (221), but lines such as (200), (110), and (222) will be present.

The basis of the fcc structure referred to the cubic cell has identical atoms at  $000; 0\frac{1}{2}\frac{1}{2};$

$$\frac{1}{2}0\frac{1}{2}; \frac{1}{2}\frac{1}{2}0.$$

From Eq. (4), the diffraction amplitude becomes

$$F(h' k' l') = f[1 + e^{\pi i(h' + l')} + e^{\pi i(h' + k')} + e^{\pi i(k' + l')}]$$

Here the expression within the square brackets is the structure factor for *fcc* crystals. The allowed reflections correspond to the cases in which  $(h' k' l')$  are either all even or all odd. No reflections can occur for which the indices are partly even and partly odd.

## CRYSTAL BONDING

### Bonding in Crystals

Most solids are crystalline, with the atoms, ions or molecules of which they are composed falling into regular, repeated three-dimensional patterns. The presence of *long-range order* is thus the defining property of a crystal. Crystals may be classified in terms of the dominant type of chemical binding force keeping the atoms together. All these bonds involve electrostatic forces, with the chief differences among them lying in the ways in which the outer electrons of the structural elements are distributed. The distinct types of bonds that provide the cohesive forces in crystals can be classified as follows: (i) the ionic bond (ii) the covalent bond (iii) the metallic bond (iv) the van der Waals bond and (v) the hydrogen bond. We briefly discuss the different types of bonds in crystals.

### Crystals of Inert Gases

The inert gases form the simplest crystals. The electron distribution is very close to that of the free atoms. The crystals are transparent insulators, weakly bound, with low melting temperatures. The atoms have very high ionization energies.

The outermost electron shells of the atom are completely filled. The distribution of electron charge in the free atom is spherically symmetric. In the crystal, the inert gas atoms pack together as closely as possible. The structures are all cubic close-packed (fcc), except  $\text{He}^3$  and  $\text{He}^4$ .

## 25.26 COHESIVE ENERGY OF INERT GAS CRYSTALS

The Cohesive energy of a crystal is defined as the energy that must be added to the crystal to separate its components into neutral free atoms at rest, at infinite separation, with the same electronic configuration.

The interaction energy of inert gas crystals is due to the following two types of interactions:

- (a) Van der Waals or London interaction, and
- (b) Repulsive interaction.

### (a) Van der Waals-London Interaction

Consider two identical inert gas atoms at a separation  $R$  large in comparison with the radii of the atoms (Fig. 25.69). The atoms induce dipole moments in each other. The induced moments cause an attractive interaction

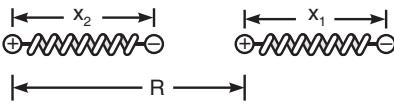


Fig. 25.69

Consider two identical linear harmonic oscillators 1 and 2 separated by  $R$ . Each oscillator bears charges  $\pm e$  with separations  $x_1$  and  $x_2$ . The particles oscillate along the  $x$  axis. Let  $p_1$  and  $p_2$  denote the momenta. The force constant is  $C$ .

The hamiltonian of the unperturbed system is

$$H_0 = \frac{1}{2m} p_1^2 + \frac{1}{2} Cx_1^2 + \frac{1}{2m} p_2^2 + \frac{1}{2} Cx_2^2 \quad \dots(1)$$

We assume that each uncoupled oscillator has the frequency  $\omega_0$  of the strongest optical absorption line of the atom. Thus  $C = m\omega_0^2$  or  $\omega_0 = (C/m)^{1/2}$ .

Let  $H_1$  be the coulomb interaction energy of the two oscillators. We use CGS units. Then

$$H_1 = \frac{e^2}{R} + \frac{e^2}{R+x_1-x_2} - \frac{e^2}{R+x_1} - \frac{e^2}{R-x_2} \quad \dots(2)$$

$$H_1 \approx -\frac{2e^2 x_1 x_2}{R^3} \quad (\because |x_1|, |x_2| \ll R) \quad \dots(3)$$

The total hamiltonian  $H = H_0 + H_1$  ...(4)

$H$  can be diagonalized by the normal mode transformation

$$x_s \equiv \frac{1}{\sqrt{2}}(x_1 + x_2); \quad x_a \equiv \frac{1}{\sqrt{2}}(x_1 - x_2). \quad \dots(5)$$

The subscripts  $s$  and  $a$  denote symmetric and antisymmetric modes of motion. Solving for  $x_1$  and  $x_2$ , we get

$$x_1 = \frac{1}{\sqrt{2}}(x_s + x_a); \quad x_2 = \frac{1}{\sqrt{2}}(x_s - x_a). \quad \dots(6)$$

The momenta  $p_s, p_a$  associated with the two modes are,

$$p_1 \equiv \frac{1}{\sqrt{2}}(p_s + p_a); \quad p_2 \equiv \frac{1}{\sqrt{2}}(p_s - p_a). \quad \dots(7)$$

$$H = \left[ \frac{1}{2m} P_s^2 + \frac{1}{2} \left( C - \frac{2e^2}{R^3} \right) x_s^2 \right] + \left[ \frac{1}{2m} P_a^2 + \frac{1}{2} \left( C + \frac{2e^2}{R^3} \right) x_a^2 \right] \quad \dots(8)$$

The two frequencies of the coupled oscillators are,

$$\omega = \left[ \left( C \pm \frac{2e^2}{R^3} \right) / m \right]^{1/2} \cong \omega_0 \left[ 1 \pm \frac{1}{2} \left( \frac{2e^2}{CR^3} \right) - \frac{1}{8} \left( \frac{2e^2}{CR^3} \right)^2 + \dots \right]. \quad \dots(9)$$

The zero point energy of the system is  $\frac{1}{2}\hbar(\omega_s + \omega_a)$ . Because of the interaction the sum is lowered from the uncoupled value  $2 \cdot \frac{1}{2}\hbar\omega_0$  by

$$\Delta U = \frac{1}{2}\hbar(\Delta\omega_s + \Delta\omega_a) = -\hbar\omega_0 \cdot \frac{1}{8} \left( \frac{2e^2}{CR^3} \right)^2 = -\frac{A}{R^6}. \quad \dots(10)$$

This attractive interaction varies as the minus sixth power of the separation of the two oscillators.

This is called the van der Waals interaction (or the London interaction or the induced dipole-dipole interaction).

### (b) Repulsive Interaction

As the two atoms are brought together, their charge distributions gradually overlap. So the electrostatic energy of the system changes. At sufficiently close separations the overlap energy is repulsive, because of the *Pauli exclusion principle*. The electron overlap increases the total energy of the system and gives a repulsive contribution to the interaction. We shall express the repulsive interaction by an empirical repulsive potential of the form  $B/R^{12}$ , where  $\mathbf{B}$  is a positive constant.

The total potential energy of two atoms at separation  $R$  is

$$U(R) = \frac{B}{R^{12}} - \frac{A}{R^6}$$

or

$$U(R) = 4\epsilon \left[ \left( \frac{\sigma}{R} \right)^{12} - \left( \frac{\sigma}{R} \right)^6 \right]. \quad \dots(11)$$

Here,  $\epsilon$  and  $\sigma$  are the new parameters, with  $4\epsilon\sigma^6 \equiv A$  and  $4\epsilon\sigma^{12} \equiv B$ .

The potential (11) is called *Lennard-Jones Potential* and is plotted in Fig. 25.70.

### Equilibrium Lattice Constants

We neglect the kinetic energy of the inert gas atoms. The cohesive energy of an inert gas crystal is given by summing the Lennard-Jones potential (11) over all pairs of atoms in the crystal. Suppose that the crystal has  $N$  atoms.

The total potential energy is

$$U_{\text{tot}} = \frac{1}{2} N (4\epsilon) \left[ \sum_j \left( \frac{\sigma}{p_{ij} R} \right)^{12} - \sum_j \left( \frac{\sigma}{p_{ij} R} \right)^6 \right]. \quad \dots(12)$$

Here,  $p_{ij} R$  is the distance between reference atom  $i$  and any other atom  $j$ , expressed in terms of the nearest neighbour distance  $R$ .

$\frac{1}{2}N$  has been used instead of  $N$  because of the fact that interaction of pair of  $ij$  atoms will be only once.

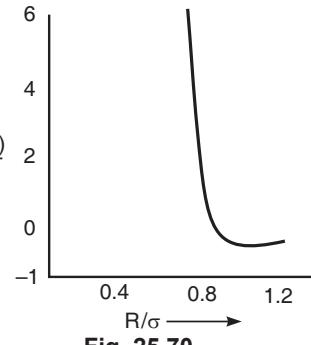


Fig. 25.70

$$\text{For the fcc structure, } \sum_j p_{ij}^{-12} = 12.13188; \quad \sum_j p_{ij}^{-6} = 14.45392. \quad \dots(13)$$

Since there are 12 nearest neighbours of an atom in *fcc* structure, the major contribution to the interaction energy arises from the nearest neighbours.

The equilibrium distance  $R_0$  is obtained from the condition that at  $R = R_0$ ,  $dU_{tot} / dR = 0$ .

$$\frac{dU_{tot}}{dR} = -2N\varepsilon \left[ (12)(12 \cdot 13) \frac{\sigma^{12}}{R_0^{13}} - (6)(14 \cdot 45) \frac{\sigma^6}{R_0^7} \right]. \quad \dots(14)$$

$$\therefore R_0/\sigma = 1.09. \quad \dots(15)$$

It is the same for all elements with an *fcc* structure.

### Cohesive Energy

The cohesive energy of inert gas crystals at absolute zero and at zero pressure is obtained by substituting (13) and (15) in (12).

$$U_{tot}(R) = 2N\varepsilon \left[ (12 \cdot 13) \left( \frac{\sigma}{R} \right)^{12} - (14 \cdot 45) \left( \frac{\sigma}{R} \right)^6 \right]. \quad \dots(16)$$

$$\text{At } R = R_0, \quad U_{tot}(R_0) = -(2.15)(4N\varepsilon), \quad \dots(17)$$

It is the same for all inert gases.

### Molecular Bond

Neutral atoms with closed electron shells are bound together weakly by the *van der Waals forces*. The van der Waals' attraction was first explained for electrically neutral gas molecules by Debye. He assumed that neighbouring molecules induced dipoles in each other because of their own changing electric fields. This interaction produces an attractive force that is inversely proportional to the seventh power of the separation. The molecules are located at the crystal lattice points and the bonds between them are developed by van der Waals' forces. Solid argon and soild methane ( $\text{CH}_4$ ) are examples of molecular crystals.

**Properties.** Van der Waals' forces are much weaker than those found in ionic and covalent bonds. As a result, molecular crystals generally have low melting and boiling points and little mechanical strength. Molecular crystals have low cohesive energies. Molecular crystals are good insulators due to non-availability of free electrons.

## IONIC CRYSTALS

Ionic crystals are made up of positive and negative ions. The ionic bond results from the electrostatic interaction of oppositely charged ions.

### 25.27 IONIC BOND

Ionic bonds are formed when atoms that have low ionization energies, and hence lose electrons readily, interact with other atoms that tend to acquire excess electrons. The former atoms give up electrons to the latter. Thus the atoms become positive and negative ions respectively. These ions come together in an equilibrium configuration in which the attractive forces between positive and negative ions predominate over the repulsive forces between similar ions.

Consider the case of sodium chloride which is a typical example of an ionic crystal. Here, a single valence electron is transferred from the sodium atom to the chlorine atom. The  $\text{Na}^+$  and  $\text{Cl}^-$  ions so formed are arranged in a face-centred cubic structure [Fig. 25.71]. A different arrangement is found in cesium chloride crystals. The body-centred cubic structure of a  $\text{CsCl}$  crystal is shown in Fig. 25.72.

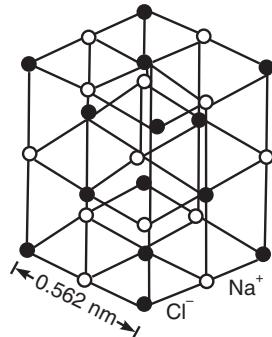


Fig. 25.71

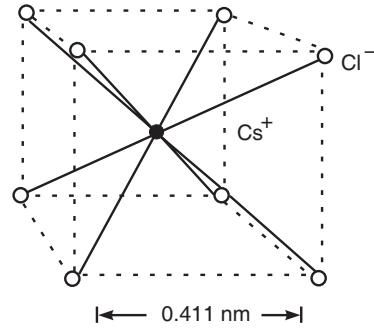


Fig. 25.72

### Expression for the cohesive energy of an ionic crystal

The *cohesive energy* of an ionic crystal is the energy that would be liberated by the formation of the crystal from individual neutral atoms. The principal contribution to the cohesive energy of an ionic crystal is the electrostatic potential energy  $U_{\text{coulomb}}$  of the ions. Let us consider an  $\text{Na}^+$  ion in  $\text{NaCl}$ . Its nearest neighbours are six  $\text{Cl}^-$  ions, each one the distance  $r$  away.

$$\left. \begin{array}{l} \text{The potential energy of the } \text{Na}^+ \text{ ion} \\ \text{due to the } 6 \text{ } \text{Cl}^- \text{ ions} \end{array} \right\} U_1 = - \frac{6e^2}{4\pi\epsilon_0 r}$$

The next nearest neighbours are 12  $\text{Na}^+$  ions, each one the distance  $\sqrt{2} r$  away.

$$\left. \begin{array}{l} \text{The potential energy of the } \text{Na}^+ \text{ ion} \\ \text{due to the } 12 \text{ } \text{Na}^+ \text{ ions} \end{array} \right\} U_2 = + \frac{12e^2}{4\pi\epsilon_0 \sqrt{2} r}$$

Then there are 8  $\text{Cl}^-$  ions at  $\sqrt{3} r$  distance, 6  $\text{Na}^+$  ions at  $2r$  distance and so on. When the summation is continued over all the + and - ions in a crystal of infinite size, the result is

$$\begin{aligned} U_{\text{coulomb}} &= - \frac{6e^2}{4\pi\epsilon_0 r} + \frac{12e^2}{4\pi\epsilon_0 \sqrt{2} r} - \frac{8e^2}{4\pi\epsilon_0 \sqrt{3} r} + \frac{6e^2}{4\pi\epsilon_0 2r} + \dots \\ &= - \frac{e^2}{4\pi\epsilon_0 r} \left( \frac{6}{\sqrt{1}} - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - \frac{6}{\sqrt{4}} + \dots \right) \\ &= - 1.748 \frac{e^2}{4\pi\epsilon_0 r} \end{aligned}$$

or in general,

$$U_{\text{coulomb}} = - \alpha \frac{e^2}{4\pi\epsilon_0 r} \quad \dots(1)$$

This result holds for the potential energy of a  $\text{Cl}^-$  ion also.  $\alpha$  is called the *Madelung constant* of the crystal. It has the same value for all crystals of the same structure. For simple crystal structures  $\alpha$  lies between 1.6 and 1.8.

Two ions cannot continuously approach each other under coulomb attraction on account of the exclusion principle. When they are at a certain small distance apart, they begin to repel each other with a force which increases rapidly with decreasing internuclear distance  $r$ . The potential energy contribution of the short-range repulsive forces can be expressed approximately in the form

$$U_{\text{repulsive}} = \frac{B}{r^n} \quad \dots(2)$$

where  $B$  is a constant and  $n$  is a number ( $n \approx 9$ ).

Therefore, the total potential energy  $U$  of each ion due to its interactions with all the other ions is

$$U = U_{\text{Coulomb}} + U_{\text{repulsive}} = -\frac{\alpha e^2}{4\pi\epsilon_0 r} + \frac{B}{r^n} \quad \dots(3)$$

We must now evaluate the constant  $B$ . At the equilibrium separation  $r_0$  of the ions,  $U$  is a minimum. So  $(dU/dr) = 0$  when  $r = r_0$ .

$$\begin{aligned} \left(\frac{dU}{dr}\right)_{r=r_0} &= \frac{\alpha e^2}{4\pi\epsilon_0 r_0^2} - \frac{nB}{r_0^{n+1}} = 0 \\ \text{or} \quad B &= \frac{\alpha e^2}{4\pi\epsilon_0 n} r_0^{n-1} \end{aligned} \quad \dots(4)$$

The total potential energy at the equilibrium separation is

$$\begin{aligned} U &= -\frac{\alpha e^2}{4\pi\epsilon_0 r_0} + \frac{\alpha e^2 r_0^{n-1}}{4\pi\epsilon_0 n r_0^n} \\ \therefore U &= -\frac{\alpha e^2}{4\pi\epsilon_0 r_0} \left(1 - \frac{1}{n}\right) \end{aligned} \quad \dots(5)$$

This is the magnitude of the energy needed to separate an ionic crystal into individual ions (not into atoms).

In an NaCl crystal, the equilibrium distance  $r_0$  between ions is  $2.81 \times 10^{-10}$  m,  $\alpha = 1.748$  and  $n \approx 9$ ,  $1/4\pi\epsilon_0 = 9 \times 10^9 \text{ Nm}^2 \text{ C}^{-2}$ .

The potential energy of an ion of either sign is

$$\begin{aligned} U &= -\frac{\alpha e^2}{4\pi\epsilon_0 r_0} \left(1 - \frac{1}{n}\right) \\ &= -\frac{(9 \times 10^9 \text{ Nm}^2 \text{ C}^{-2})(1.748)(1.60 \times 10^{-19} \text{ C})^2}{2.81 \times 10^{-10} \text{ m}} \left(1 - \frac{1}{9}\right) \\ &= -1.27 \times 10^{-18} \text{ J} = -7.96 \text{ eV.} \end{aligned}$$

Because we may not count each ion more than once, only half this potential energy, or  $-3.98$  eV, represents the contribution *per ion* to the cohesive energy of the crystal.

Some energy is needed to transfer an electron from a Na atom to a Cl atom to form a  $\text{Na}^+ - \text{Cl}^-$  ion pair. This electron transfer energy is the difference between the  $+5.14$  eV ionization energy of Na and the  $-3.61$  eV electron affinity of Cl, or  $+1.53$  eV. Each atom thus contributes  $+0.77$  eV to the cohesive energy.

$\therefore$  the total cohesive energy per atom in the NaCl crystal is

$$E_{\text{cohesive}} = (-3.98 + 0.77) \text{ eV/atom} = -3.21 \text{ eV/atom.}$$

**Properties :** (i) Most ionic solids are hard, brittle and have high melting points.

(ii) They are soluble in polar liquids like water.

(iii) Their electrical conductivity is much smaller than that of metals at room temperature. But in contrast to metals, the conductivity of ionic crystals increases with increasing temperature. At elevated temperatures, the ions themselves become mobile and ionic conductivity results.

(iv) Ionic solids crystallize in close-packed structures, of which the NaCl and CsCl structures are the commonest.

### 25.27.1. Calculation of Repulsive Exponent $n$

Born determined the repulsive exponent ( $n$ ) from measurements of the compressibility of the crystal as follows:

The compressibility  $C$  (i.e., reciprocal of bulk modulus  $K$ ) is defined as

$$\frac{1}{C} = K = -V \left( \frac{dP}{dV} \right) \quad \dots(1)$$

where  $V$  is volume,  $P$  is pressure.

From the first law of thermodynamics

$$dQ = dU + P dV$$

At absolute zero there is no absorption of heat energy by the ions.

$$\therefore dQ = 0.$$

$\therefore$  change in internal energy,  $dU = -P dV$

$$\frac{dU}{dV} = -P$$

$$\text{or} \quad \frac{d^2 U}{dV^2} = -\frac{dP}{dV} \quad \dots(2)$$

$$\therefore \frac{1}{C} = V \frac{d^2 U}{dV^2} \quad \dots(3)$$

$$\text{Also} \quad \frac{dU}{dV} = \frac{dU}{dr} \cdot \frac{dr}{dV} \text{ i.e., } \frac{d}{dV} = \frac{dr}{dV} \cdot \frac{d}{dr} \quad \dots(4)$$

$$\begin{aligned} \text{and} \quad \frac{d^2 U}{dV^2} &= \frac{d}{dV} \left( \frac{dU}{dV} \right) = \frac{d}{dV} \left( \frac{dU}{dr} \cdot \frac{dr}{dV} \right) \\ &= \frac{dU}{dr} \cdot \frac{d^2 r}{dV^2} + \frac{dr}{dV} \cdot \frac{d}{dV} \left( \frac{dU}{dr} \right) \\ &= \frac{dU}{dr} \cdot \frac{d^2 r}{dV^2} + \left( \frac{dr}{dV} \right) \cdot \frac{d}{dV} \frac{d}{dr} \left( \frac{dU}{dr} \right) \\ &= \frac{dU}{dr} \cdot \frac{d^2 r}{dV^2} + \left( \frac{dr}{dV} \right)^2 \cdot \frac{d^2 U}{dr^2} \end{aligned} \quad \dots(5)$$

For NaCl crystal the volume of unit cell is

$$(2r)^3 = 8r^3$$

The unit cell has 4 sodium ions and 4 chloride ions. The volume  $8r^3$  corresponds to 8 ions. So each ion corresponds to the volume  $r^3$ . If the crystal contains  $N$ -Na and  $N$ -Cl atoms, then  $V = 2Nr^3$ .

$$\therefore \frac{dV}{dr} = 6Nr^2 \text{ or } \frac{dr}{dV} = \frac{1}{6Nr^2}$$

$$\frac{d^2 r}{dV^2} = \frac{d}{dV} \left( \frac{dr}{dV} \right) = \frac{dr}{dV} \cdot \frac{d}{dr} \left( \frac{dr}{dV} \right) = \frac{1}{6Nr^2} \times \frac{d}{dr} \left( \frac{1}{6Nr^2} \right) = -\frac{1}{18N^2 r^5}$$

Substituting these values in Eq. (5),

$$\frac{d^2 U}{dV^2} = -\frac{1}{18N^2 r^5} \frac{dU}{dr} + \frac{1}{36N^2 r^4} \frac{d^2 U}{dr^2} \quad \dots(6)$$

At  $r = r_0$ ,  $U$  is minimum and  $dU/dr = 0$ .

$$\left( \frac{d^2 U}{dV^2} \right)_{r=r_0} = \frac{1}{36N^2 r_0^4} \left( \frac{d^2 U}{dr^2} \right)_{r=r_0} \quad \dots(7)$$

Substituting the above value in Eq. (3),

$$\begin{aligned} \frac{1}{C} = & \left( V \frac{d^2 U}{dV^2} \right)_{r=r_0} = 2Nr_0^3 \times \frac{1}{36N^2 r_0^4} \left( \frac{d^2 U}{dr^2} \right)_{r=r_0} \\ & = \frac{1}{18Nr_0} \cdot \left( \frac{d^2 U}{dr^2} \right)_{r=r_0} \end{aligned} \quad \dots(8)$$

But

$$U = NU_i = N \left[ \frac{B}{r^n} - \frac{\alpha e^2}{4\pi\epsilon_0 r} \right] \quad \dots(9)$$

$$\therefore \frac{dU}{dr} = N \left[ -\frac{Bn}{r^{n+1}} + \frac{\alpha}{4\pi\epsilon_0} \times \frac{e^2}{r^2} \right]$$

When

$$r = r_0, \frac{dU}{dr} = 0$$

$$-\frac{Bn}{r_0^{n+1}} + \frac{\alpha}{4\pi\epsilon_0} \frac{e^2}{r_0^2} = 0$$

$\therefore$

$$B = \frac{\alpha e^2 r_0^{n-1}}{4\pi\epsilon_0 n} \quad \dots(10)$$

Also

$$\frac{d^2 U}{dr^2} = N \left[ \frac{n(n+1)B}{r_0^{n+2}} - \frac{2\alpha e^2}{4\pi\epsilon_0 r_0^3} \right]$$

Substituting the value of  $B$ ,

$$\left( \frac{d^2 U}{dr^2} \right)_{r=r_0} = \frac{N\alpha e^2 (n-1)}{4\pi\epsilon_0 r_0^3} \quad \dots(11)$$

Substituting this value in Eq. (8),

$$\frac{1}{C} = \frac{1}{18Nr_0} \frac{N\alpha e^2 (n-1)}{4\pi\epsilon_0 r_0^3}$$

$\therefore$

$$n = 1 + \frac{(4\pi\epsilon_0)18r_0^4}{C\alpha e^2} \quad \dots(12)$$

Thus the value of  $n$  may be determined.

### 25.27.2. Born-Haber Cycle

**Lattice energy :** The lattice energy is defined as the energy released in the process when the constituent ions are placed in their respective positions in the crystal lattice or this is the amount of energy which is spent to separate the solid ionic crystal into its constituent ions.

An experimental check on the calculated values of the lattice energies may be obtained from Born-Haber cycle. The Born-Haber cycle for NaCl is represented as follows [Fig. 25.73] :

Here  $S$  = Heat of sublimation,

$D$  = Dissociation energy

$I$  = Ionisation energy,  $E$  = Electron affinity

$\Delta H$  = Heat of dissociation.  $U$  = Lattice energy.

The cycle can be explained as follows :

(i) Initially we start with solid sodium and chlorine molecule. The solid sodium is subjected to the sublimation energy ( $S$ ) and it is vaporised. To the chlorine molecule, dissociation energy ( $D$ ) is supplied and as a result it is dissociated into its constituent atoms. Here  $\frac{D}{2}$  is the dissociation energy per chlorine atom.

(ii) Next ionisation energy  $I$  is supplied. The outer electron of Na gaseous atom is removed. This electron is added to Cl atom. As the Cl atom has the electron affinity  $E$ , an energy  $E$  is given out.

(iii) The two ions ( $\text{Na}^+$  and  $\text{Cl}^-$ ) are arranged in the lattice and hence the *lattice energy*  $U$  is released.

(iv) We again reach the starting point by supplying an energy ( $\Delta H$ ) known as *heat of dissociation*.

$$\text{Now } S + \frac{D}{2} + I - E - U + \Delta H = 0$$

$$\therefore U = S + \frac{D}{2} + I - E + \Delta H$$

All quantities excepting  $U$  are experimentally determinable. Hence the lattice energy  $U$  can be calculated.

### 25.27.3. Binding Energy of Ionic Crystals (Alternative Analysis)

The energy of the crystal is lower than that of the free atoms by an amount equal to the energy required to pull the crystal apart into a set of free atoms. This is called the binding energy (or Cohesive energy) of the Crystal.

The theory developed by Born and Madelung is based on the assumption that the ionic crystals are built up of positive and negative ions each having a spherically symmetric charge distribution. The force between the ions is assumed to be mainly electrostatic. Thus the main contribution to the binding energy arises from electrostatic interaction and is called the *Madelung energy*. The ions tend to acquire such an arrangement in a crystal structure which results in the maximum attractive interaction amongst themselves.

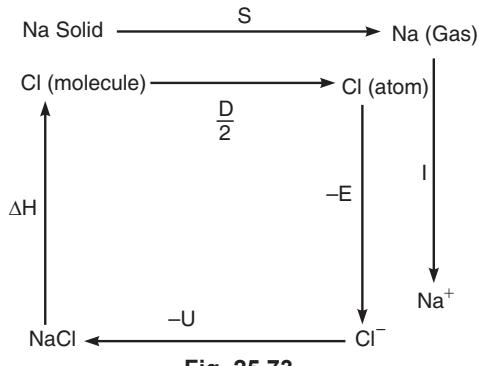


Fig. 25.73

There exist two types of interactions in ionic crystals.

- (i) The Coulomb's electrostatic interaction energy between two ions with charges  $\pm q$  is given by

$$\frac{\pm q^2}{4\pi\epsilon_0 r}$$

Here,  $r$  is the distance of separation between them.

- (ii) There is an interaction due to the overlap of the electronic shells of neighbouring atoms. The repulsive energy is of the form

$$\lambda e^{-r/\rho}$$

Here,  $\lambda$  represents the strength and  $\rho$  the range of repulsive interaction.

This expression for the repulsive energy is called the *central field repulsive potential*.

Let  $U_{ij}$  be the interaction energy between ions  $i$  and  $j$ . Then,

$$U_{ij} = \lambda \exp(-r_{ij}/\rho) \pm \frac{q^2}{4\pi\epsilon_0 r_{ij}} \quad \dots(1)$$

We define a sum  $U_i$  which includes all interactions involving the ion  $i$ :

$$U_i = \sum_j U_{ij} \quad \dots(2)$$

Here, the summation includes all ions except  $j = i$ .

The total lattice energy  $U_{\text{tot}}$  of a crystal composed of  $N$  molecules or  $2N$  ions is

$$U_{\text{tot}} = NU_i \quad \dots(3)$$

Here  $N$ , rather than  $2N$ , occurs because we must count each *pair* of interactions only once. The total lattice energy is defined as the energy required to separate the crystal into individual ions at an infinite distance apart.

We introduce quantities  $p_{ij}$  such that  $r_{ij} \equiv p_{ij} R$ . If ions  $i$  and  $j$  are the nearest neighbours, then  $r_{ij} = R$  or  $p_{ij} = 1$ .

Here,  $R$  is the nearest-neighbour separation in the crystal.

If we include the repulsive interaction only among nearest neighbours, we have

$$U_{ij} = \begin{cases} \lambda \exp(-R/\rho) - \frac{q^2}{4\pi\epsilon_0 R} & (\text{nearest neighbours}) \\ \pm \frac{1}{p_{ij}} \frac{q^2}{4\pi\epsilon_0 R} & (\text{otherwise}) \end{cases} \quad \dots(4)$$

$$\therefore U_{\text{tot}} = NU_i = N \left( z\lambda e^{-R/\rho} - \frac{\alpha q^2}{4\pi\epsilon_0 R} \right) \quad \dots(5)$$

Here,  $z$  is the number of nearest neighbours of any ion and

$$\alpha \equiv \sum_j \frac{(\pm)}{p_{ij}} \equiv \text{Madelung constant} \quad \dots(6)$$

The sum should include the nearest-neighbour contribution, which is just  $z$ .

The choice of sign depends upon the type of the reference ion. If the reference ion is negative, the positive sign is used for a positive ion and the negative sign for a negative ion. Thus the value of the

Madelung constant depends on the lattice structure. It is basically a correction factor that determines the magnitude of the error introduced by considering only the nearest neighbour interaction.

At the equilibrium separation,  $R_0$ ,  $dU_{tot}/dR = 0$ .

$$\begin{aligned} N \frac{dU_i}{dR} \Big|_{R=R_0} &= -\frac{Nz\lambda}{\rho} e^{-R_0/\rho} + \frac{N\alpha q^2}{4\pi\epsilon_0 R_0^2} = 0 \\ \text{or } R_0^2 e^{(-R_0/\rho)} &= \frac{\rho\alpha q^2}{4\pi\epsilon_0 z\lambda} \end{aligned} \quad \dots(7)$$

This determines the equilibrium separation  $R_0$  if the parameters  $\rho$ ,  $\lambda$  of the repulsive interaction are known.

The total lattice energy of the crystal of  $2N$  ions at their equilibrium separation  $R_0$  may be written, using (5) and (7).

$$U_{tot} = -\frac{N\alpha q^2}{4\pi\epsilon_0 R_0} \left( 1 - \frac{\rho}{R_0} \right) \quad \dots(8)$$

The term  $\left( \frac{-N\alpha q^2}{4\pi\epsilon_0 R_0} \right)$  is the *Madelung energy*.

The equilibrium distance  $R_0$  is determined empirically, the range  $\rho$  is determined from the knowledge of the bulk modulus of the crystal, and the Madelung constant  $\alpha$  is determined theoretically from the geometry of the crystal structure.

Thus the lattice energy of a crystal can be determined.

#### 25.27.4. Evaluation of the Madelung Constant

The definition of the Madelung constant  $\alpha$  is,

$$\alpha = \sum_j \frac{(\pm)}{p_{ij}} \quad \dots(1)$$

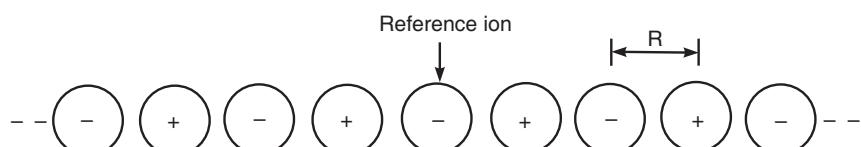
An equivalent definition is

$$\frac{\alpha}{R} = \sum_j \frac{(\pm)}{r_j} \quad \dots(2)$$

Here,  $r_j$  is the distance of the  $j$ th ion from the reference ion and  $R$  is the nearest-neighbour distance.

Consider a one-dimensional crystal consisting of alternate positive and negative ions (Fig. 25.74).

Let  $R$  be interionic spacing. Take a negative ion as the reference ion. Then



**Fig. 25.74**

$$\frac{\alpha}{R} = 2 \left[ \frac{1}{R} - \frac{1}{2R} + \frac{1}{3R} - \frac{1}{4R} + \dots \right],$$

or 
$$\alpha = 2 \left[ 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots \right] \quad \dots(3)$$

The factor of 2 occurs because there are two equivalent ions, one to the right and one to the left each at distance  $r_j = nR$ .

Using the series expansion

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

and putting  $x = 1$  in it, we get

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$

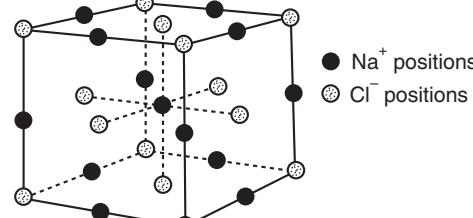
Therefore, Eq. (3) gives

$$\alpha = 2 \ln 2 = 1.38 \quad \dots(4)$$

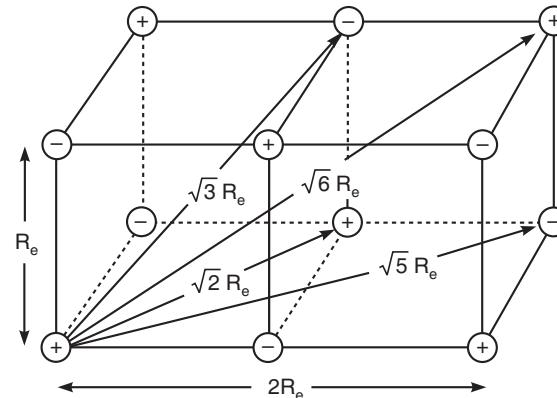
### Evaluation of Madelung constant for NaCl Crystal

The **Madelung constant**,  $\alpha$  is a function of crystal structure. It can be calculated from the geometrical arrangement of ions in the crystal.

Fig. 25.75 shows the equilibrium positions of ions in NaCl crystal.



**Fig. 25.75**



**Fig. 25.76**

Consider a central  $\text{Na}^+$  ion as the reference ion having a single positive charge on it. This ion is surrounded by 6  $\text{Cl}^-$  ions as first nearest neighbours. Let us consider them at unit distance (Fig. 25.76).

- Twelve  $\text{Na}^+$  ions are the second nearest neighbours at a distance  $\sqrt{2}$ .
- Eight  $\text{Cl}^-$  ions are the third nearest neighbours at a distance  $\sqrt{3}$  and so on.

The Madelung constant for the NaCl structure can be written as a summation series.

$$\alpha = \frac{6}{1} - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - \frac{6}{\sqrt{4}} + \frac{24}{\sqrt{5}} - \frac{24}{\sqrt{6}} + \dots$$

This converges to a value 1.74756.

## 25.28 COVALENT BOND

In the covalent bond, atoms are held together by the sharing of electrons. Each atom participating in a covalent bond contributes an electron to the bond. These electrons are shared by both atoms rather than being the virtually exclusive property of one of them as in an ionic bond. Diamond

is an example of a crystal whose atoms are linked by covalent bonds. Fig. 25.77 shows the structure of a diamond crystal. The tetrahedral arrangement is a consequence of the ability of each carbon atom to form covalent bonds with four other atoms.

**The hydrogen molecule.** The binding of molecular hydrogen ( $H_2$ ) is a simple example of covalent bond. In the  $H_2$  molecule, two electrons are shared by the two atoms (Fig. 25.78). As these electrons circulate, they spend more time between the atoms (in fact between the protons) than elsewhere and this produces an attractive force. Thus covalent crystals are composed of neutral atoms having slightly overlapping electron clouds. Diamond, germanium, silicon, and silicon carbide (SiC) are examples of covalent crystals. In SiC each atom is surrounded by four atoms of the other kind in the same tetrahedral structure as that of diamond.

Fig. 25.79 shows the variation of the energy of the system  $H + H$  with their distances apart when the electron spins are parallel and antiparallel.

The difference between the two curves is due to the Pauli exclusion principle, which leads to a dominating repulsion when the spins are parallel. The strongest binding occurs when the spins of the two electrons are anti parallel. The binding depends on the relative spin orientation not because there are strong magnetic dipole forces between the spins, but because the Pauli principle modifies the distribution of charge according to the spin orientation. A spin-dependent coulomb energy of the system is found. This is called the *exchange interaction*.

#### Directional Nature of a Covalent Bond

The directional nature of the covalent bond results from the restricted orbital motion of the electrons. Fig. 25.80 shows the shapes of atomic orbitals  $s(l=0)$  and  $p(l=1)$ .

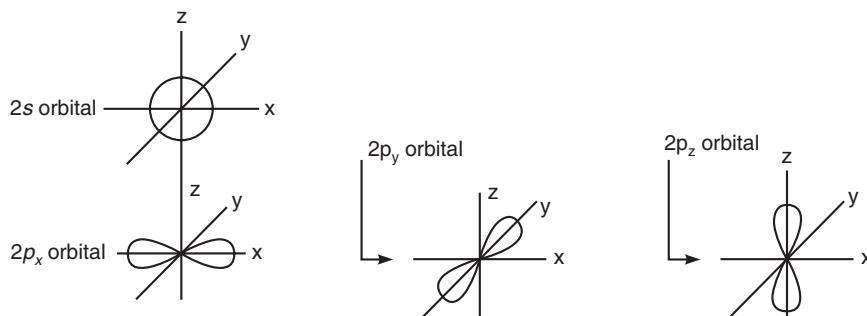


Fig. 25.80

In  $p$ -orbitals, the shapes represent the angular dependence (directional) distribution of electron density.

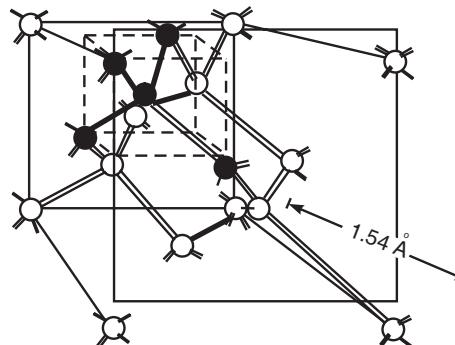


Fig. 25.77

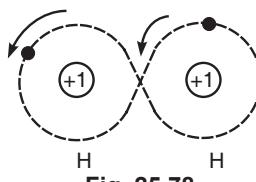


Fig. 25.78

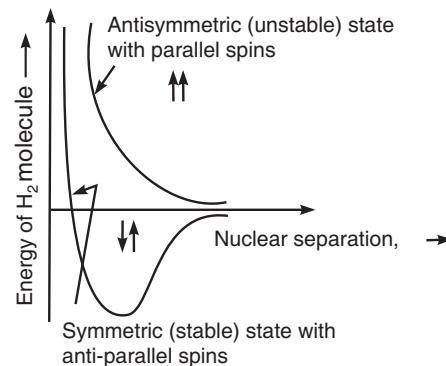


Fig. 25.79

**Hybridisation.** Covalent bonds are formed not only due to the overlap of either pure *s* orbitals or pure *p* orbitals but also due to the overlap of *s* and *p* orbitals. Such bonding is called **hybrid bonding**. Carbon exhibits such a bonding. The electronic configuration of normal carbon atom is  $1s^2 2s^2 2p^2$ . The electron spin distribution is

$$\begin{array}{c} \uparrow\downarrow \quad \uparrow\downarrow \quad \uparrow \quad \uparrow \\ 1s \quad 2s \quad 2p_x \quad 2p_y \quad 2p_z \end{array}$$

- The electron spins are paired in *1s* and *2s* orbitals only.
- The *2p* electrons remain unpaired.

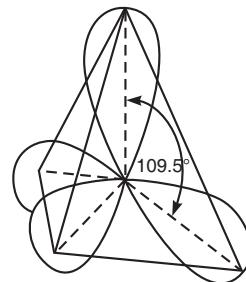
It is these unpaired electrons which are available for bond formation. Accordingly, the carbon atom is expected to form only two bonds.

- However, when the carbon atoms approach each other, one electron from *2s* orbital is excited to the *p* level. The resulting configuration becomes  $1s^2 2s^1 2p^3$ .

The electron spin distribution is

$$\begin{array}{c} \uparrow\downarrow \quad \uparrow \quad \uparrow \quad \uparrow \quad \uparrow \\ 1s \quad 2s \quad 2p_x \quad 2p_y \quad 2p_z \end{array}$$

This configuration shows that there are four unpaired electron spins. The favourable bonding directions of these orbitals are directed towards four corners of a regular tetrahedron with the bond angles  $109.5^\circ$  (Fig. 25.81).



**Fig. 25.81.**  $sp^3$  orbitals directed symmetrically towards the corners of a tetrahedron

This arrangement of orbitals is called **hybridization**.

The four orbitals are called  $sp^3$  hybrids.

- In diamond, carbon atoms exhibit  $sp^3$  tetrahedral covalent bonding. Because of these strong, directional, primary valence forces, the crystal becomes strong with high melting point and low thermal expansion coefficient.

Since it is extremely hard, it is used as an abrasive.

Since valence electrons are strongly locked in covalent bonds, it is an electrical insulator.

#### Characteristics of Covalent Crystals:

- (1) Covalent bond is a strong bond. Cohesive energies of 6 to 12 eV/atom are typical of covalent crystals, which is more than the usual cohesive energies in ionic crystals. All covalent crystals are hard, have high melting points, and are insoluble in all ordinary liquids.
- (2) Covalent bonds are *strongly directional*.
- (3) The other important characteristic property of the covalent bond is its *saturability*. Saturability means that each atom can form covalent bonds only with a limited number of its neighbours.
- (4) The conductivity of covalent crystals varies over a wide range. Some crystals are insulators (diamond) and some are semiconductors (*Ge*). The conductivity increases with the increase of temperature.
- (5) The optical properties of the covalent crystals are characterized by high refractive index and high dielectric constant. Covalent crystals are transparent to long-wavelength radiation but opaque to shorter wavelengths.

### 25.29 METALLIC BOND

In metallic crystals, the metallic bond arises when all of the atoms share all of the valence electrons. The valence electrons of the atoms comprising a metal are common to the entire aggregate, so that a kind of "gas" of free electrons pervades it. The crystal is held together by the electrostatic attraction between the negative electron gas and the positive metal ions. The best example for a metallic crystal is sodium.

The cohesion (*i.e.*, the ability to remain a solid) of the metallic crystal results from a combination of forces :

- (i) the attraction of the electron cloud for the ion cores,
- (ii) the mutual repulsion of the electrons, and
- (iii) the mutual repulsion of the ion cores.

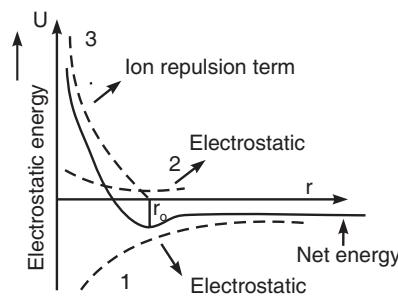
Fig. 25.82 shows a sketch of a reasonable form for the potential (free) energy contribution, as a function of nuclear spacing, from each of these charge interactions.

The following are the characteristics of metal crystals:

(1) The presence of free electrons accounts for the high electrical and thermal conductivities of metals. The high electrical conductivity of metals is in turn responsible for their high optical reflection and absorption coefficients.

(2) The other characteristic properties of metals are their ductility and metallic lustre. Since the metallic bonds are not localised between adjacent atoms, the atoms of a metal can be rearranged in position without rupturing the crystal. This explains ductility of metals. When light shines on a metal, the free electrons oscillate under the electromagnetic field of the incident light and become sources of light. This gives the metal its surface lustre.

(3) Metallic bonds are weaker than ionic and covalent bonds.



**Fig. 25.82**

### 25.30 HYDROGEN BOND

Hydrogen bond is formed under certain conditions when a single hydrogen atom appears to be bonded to two distinct electronegative atoms. This configuration is represented as  $X-H \dots Y$  where  $X$  is called 'donor' while  $Y$  is called 'acceptor'. The weaker of the two bonds (shown by dotted line) is the hydrogen bond while the other (shown by a full line) is a strong covalent bond.  $H_2O$  (ice),  $NH_3$  and  $HF$  are examples of hydrogen-bonded crystals. These bonds are stronger than van der Waals bonds but weaker than ionic or covalent bonds.

**Table 25.3. Types of crystalline Solids**

Type	Ionic	Covalent	Molecular	Metallic
Bond	Electric attraction	Shared electrons	Van der Waals forces	Electron gas
Properties	Hard; high melting points; may be soluble in polar liquids such as water; electrical insulators (but conductors in solution)	Very hard; high melting points; insoluble in nearly all liquids; semiconductors (except diamond, which is an insulator)	Soft; low melting and boiling points; soluble in covalent liquids; electrical insulators.	Ductile; metallic luster, high electrical and thermal conductivity

Example	Sodium chloride, NaCl $E_{\text{cohesive}} = 3.28 \text{ eV/atom}$	Diamond, C $E_{\text{cohesive}} = 7.4 \text{ eV/atom}$	Methane, CH <sub>4</sub> $E_{\text{cohesive}} = 0.1 \text{ eV/molecule}$	Sodium, Na $E_{\text{cohesive}} = 1.1 \text{ eV/atom}$
	<ul style="list-style-type: none"> <li>The cohesive energy is the work needed to remove an atom (or molecule) from the crystal and so indicates the strength of the bonds holding it in place.</li> </ul>			

## INTERATOMIC POTENTIAL

### 25.31 INTERATOMIC POTENTIAL

The potential between a pair of atoms is called the *interatomic potential*.

From the very existence of solids we draw the following two general conclusions:

- There must act attractive forces between the atoms or molecules in a solid which keep them together.
- There must be repulsive forces acting between the atoms as well, since large external pressures are required to compress a solid to appreciable extent.

Consider a single pair of atoms *A* and *B* which form a stable chemical compound.

Assume that the potential energy of atom *B* due to the presence of atom *A* is given by

$$U(r) = -\frac{\alpha}{r^n} + \frac{\beta}{r^m} \quad \dots(1)$$

Here *r* is the distance between the nuclei of the two atoms.  $\alpha$ ,  $\beta$ ,  $m$  and  $n$  are constants characteristic for the molecule *AB*.

The first term,  $\frac{\alpha}{r^n}$ , which is negative, corresponds to the energy associated with the forces of attraction.

The second term,  $\frac{\beta}{r^m}$ , which is positive, corresponds to the forces of repulsion.

The force between the two atoms *A* and *B* as function of *r* is given by

$$F(r) = -\frac{dU}{dr} = -\frac{n\alpha}{r^{n+1}} + \frac{m\beta}{r^{m+1}} \quad \dots(2)$$

The energy (*U*) and force (*F*) between two atoms *A* and *B* as function of their separation *r* are plotted in Fig. 25.83. Dotted curves are the sums of the attractive and repulsive curves.

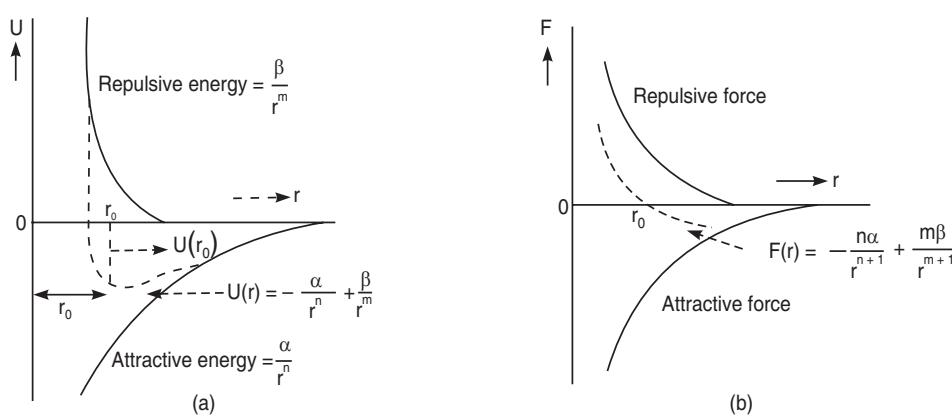


Fig. 25.83

The attraction between two atoms brings them closer together until the individual electron clouds begin to overlap and a strong repulsive force arises to comply with Pauli's exclusion principle. When the attractive force and the repulsive force between any two atoms are equal, the two atoms should be in a stable situation with a minimum potential energy corresponding to the equilibrium interatomic distance,  $r_0$ .

- In solids, forces between the atoms are of two types: attractive to prevent them from moving away from each other; and repulsive to prevent them from collapsing. These forces are equal at equilibrium separation (minimum potential energy configuration).

At the equilibrium separation  $r = r_0$ , the potential energy must be a minimum and the first derivative of Eq. (1) must vanish.

$$\frac{dU}{dr} \Big|_{r=r_0} = 0 = \frac{n\alpha}{r^{n+1}} - \frac{m\beta}{r^{m+1}} \quad \dots(3)$$

$$\therefore \beta = \alpha \cdot \frac{n}{m} (r_0)^{m-n} \quad \dots(4)$$

and

$$(r_0)^{m-n} = \left(\frac{\beta}{\alpha}\right) \left(\frac{m}{n}\right) \quad \dots(5)$$

According to Eq. (2), this condition is equivalent to the requirement that the attractive and repulsive forces balance, i.e.,  $F(r_0) = 0$ .

From Eqs. (1) and (4), we obtain for the energy in the equilibrium state

$$\begin{aligned} U(r_0) &= -\frac{\alpha}{r_0^n} + \frac{\beta}{r_0^m} \\ \therefore U(r_0) &= -\frac{\alpha}{r_0^n} \left(1 - \frac{n}{m}\right) \end{aligned} \quad \dots(6)$$

The energy corresponding to  $r = r_0$  position, denoted by  $U(r_0)$  is called the binding energy or the energy of the cohesion of the molecule. This is the energy required to remove the two atoms of the molecule ( $AB$ ) into an infinite separation.

**Dissociation Energy.** The positive quantity  $D = -U(r_0)$  is the *dissociation energy* of the molecule, i.e., the energy required to separate the two atoms. Dissociation may occur, for example, at high temperatures or as a result of other processes in which the molecule can absorb sufficient energy. The dissociation energies are of the order of 1 eV to a few eV.

A minimum in the energy curve is possible only if  $m > n$ .

Thus the formation of a chemical bond requires that the repulsive forces be of shorter range than the attractive forces.

This may be shown by employing the condition for energy minimum,

$$\frac{d^2U}{dr^2} \Big|_{r=r_0} > 0 \quad \dots(7)$$

This condition leads to

$$\frac{-n(n+1)\alpha}{r_0^{n+2}} + \frac{m(m+1)\beta}{r_0^{m+2}} > 0 \quad \dots(8)$$

Substituting the value of  $r_0$  from Eq. (5) in Eq. (8), we obtain

$$m > n. \quad \dots(9)$$

**EXAMPLE 1.** If the potential energy function is expressed as

$$U(r) = -\frac{\alpha}{r^6} + \frac{\beta}{r^{12}}$$

show that

(a) The intermolecular distance  $r_0$  for which the potential energy is minimum is given by

$$\left(\frac{2\beta}{\alpha}\right)^{1/6}.$$

(b) The minimum potential energy is given by,  $U_{min} = -\frac{\alpha^2}{4\beta}$ .

**SOL.** (a) The condition for the minimum in the potential energy is

$$\left(\frac{dU}{dr}\right)_{at r=r_0} = 0$$

$$\therefore \left(\frac{dU}{dr}\right)_{r=r_0} = 0 = \left[\frac{6\alpha}{r^7} - \frac{12\beta}{r^{13}}\right]_{r=r_0}$$

or  $\frac{6\alpha}{r_0^7} = \frac{12\beta}{r_0^{13}}$  at  $r = r_0$

or  $r_0 = \left(\frac{2\beta}{\alpha}\right)^{1/6}$  ... (1)

(b) The minimum potential energy is

$$\begin{aligned} U_{min} &= \left[-\frac{\alpha}{r_0^6} + \frac{\beta}{r_0^{12}}\right] \\ &= -\frac{\alpha}{r_0^6} \left[1 - \frac{\beta}{\alpha} \cdot \frac{1}{r_0^6}\right] \\ &= -\frac{\alpha}{\left(\frac{2\beta}{\alpha}\right)^{1/6}} \left[1 - \frac{\beta}{\alpha} \cdot \frac{\alpha}{2\beta}\right] \quad \text{using Eq. (1)} \\ U_{min} &= -\frac{\alpha^2}{2\beta} \left[1 - \frac{1}{2}\right] = -\frac{\alpha^2}{4\beta} \end{aligned}$$

## DEFECTS IN CRYSTALS

### 25.32 INTRODUCTION

In an *ideal crystal*, the atomic arrangement is perfectly regular and continuous throughout. Thus an ideal crystal is a perfect one in all respects.

**Real crystals** usually contain several types of imperfections, such as substitutional and interstitial atoms, as well as vacancies or holes. Dislocations and surface defects are also usually present in crystals. The mechanical, electrical and magnetic properties of engineering crystalline solids, particularly metals and alloys, are profoundly affected by the imperfections in the crystals.

The properties of solids such as, strength of crystals, luminescence, colour of crystals, diffusion, crystal growth, plasticity etc., cannot be explained on the basis perfect periodicity concept. Such properties are greatly affected by the lattice defects or atomic imperfections arising due to irregularities of atomic arrays in crystals. These properties are called defect or structure sensitive properties.

The term imperfection or defect is used to describe any deviation from the perfect periodic array of atoms in the crystals.

#### **Classification of Crystal Imperfections (or Defects)**

Crystalline imperfections can be classified on the basis of their geometry under the following four main divisions.

1. POINT DEFECTS (zero dimensional defects)
 

(a) Vacancies	(b) Interstitialies
(c) Impurities	(d) Electronic defects
2. LINE DEFECTS (one dimensional defects)
 

(a) Edge dislocation	(b) Screw dislocation
----------------------	-----------------------
3. SURFACE DEFECTS (two dimensional defects)
 

(a) Grain boundaries	(b) Tilt boundaries
(c) Twin boundaries	(d) Stacking faults
(e) Ferromagnetic domain walls	
4. VOLUME DEFECTS (three dimensional defects)
 

(a) Large voids	(b) Cracks
-----------------	------------

#### **25.33 POINT DEFECTS**

Point imperfections are also called zero dimensional imperfections. They are imperfect point-like regions in the crystal. One or two atomic diameters is the typical size of a point imperfection. In a crystal lattice, point defect is one which is completely local in its effect, e.g., a vacant lattice site.

The introduction of point defect into a crystal increases its internal energy as compared to that of the perfect crystal.

(i) If the point defect is a vacancy, then there is no bonding. Hence the value of the mechanical strength at that point is reduced.

(ii) If an impurity atom is present instead of the original atom, in the lattice, there must be some strain due to the different size of the impurity atom.

\*The number of defects at equilibrium at a certain temperature can be determined from the following equation :

$$n_d = Ne^{-E_d/kT}$$

Here,  $n_d$  is the number of defects,

$N$  is the total number of atomic sites per cubic metre or per mole,

$E_d$  is the energy of activation necessary to form the defect,

$k$  is the Boltzmann constant, and

$T$  is the absolute temperature.

### 25.33.1. Vacancies

An empty site of an atom in a crystal is called vacancy.

A vacancy implies an unoccupied atom position within a crystal lattice (Fig. 25.84).

The atoms surrounding a vacancy tend to be closer together, thereby distorting the lattice planes.

Vacancies may occur as a result of imperfect packing during the original crystallization or they may arise from **thermal vibrations** of atoms at elevated temperatures, because as thermal energy is increased there is a higher probability that individual atoms will jump out of their position of lowest energy.

It may be shown by thermodynamic reasoning that lattice vacancies are a stable feature of metals at all temperatures above absolute zero.

The everyday industrial processes of annealing, homogenization, precipitation, sintering, surface hardening, oxidation and creep, all involve to varying degrees, the transport of atoms through the lattice with the help of vacancies.

Vacancies exist in a certain proportion in a crystal in thermal equilibrium, leading to an increase in the randomness of the structure.

**Schottky defect.** Schottky defect is closely related to vacancies. It is formed when an atom or an ion is removed from a normal lattice site and replaced in an average position on the surface of the crystal (Fig. 25.85). In other words, when vacancies are created by movements of atoms from positions inside the crystal to positions on the surface of the crystal, a *Schottky defect* is said to have been formed.

#### Schottky defects in Ionic crystals

In ionic crystals, formation of point imperfections is subject to the requirement that the overall electrical neutrality is maintained. A pair of one cation and one anion can be missing from an ionic crystal as shown in Fig. 25.86.

When vacancies are created by movements of one anion and one cation from positions inside the crystal to positions on the surface of the crystal, a Schottky defect is said to have been formed.

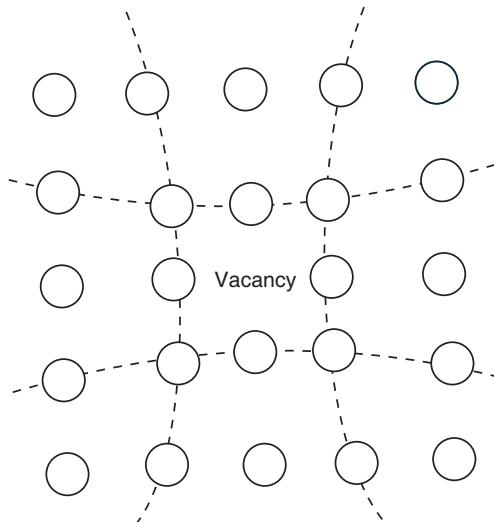


Fig. 25.84

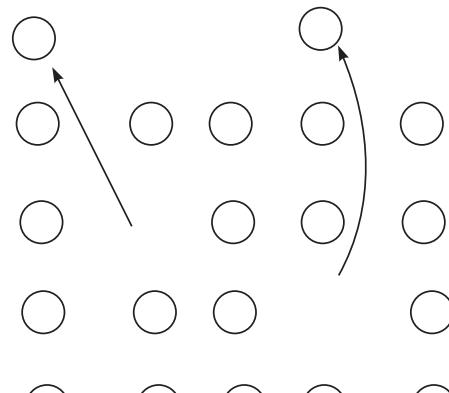


Fig. 25.85

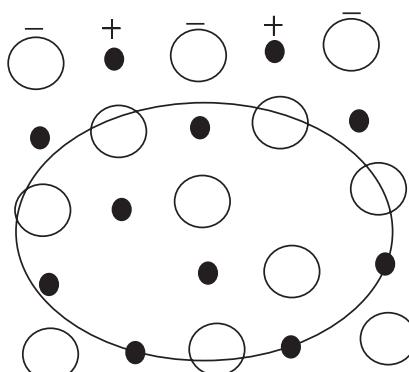


Fig. 25.86

The valency of the missing pair of ions should be equal to maintain the electrical neutrality. So these are normally generated in equal numbers of anion and cation vacancies in a crystal.

Schottky defects are dominant in alkali halides.

### 25.33.2. Interstitialcies

An interstitial defect arises when an atom occupies a definite position in the lattice that is not normally occupied in the perfect crystal.

In interstitialcies, atoms occupy vacant positions between the atoms of the ideal crystal (Fig. 25.87).

The interstitial (atom) may be either a normal atom of the crystal or a foreign atom.

Interstitialcy produces atomic distortion because interstitial atom tends to push the surrounding atoms farther apart, unless the interstitial atom is smaller than the rest of the atoms in the crystal.

**Frenkel Defect.** *Frenkel defect* is closely related to interstitialcies. An ion displaced from the lattice into an interstitial site is called a Frenkel defect (Fig. 25.88). When the defect migrates through the crystal, it does so by a sequence of jumps of the kind shown in Fig. 25.88, rather than by the migration of a single ion from one interstitial position to another.

#### Frenkel defect in Ionic Crystals

In the case of ionic crystals, an ion displaced from the lattice into an interstitial site is called a *Frenkel defect*.

As cations are generally the smaller ions, it is possible for them to get displaced into the void space present in the lattice. Anions do not get displaced like this, as the void space is just too small for their size. A Frenkel imperfection does not change the overall electrical neutrality of the crystal.

Theoretical and experimental evidence shows that Frenkel defects are most commonly found in silver halides. The density of crystal does not change due to the presence of Frenkel defects because, to the first order of approximation, there is no net change of volume of the crystal.

The Schottky and Frenkel defects in an ionic crystal are shown in Fig. 25.89.

The arrows show how the ions are displaced. In Schottky defect *S*, the ion is removed to the surface of the crystal. In a Frenkel defect *F*, the ion moves to such an interstitial position which is not a normal position of any ion.

The Frenkel and Schottky defects together are called '*Intrinsic defects*'.

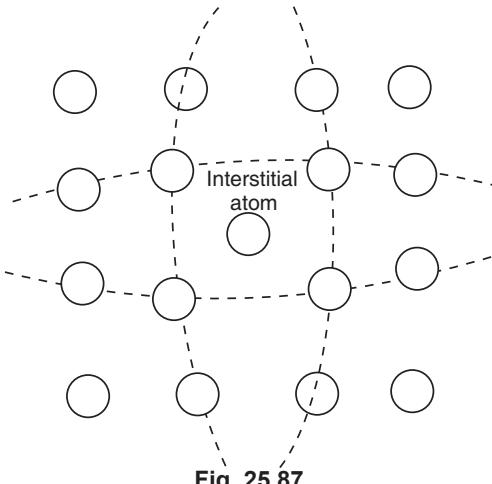


Fig. 25.87

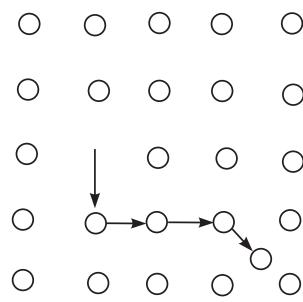


Fig. 25.88

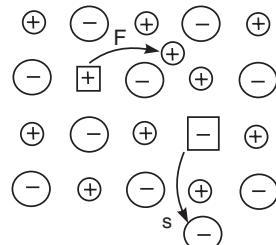


Fig. 25.89

### 25.33.3. Impurities

Impurities give rise to compositional defects. Impurities may be small particles (such as slag inclusions in metals) embedded in the structure, or foreign (metal) atoms in the lattice. Foreign atoms generally have atomic radii and electronic structures differing from those of the host atoms and therefore act as centres of distortion. Impurities may considerably distort the lattice.

There are two types of impurity defects.

(i) **Substitutional impurity.** A substitutional impurity refers to a foreign atom that substitutes for or replaces a parent atom in the lattice (Fig. 25.90).

**EXAMPLES.** 1. In ionic solids (e.g., in NaCl), the substitution of  $\text{Na}^+$  by  $\text{Li}^+$  produces a substitutional impurity.

2. In semiconductor technology, aluminium and phosphorus doped in silicon are substitutional impurities in the crystal.

A controlled addition of impurity to a very pure crystal is the basis of producing many electronic devices.

3. During the production of brass alloy, the zinc atoms are doped in copper lattice. Here the zinc atoms are called substitutional impurities.

#### (ii) Interstitial impurity

An interstitial impurity is a small sized atom occupying the void space in the parent crystal, without dislodging any of the parent atoms from their sites (Fig. 25.91).

An atom can enter the interstitial or void space only when it is substantially smaller than the parent atom.

**EXAMPLE.** In FCC iron, the atomic radius of iron atom is 0.225 nm. Carbon atoms with atomic radius 0.0777 nm can occupy the octahedral void spaces in FCC lattice as interstitial impurities.

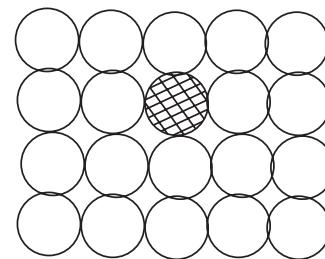


Fig. 25.90

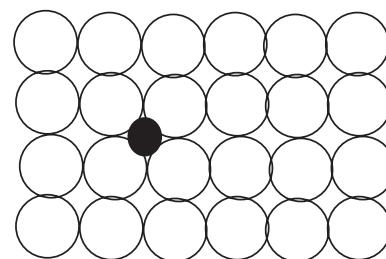


Fig. 25.91

### 25.33.4. Electronic Defects

Electronic defects are the result of errors in charge distribution in solids.

These defects are free to move in the crystal under the influence of an electrical field. This accounts for some electronic conductivity of certain solids and their increased reactivity.

**EXAMPLE.** In zinc oxide ( $\text{ZnO}$ ), the zinc ions occupy interstitials. This leads to a large number of positive charges at interstitials. Suppose in some places zinc ions are missing. Then, at these places, there is a gain of negative charges due to loss of positive charges. Thus a vacancy or an interstitial impurity may produce an excess or deficiency of positive or negative charges.

**25.34**

### TO DERIVE AN EXPRESSION FOR THE NUMBER OF SCHOTTKY DEFECTS AT A GIVEN TEMPERATURE

Fig. 25.92 shows two kinds of Schottky defects most commonly found in crystals, namely, the individual positive and negative vacancy, presumed to form near each other, and a coupled pair.

Consider a perfect crystal composed of equal numbers of positively and negatively charged ions. In order for a cation vacancy to occur, a positive ion must somehow migrate out of its proper position

in the structure to the crystal's exterior. If only positive ions migrate out of the crystal and collect on its surface, the surface will become positively charged. This positive surface charge opposes the migration of additional positive ions out of the crystal's interior. Simultaneously, the excess negative charge created inside the crystal is conducive to the formation of negative vacancies. In the absence of external forces, therefore, the number of oppositely charged vacancies inside a crystal tends to be equal.

Suppose that the crystal contains a total of  $N$  atoms.

Suppose that  $n$  Schottky defects are produced by removing  $n$  cations and  $n$  anions from the crystal's interior.

The different ways in which each kind of ion can be removed is given by

$$\frac{N(N-1)(N-2)\dots(N-n+1)}{n!} = \frac{N!}{(N-n)!n!}. \quad \dots(1)$$

The number of cation and anion vacancies are equal.

So different ways in which  $n$  Schottky defects can be formed is obtained by squaring the expression in Eq. (1).

$$\therefore P = \left[ \frac{N!}{(N-n)!n!} \right]^2 \quad \dots(2)$$

The creation of  $n$  Schottky defects increases the crystal's entropy, according to the Boltzmann relation, by an amount

$$S = k \ln \left[ \frac{N!}{(N-n)!n!} \right]^2 \quad \dots(3)$$

Let  $E_p$  be the energy required to remove a pair of atoms from the crystal's interior to the surface.

Total change in internal energy =  $E = nE_p$ .

Here,  $n$  is the number of vacancy pairs.

This, in turn, produces a change in Helmholtz free energy

$$\begin{aligned} F &= E - TS \\ &= nE_p - kT \ln \left[ \frac{N!}{(N-n)!n!} \right]^2 \end{aligned} \quad \dots(4)$$

The logarithmic term in Eq. (4) containing factorials can be simplified by using Stirling's formula,  $\ln x! = x \ln x - x$ .

$$\begin{aligned} \ln \left[ \frac{N!}{(N-n)!n!} \right]^2 &\simeq 2[\ln N! - \ln (N-n)! - \ln n!] \\ &= 2[N \ln N - N - (N-n) \ln (N-n) + (N-n) - n \ln n + n] \\ &= 2[N \ln N - (N-n) \ln (N-n) - n \ln n]. \end{aligned} \quad \dots(5)$$

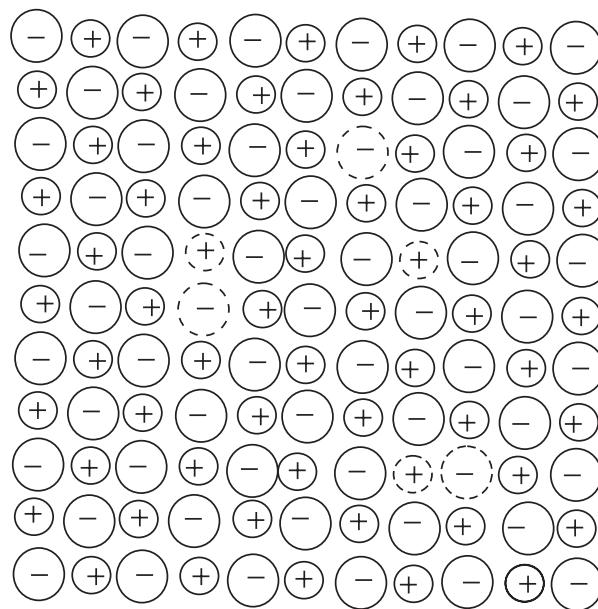


Fig. 25.92

When equilibrium is attained at a given temperature  $T$ , the Helmholtz free energy is constant and its first derivative is, therefore, equal to zero.

$$\begin{aligned} \left( \frac{\partial F}{\partial n} \right)_T &= 0 = E_p - 2kT [\ln(N-n) - \ln n] \\ &= E_p - 2kT \ln \frac{N-n}{n} \end{aligned} \quad \dots(6)$$

Here the partial differentiation is with respect to  $n$  since the total number of atomic positions in the crystal  $N$  is not altered.

$$E_p = 2kT \ln \frac{N-n}{n} \quad \dots(7)$$

$$\text{or } \frac{N-n}{n} = e^{\frac{E_p}{2kT}} \quad \dots(8)$$

The number of Schottky defects in a crystal is much smaller than the number of atoms; that is,  $n \ll N$  and  $N-n \approx N$ .

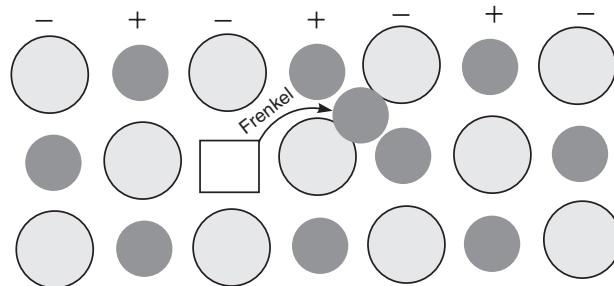
Eq. (8) reduces to

$$n \approx Ne^{-E_p/2kT} \quad \dots(9)$$

Eq. (9) is used to determine the approximate number of defects present at any temperature.

### 25.35 TO DERIVE AN EXPRESSION FOR THE NUMBER OF FRENKEL DEFECTS AT A GIVEN TEMPERATURE

In Frenkel defect, an atom is transferred from a lattice site to an *interstitial position*, a position not normally occupied by an atom (Fig. 25.93).



**Fig. 25.93**

In a perfect crystal, the energy required to displace an atom from its proper position to an interstitial position is  $E_i$ .

Suppose there are  $N$  atoms in the crystal and  $N_i$  interstitial positions in its structure.

The total number of ways in which  $n$  Frenkel defects can be formed is given by

$$P = \frac{N!}{(N-n)!n!} \frac{N_i!}{(N_i-n)!n!} \quad \dots(1)$$

The corresponding increase in entropy due to the creation of Frenkel defects is given by

$$S = k \ln \left\{ \left[ \frac{N!}{(N-n)!n!} \right] \times \left[ \frac{N_i!}{(N_i-n)!n!} \right] \right\}$$

The amount of energy required to produce  $n$  Frenkel defects is

$$E = nE_i$$

The change in the Helmholtz free energy produced by the creation of  $n$  Frenkel defects is

$$F = E - TS$$

$$\therefore F = nE_i - kT \ln \frac{N!}{(N-n)!n!} \frac{N_i!}{(N_i-n)!n!} \quad \dots(2)$$

The factorial terms are simplified by using Stirling's formula

$$\ln x! = x \ln x - x.$$

$$\begin{aligned} \ln \frac{N!}{(N-n)!n!} + \ln \frac{N_i!}{(N_i-n)!n!} &\approx N \ln N + N_i \ln N_i \\ &\quad - (N-n) \ln (N-n) - (N_i-n) \ln (N_i-n) - 2n \ln n. \end{aligned} \quad \dots(3)$$

Substituting (3) into (2), we get

$$F = nE_i - kT [N \ln N + N_i \ln N_i - (N-n) \ln (N-n) - (N_i-n) \ln (N_i-n) - 2n \ln n] \quad \dots(4)$$

Differentiating Eq. (4) with respect to  $n$ ,

$$\left( \frac{\partial F}{\partial n} \right)_T = E_i - kT \ln \frac{(N-n)(N_i-n)}{n^2} \quad \dots(5)$$

At equilibrium  $(\partial F / \partial n)_T = 0$ .

$$\therefore E_i = kT \ln \left[ \frac{(N-n)(N_i-n)}{n^2} \right]$$

Now,  $N \gg n$  and  $N_i \gg n$ .

$$\begin{aligned} \therefore \frac{E_i}{kT} &\approx \ln \frac{NN_i}{n^2} \\ &= \ln (NN_i) - 2 \ln n \end{aligned}$$

Rearranging the terms, we have

$$\begin{aligned} \ln n &= \frac{1}{2} \ln (NN_i) - \frac{E_i}{2kT} \\ \therefore n &= (NN_i)^{\frac{1}{2}} e^{-E_i/2kT} \end{aligned} \quad \dots(6)$$

In silver halides, the most prevalent defects below 700 K are Frenkel defects.

### 25.36 TO DERIVE AN EXPRESSION FOR THE NUMBER OF VACANCIES AT A GIVEN TEMPERATURE

In all crystals vacancies are present. The main cause for these defects is thermal agitation. Suppose  $E_v$  is the energy required to take an atom from a lattice site inside the crystal to a lattice site on the surface. Suppose there are  $N$  atoms.

The amount of energy required to produce  $n$  number of isolated vacancies is

$$E = nE_v \quad \dots(1)$$

The total number of ways in which  $n$  number of atoms can be removed out of  $N$  number of atoms in a crystal on to the surface is

$$P = \frac{N!}{(N-n)!n!} \quad \dots(2)$$

The increase in entropy due to formation of  $n$  vacancies is

$$S = k \ln P = k \ln \left\{ \frac{N!}{(N-n)!n!} \right\}$$

The change in the Helmholtz free energy produced by the creation of  $n$  vacancies is

$$\begin{aligned} F &= E - TS \\ \therefore F &= nE_v - kT \ln \left\{ \frac{N!}{(N-n)!n!} \right\} \end{aligned}$$

$$F = nE_v - kT [\ln N! - \ln (N-n)! - \ln n!]$$

The factorial terms are simplified by using Stirling approximation,

$$\ln x! = x \ln x - x.$$

$$\begin{aligned} F &= nE_v - kT [N \ln N - N - (N-n) \ln (N-n) + (N-n) - n \ln n + n] \\ F &= nE_v - kT [N \ln N - (N-n) \ln (N-n) - n \ln n] \end{aligned} \quad \dots(3)$$

At thermal equilibrium, free energy is constant and minimum.

Differentiating Eq. (3) with respect to  $n$ ,

$$\begin{aligned} \left( \frac{\partial F}{\partial n} \right)_T &= 0 = E_v - kT \left[ 0 - (N-n) \frac{-1}{(N-n)} - (-1) \ln (N-n) - n \frac{1}{n} - 1 \ln n \right] \\ E_v &= kT [1 + \ln (N-n) - 1 - \ln n] \\ E_v &= kT \ln \left[ \frac{(N-n)}{n} \right] \\ \frac{N-n}{n} &= \exp \left( \frac{E_v}{kT} \right) \\ n &\equiv (N-n) \exp \left[ \frac{-E_v}{kT} \right] \end{aligned}$$

If  $n \ll N$ ,  $n$  can be neglected.

$$\therefore n \equiv N \exp \left[ \frac{-E_v}{kT} \right]$$

If

$E_v = 1\text{eV}$  and  $T = 1000\text{ K}$ , then

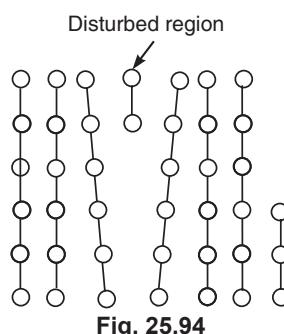
$$\frac{n}{N} = e^{-11.6} = 10^{-5}.$$

The equilibrium concentration of vacancies decreases as the temperature decreases.

### 25.37 LINE DEFECTS

Line defects are called dislocations. These are one-dimensional imperfections in the geometrical sense. A *dislocation* may be defined as a disturbed region between two substantially perfect parts of a crystal (Fig. 25.94).

Dislocation is a line defect in a crystal structure whereby a part-plane of atoms is displaced from its symmetrically stable positions in the array. It is surrounded within the structure by an extensive elastic strain field and its associated stresses.



There are two basic types of dislocations :

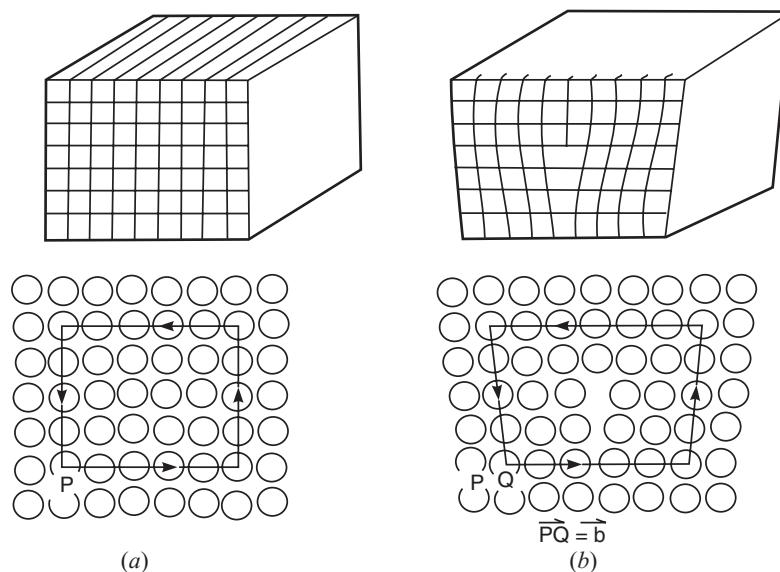
1. Edge dislocation, and
2. Screw dislocation

### 25.37.1. Edge Dislocation

Fig. 25.95 (a) shows a perfect crystal.

The top sketch shows a three dimensional view of a perfect crystal.

The bottom sketch shows the arrangement of atoms on its front face.



**Fig. 25.95**

Fig. 25.95 (b) shows an imperfect crystal.

The top sketch shows a three dimensional view of an imperfect crystal with an edge dislocation.

The bottom sketch shows the arrangement of atoms on its front face. The atoms above the edge of the extra plane are squeezed together and are in a state of compression. The bond lengths have been compressed to smaller than the equilibrium value. Just below the edge of the extraplane, the atoms are pulled apart and are in a state of tension. Here the bond lengths have been stretched to above the normal values. This distorted configuration extends all along the edge into the crystal. There is an extra energy due to the distortion in the region immediately surrounding the edge of the incomplete plane. As the region of maximum distortion is centred around the edge of the incomplete plane, this distortion represents a line imperfection and is called an *edge dislocation*.

The vector  $\vec{b} = \overrightarrow{PQ}$  connecting the end point  $Q$  with the starting point  $P$  is the Burgers vector of the dislocation.

Edge dislocation can be classified as *positive edge dislocation* and *negative edge dislocation*.

An edge dislocation involves an extra row of atoms, either above (positive sign) or below (negative sign) the slip plane (Fig. 25.96).

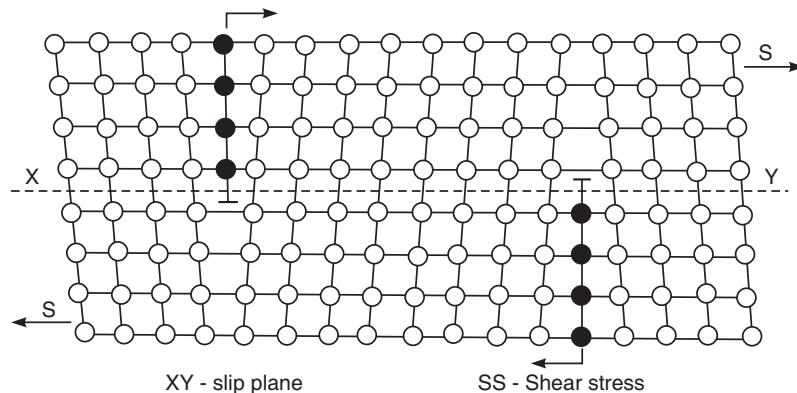


Fig. 25.96

The horizontal line  $XY$  is called the slip plane.

#### (i) Positive dislocation

If an extra plane of atoms is above the line  $XY$ , the edge dislocation is said to be positive. It is denoted by the symbol  $\perp$ .

#### (ii) Negative dislocation

If an extra plane of atoms is below the line  $XY$ , the dislocation is said to be negative. It is denoted by the symbol  $T$ .

### 25.37.2. Movement of $\perp$ and $T$ Dislocations Under Shear Stress

The presence of an extra plane of atoms means that adjacent atoms are displaced elastically. Consequently, from both sides elastic forces are exerted on the dislocation. These forces balance out, so that it is easy to move the dislocation from one position to another.

Under a shear stress sense  $\rightleftharpoons$  a positive dislocation ( $\perp$ ) moves to the right and a negative dislocation ( $T$ ) to the left (Fig. 25.96).

#### Slip caused by the movement of edge dislocation

Fig. 25.97 shows how edge dislocation ( $D$ ) gets glided along the slip plane under the action of shear stress ( $S$ ) and how the slip is caused by the movement of an edge dislocation.

As the dislocation glides out of the crystal completely, it produces a *slip step* of one atom width at the edge of the crystal.

The edge dislocation is particularly useful in explaining slip in plastic flow during mechanical working.

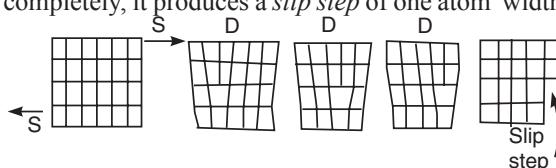


Fig. 25.97

### 25.37.3. Burger's Vector

The Burger's vector indicates how much and in what direction the lattice above the slip plane appears to have been shifted with respect to the lattice below the slip plane. The Burger's vector is perpendicular to the edge dislocation.

**Burger's Vector** marks the magnitude and direction of the strain component of dislocation.

#### Method of determining Burger's Vector for Edge Dislocation

Fig. 25.98 shows the method of determining the Burger's Vector.

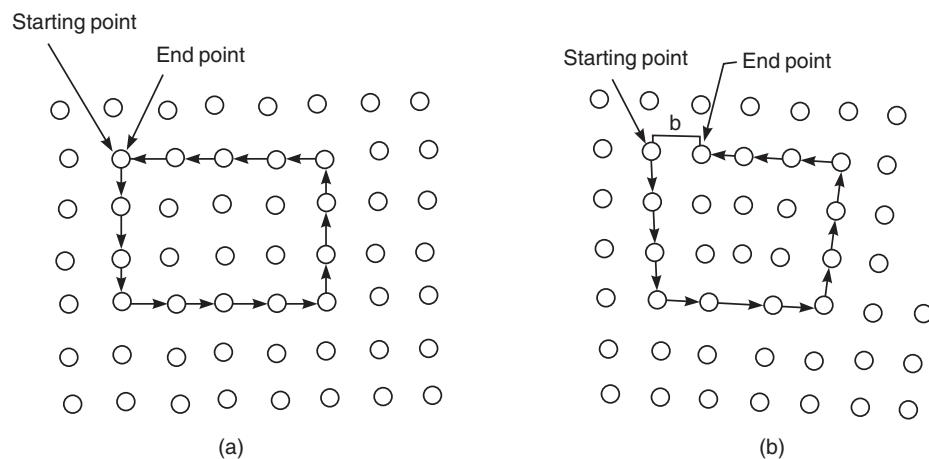


Fig. 25.98

The Burger's circuit is formed by proceeding through the undisturbed region surrounding a dislocation in steps which are integral multiples of a lattice translation. The loop is completed by going an equal number of translations in a positive sense and negative sense in a plane normal to the dislocation line.

(i) The Burger's loop closes upon itself if it does not enclose a dislocation [Fig. 25.98 (a)].

(ii) The Burger's circuit surrounding an edge dislocation is shown in Fig. 25.98 (b). The end point does not coincide with the starting point.

Therefore, the vector  $\vec{b}$  connecting the end point with the starting point is the Burger's Vector of the dislocation.

#### 25.37.4. Screw Dislocation

Screw dislocation results from a displacement of the atoms in one part of a crystal relative to the rest of the crystal, forming a spiral ramp around the dislocation line.

Fig. 25.99 shows what happens when one part of the crystal is displaced relative to the rest of the crystal and the displacement terminates within the crystal.

The row of atoms marking the termination of the displacement is the screw dislocation.  $E F$  indicates the dislocation line.

In screw dislocation, the Burger's vector lies *parallel* to the dislocation line along the axis of a line of atoms in the same plane. The Burger's vector (Fig. 25.100) determines the magnitude and direction of the screw dislocation. The screw dislocation may be thought of as produced by cutting the

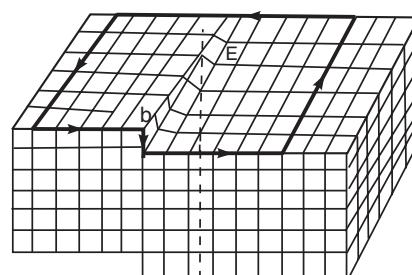


Fig. 25.99

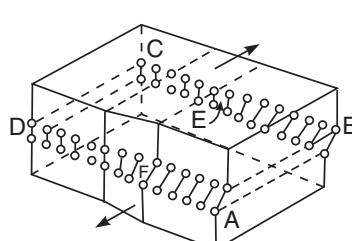


Fig. 25.100

crystal partway through with a knife and shearing it parallel to the edge of the cut by one atom spacing. A screw dislocation transforms successive atom planes into the surface of a helix.

A screw dislocation is sketched in Fig. 25.100.

A part *ABEF* of the slip plane has slipped in the direction parallel to the dislocation line *EF*.

A screw dislocation may be visualized as a helical arrangement of lattice planes, such that we change planes on going completely around the dislocation line.

Fig. 25.101 shows another view of a screw dislocation. The broken vertical line that marks the dislocation is surrounded by strained material.

#### 25.37.5. Differences between Edge Dislocation and Screw Dislocation

The following Table gives the differences between edge dislocation and screw dislocation.

<b>Edge Dislocation</b>	<b>Screw Dislocation</b>
1. An edge of an atomic plane is formed internal to the crystal.	Only a distortion of the lattice cells in the immediate vicinity is produced.
2. The Burger's vector of an edge dislocation is perpendicular to the dislocation line and lies in the slip plane.	The Burger's vector of a screw dislocation is parallel to the dislocation line.
3. An edge dislocation moves (in its slip plane) in the direction of the Burger's vector (slip direction).	A screw dislocation moves (in the slip plane) in a direction perpendicular to the Burger's vector (slip direction).
4. An edge dislocation involves an extra row of atoms either above (positive sign) or below (negative sign) the slip plane. Symbols : $\perp$ (or) $T$ .	In the screw dislocation, the distortion follows a helical or screw path. The pitch of the screw may be left-handed or right-handed. Symbols : $\curvearrowleft$ or $\curvearrowright$ .
5. The forces required to form and to move an edge dislocation are smaller in their values.	The forces required to form and move a screw dislocation are somewhat greater than those required to initiate an edge dislocation.
6. Speed of movement of an edge dislocation is greater than that of a screw dislocation.	Speed of movement of a screw dislocation is less than that of an edge dislocation.
7. The edge dislocation is particularly useful in explaining slip in plastic flow during mechanical working.	Screw dislocation is especially useful in explaining crystal growth as well as slip in plastic deformation.

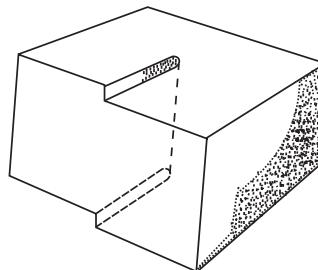


Fig. 25.101

### 25.38 SURFACE DEFECTS

In a surface defect, the crystalline irregularity extends in two dimensions. Surface defects are also called plane defects.

Surface imperfections refer to regions of distortions that lie about a surface having a thickness of a few atomic diameters.

Surface imperfections of a structural nature arise from a change in the stacking of atomic planes on or across a boundary. The change may be one of the orientations or of the stacking sequence of the planes. There are two types of surface imperfections.

1. External surface imperfections.
2. Internal surface imperfections.

### 25.38.1. External Surface imperfections

The external surface of a crystal is an imperfection in itself, as the atomic bonds do not extend beyond the surface. Although we may visualise an external surface as simply a terminus of the crystal structure, the atoms on the surface cannot be compared with the atoms within a crystal. The surface atoms have neighbours on one side only, while atoms inside the crystal have neighbours on either side of them (Fig. 25.102). Since the external surface atoms are not entirely surrounded by others, they possess higher energy than that of internal atoms. This energy of the surface atom, for most metals, is of the order  $1\text{J/m}^2$ .

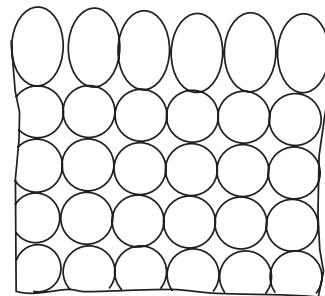


Fig. 25.102

### 25.38.2. Internal Surface Imperfections

Internal surface imperfections arise from a change in the stacking of atomic planes across a boundary. The change may be one of the orientation or of the stacking sequence of the planes.

The important internal surface imperfections are given below :

- (i) Grain boundaries
- (ii) Tilt boundaries and twist boundaries
- (iii) Twin boundaries
- (iv) Stacking faults and
- (v) Ferromagnetic domain walls.

### 25.38.3. Grain Boundaries

Most of the engineering materials are polycrystalline in nature. Grain boundaries are those surface imperfections which separate crystals or grains of different orientation in a polycrystalline aggregation during nucleation or crystallisation. During solidification or during recrystallization of polycrystalline crystals, new crystals are randomly oriented with respect to one another. They grow by the addition of atoms from the adjacent regions and eventually impinge on each other. When two crystals impinge in this manner, the atoms that are caught in between the two crystals are being pulled by each of the two crystals to join its own configuration. They can join neither crystal due to the opposing forces. Therefore, they take up a compromise position. The thickness of this region is only a few atomic diameters, because the opposing forces from neighbouring crystals are felt by the intervening atoms only at such short distances. The boundary region is called a crystal boundary or a grain boundary. The crystal orientation changes sharply at the grain boundary (Fig. 25.103).

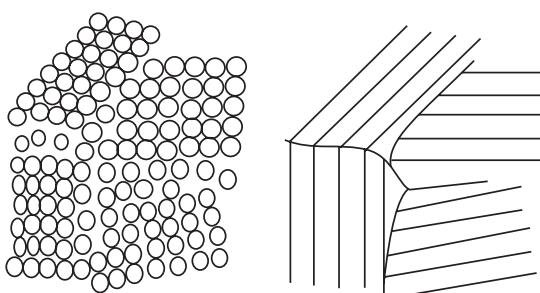


Fig. 25.103

**Definition.** *Grain boundaries are those planar imperfections in polycrystalline materials that separate crystals (grains) of different orientation.*

*A grain boundary is formed when two growing grain surfaces meet.*

For example if we take a piece of iron or copper, it is not in a single crystalline form, but it consists of many small interlocking crystals or grains having random orientations. The boundary between these interlocking crystals must have a structure common to both the orientations and structures of the adjacent crystals. Therefore, the grain boundary forms a discontinuity in the periodicity of the lattice of crystallite or grain. This type of lattice imperfection is called *grain boundary*. The crystal orientation changes sharply at the grain boundary (Fig. 25.103).

The grain boundary has a definite thickness of 2 to 10 or more atomic distances. A grain boundary thickness of a few atoms has been confirmed by the field ion microscopy. Boundaries between grains in commercial metals are often wider, because impurities commonly concentrate at the grain boundaries during solidification.

In grain boundaries, the atomic packing is imperfect. At the grain boundary between two adjacent grains, there is a transition zone which is not aligned with either grain (Fig. 25.104).

Most atoms at the boundaries are located in highly strained and distorted positions and their free energy is higher than that of the atoms in the regular, undisturbed part of the crystal lattice.

**High angle grain boundary.** If the misorientation angle ( $\theta$ ), between the grains is greater than  $10^\circ$ – $15^\circ$ , the resultant grain boundaries are called *High-angle grain boundaries*. These imperfections in the grain boundary enable the microscopist to see them, for in a transparent crystalline material they may scatter light and in opaque material they can be etched chemically.

**Low angle grain boundary.** Low angle grain boundaries have angular misorientation of the order of few degrees but less than  $10^\circ$ .

#### 25.38.4. Tilt Boundaries

Tilt boundaries is another surface imperfection. This is called low-angle boundary as the orientation difference between two neighbouring crystals is less than  $10^\circ$ .

Tilt boundary may be regarded as an array of edge dislocations ( $\perp$ ) (Fig. 25.105).

By rotation of an axis in the boundary it is possible to bring the axis of two bordering grains into coincidence, i.e., a **tilt boundary**.

The angle of tilt is given by

$$\tan \theta = \frac{b}{D} \text{ or } \theta = \frac{b}{D}$$

(because in tilt boundary, the degree of misorientation of the grain boundary or angle of tilt, i.e.,  $\theta$  is less than  $10^\circ$ ).

$D$  is the dislocation spacing and  $b$  is the length of Burger's vector.

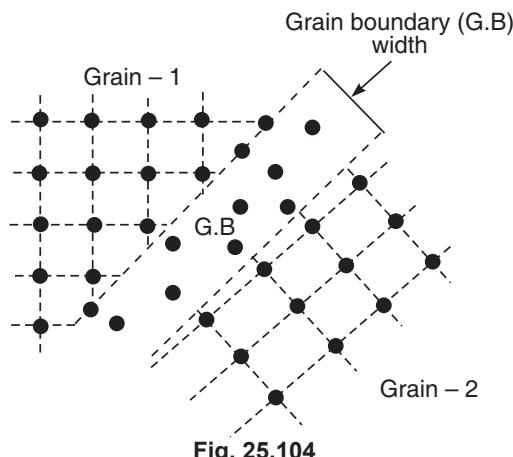


Fig. 25.104

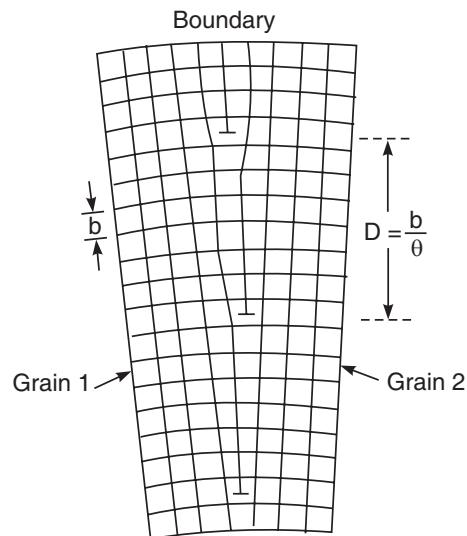


Fig. 25.105

### 25.38.5. Twist Boundaries

Twist boundaries are the second class of low angle boundaries resulting from the set of screw dislocations. In a twist boundary, the rotation is about an axis normal to the boundary.

### 25.38.6. Twin Boundaries

Surface defects which separate two orientations that are mirror images of one another are called *twin boundaries*. Twin boundaries occur in pairs, such that the orientation change introduced by one boundary is restored by the other (Fig. 25.106).

The region between the pair of boundaries is called the twinned region. Twin boundaries are easily identified under an optical microscope. Twins which form during the process of recrystallization are called *annealing twins*. Twins which form during plastic deformation of the material are called *deformation twins*.

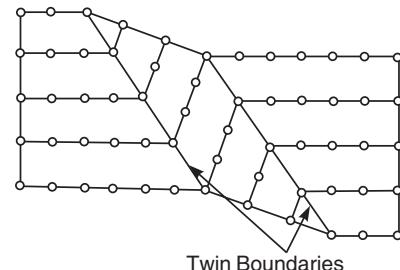


Fig. 25.106

### 25.38.7. Stacking Faults

*A stacking fault is a surface imperfection that results from the stacking of one atomic plane out of sequence on another, while the lattice on either side of the fault is perfect. The stacking fault is a discrepancy in the packing sequence of the layers of atoms (although all the lattice sites are occupied).*

It is usually produced during the growth of the crystals and plastic deformation of crystals.

**Example :** In the case of close packed FCC structure the stacking sequence can be written as ABC, ABC.....In that sequence it is possible in one atom layer ‘A’, the atoms are not positioned properly in a small region and hence deviates from the sequence, relative to the atoms of the layers above and below giving a defect, since now there is a sequence of ....BCBC.....which belongs to HCP structure instead of ....ABC ABC.....

Fig. 25.107 shows the stacking fault in a FCC metal. So one may conclude that stacking fault may arise when there is only small dissimilarity (electrostatically) between the stacking sequences of close-packed planes in FCC and HCP metals.

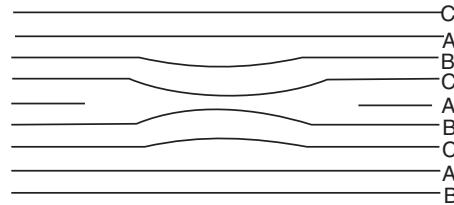


Fig. 25.107

### 25.38.8. Ferromagnetic Domain Walls

When two ferromagnetic regions differ from one another only in the direction of magnetisation, the boundary between them is an imperfection and is called a ferromagnetic domain wall. These domain walls determine the magnetic properties of ferromagnetic materials.

### 25.38.9. Volume Defects

Volume defects such as cracks may arise when there is only small electrostatic dissimilarity between the stacking sequence of close packed planes in metals.

When clusters of atoms are missing, a large vacancy or void is created which is also a volume imperfection.

Foreign particle inclusions, large voids or non-crystalline regions which have the dimensions of at least 1 to 3 nm are also called volume imperfections.

### 25.39 EFFECTS OF CRYSTAL IMPERFECTION

**1.** The addition of impurities in the case of semiconductors increases the electrical conductivity. But in the case of metals, it decreases the electrical conductivity.

#### 2. Colour Centres

The colour of a crystal is due to the selective absorption of some component of the visible spectrum by certain imperfections, called *colour centres*, which are present in the crystal. These imperfections may be interstitial impurity atoms such as transition-metal ions. Alternatively, they may be vacancies in the structure produced by deviations from stoichiometry so that there are present excess positive ions accompanied by negative vacancies or excess negative ions accompanied by positive vacancies. Thus, when excess Zn is present in ZnO it takes on a yellow colour. Excess lithium in LiF turns the crystal pink. Excess potassium in KCl makes the crystal appear violet.

#### 3. The imperfections account for

- (i) Flow and fracture characteristics
- (ii) Crystal growth
- (iii) Diffusion mechanisms
- (iv) Creep characteristics of real metals and alloys
- (v) Annealing and precipitation
- (vi) Oxidation and corrosion
- (vii) Yield strength, fracture strength, plasticity, thermal conductivity, dielectric strength, etc.

#### 4. The defects influence the properties of the solids in the following different ways :

- (i) The defects may scatter conduction electrons in a metal, increasing its electrical resistance by several percent in many pure metals and much more in alloys.
- (ii) Some defects, even if present sparingly, decrease the strength of the crystal.
- (iii) Pure salts having impurities and imperfections are often coloured.
- (iv) Hysteresis loss of ferromagnets.

**5.** Imperfections do not always affect the properties of materials adversely. There are many situations in which controlled amount of imperfection can bring about specific characteristics desired in a material. For example, carbon steel is a material whose properties are improved by imperfections. The interstitial carbon atoms which may be thought of as imperfections in the cubic crystal structure of iron, increase the mechanical strength of iron.

### EXERCISES

#### Choose the Correct Answer (Bharathiar University B.U.)

1. Solids are classified into crystalline solids and \_\_\_\_\_ based on the arrangement of atoms or molecules.  
 (a) conductors      (b) insulators      (c) semiconductors      (d) amorphous solids  
 (B.U. 2011)
2. In tetragonal crystal system, the angle  
 (a)  $\alpha = 90 \neq \beta \neq \gamma$       (b)  $\beta = 90 \neq \alpha \neq \gamma$       (c)  $\alpha = \beta = \gamma = 90$       (d)  $\alpha = \beta = 90 \neq \gamma$   
 (B.U. 2013)
3. In a crystal, the Miller indices of a given plane is 3 : 2 : 1. What are the intercepts?  
 (a) 1a, 2b, 3c      (b) 2a, 3b, 6c      (c) 2a, 1b, 3c      (d) 3a, 2b, 1c  
 (B.U. 2012)

4. Atomic radius of Simple Cubic (SC) cell is  
 (a)  $\frac{a}{2}$       (b)  $\frac{a\sqrt{3}}{4}$       (c)  $\frac{a}{2\sqrt{2}}$       (d)  $\frac{a\sqrt{3}}{8}$       (B.U. 2011)
5. The atomic packing factor for a simple cubic cell is  
 (a)  $\pi$       (b)  $\frac{\sqrt{2}}{6}\pi$       (c)  $\frac{\sqrt{2}}{8}\pi$       (d)  $\frac{\pi}{6}$       (B.U. 2012)
6. Atomic radius of bcc unit cell \_\_\_\_\_  
 (a)  $\frac{\sqrt{3}}{2}a$       (b)  $\frac{\sqrt{3}}{4}a$       (c)  $\frac{2}{\sqrt{3}}a$       (d)  $\frac{4}{\sqrt{3}}a$       (B.U. 2013)
7. The coordination number of body centered cubic crystal is \_\_\_\_\_  
 (a) 6      (b) 8      (c) 12      (d) 10      (B.U. 2015)
8. For unit cell of fcc, the atomic radius \_\_\_\_\_  
 (a)  $r = \frac{\sqrt{2}}{4}a$       (b)  $r = \frac{\sqrt{2}}{6}a$       (c)  $r = \frac{\sqrt{2}}{8}a$       (d)  $r = \frac{\sqrt{2}}{10}a$       (B.U. 2013)
9. In hexagonal crystal, the cell axes are \_\_\_\_\_  
 (a)  $a \neq b = c$       (b)  $a \neq b \neq c$       (c)  $a = b \neq c$       (d)  $a = b = c$       (B.U. 2013)
10. The atomic packing factor of a cadmium crystal is  
 (a) 74%      (b) 52%      (c) 38%      (d) 68%      (B.U. 2010)
11. An example for Hexagonal close packed structure is \_\_\_\_\_  
 (a) Quartz      (b) Lead      (c) Gold      (d) Silver      (B.U. 2015)
12. The atomic packing factor of a diamond cube structure is  
 (a) 78%      (b) 68%      (c) 34%      (d) 52%      (B.U. 2011)
13. The coordination number of a NaCl crystal is  
 (a) 4      (b) 6      (c) 12      (d) 8      (B.U. 2010)
- [Ans. 1. (d), 2. (c), 3. (b), 4. (a), 5. (d), 6. (b), 7. (b), 8. (a), 9. (c), 10. (a), 11. (a), 12. (c), 13. (b)]

### Long Answer Questions

14. Distinguish between crystalline and amorphous solids.      (B.U. 2013)
15. Explain different features of the crystal.      (B.U. 2013)
16. Define the following terms:  
 (a) space lattice      (b) basis      (c) unit cell      (B.U. 2011)
17. With neat diagrams, enumerate the seven crystal systems pointing out their characteristic features.      (B.U. 2015)
18. Explain the crystal classification into 7 systems and 14 Bravais lattices.      (B.U. 2012)
19. What are Miller Indices? Write the procedure for finding Miller Indices of a given plane.      (B.U. 2012)
20. Calculate the Packing factor of BCC structure.      (B.U. 2012)
21. Describe the structure of bcc crystal.      (B.U. 2013)
22. Calculate the atomic packing factor of FCC crystal lattice.      (B.U. 2015)

23. Describe the crystal structure of *fcc* crystal. (B.U. 2013)
24. Calculate the atomic packing factor of HCP structure. (B.U. 2011)
25. Explain the structure of diamond crystal. (B.U. 2011)
26. Explain the following structures with their diagram. (B.U. 2012)  
 (a) NaCl structure      (b) Zinc blende structure  
 (c) Diamond structure
27. Describe the crystal structure of NaCl crystal. (B.U. 2015)
28. Explain the structure of KCl crystal. (B.U. 2011)
29. Draw the structure of Caesium chloride (Madras 2006)
30. Describe the rotating crystal method for crystal structure analysis. What are the advantages of this method? (Meerut 2007)
31. Explain the experimental method of X-ray diffraction studies using powder photograph method. (Madras 2006)
32. What are Laue equations for diffraction of X-rays by a crystalline solid? Show that these equations lead to Bragg's Law for x-ray diffraction. (Raj. 83)
33. What do you mean by reciprocal lattice? How would you realise it? Discuss the properties of reciprocal lattice. Obtain Bragg's condition for reciprocal lattice. (Meerut 1993)
34. Derive the Bragg conditions in terms of the reciprocal lattice vectors. (Raj., 1974)
35. Show that the condition for the X-ray diffraction by a lattice can be written
- $$\vec{2k} \cdot \vec{G} + G^2 = 0$$
- where  $\mathbf{k}$  is the wave vector of the incident beam and  $\mathbf{G}$  is a reciprocal lattice vector. (Meerut M.Sc. 1977)
36. Show that the reciprocal lattice of a simple cubic lattice is another simple cubic lattice. Hence deduce the first Brillouin zone of the simple cubic crystal lattice.
37. Show that the reciprocal lattice for a body centred cubic crystal is face centred cubic. (Raj., 1974, Jodhpur, 1968)
38. What do you mean by reciprocal lattice? Calculate the reciprocal lattice for a direct *fcc* lattice. What is meant by the first Brillouin Zone? (Delhi 1983)
39. Discuss the origin of the geometric structure factor and the atomic scattering factor in x-ray diffraction and derive suitable mathematical expressions for them. What is the importance of the geometrical structure factor in the analysis of the crystal structure? (Raj., 1977)
40. Explain the origin of van der Waals forces in molecular crystal. Show that for van der Waals forces the interaction energy varies as  $1/R^6$ , where  $R$  is the separation of two interacting atoms. (Raj., 1984)
41. Define Cohesive energy and determine its value for crystals of inert gases.
42. Obtain the binding energy of an ionic crystal and derive the expression for the Madelung constant. Calculate the value of this constant for a linear ionic lattice. (Raj., 1970)
43. Obtain the expression for the total cohesive energy of an ionic crystal in terms of Madelung constant and other parameters.
44. What is Madelung constant? Show that Madelung constant for one dimensional array of ions of alternating sign with a distance  $a$  between successive ions is equal to  $2 \log_e 2$ .
45. What is Madelung constant? Calculate it for NaCl crystal.
46. (i) What are the characteristics of covalent bond? Discuss the directional nature of a covalent bond.  
 (ii) What do you mean by hybridization? Discuss the hybridization of carbon.
47. Discuss the various types of binding in crystals. How do they differ from each other? Explain giving at least one example of each. (Rohilkhand 1990)

48. Describe the nature and origin of various forces existing between the atoms of a crystal. Explain the formation of a stable bond using the potential energy versus interatomic distance curve.
49. What do you mean by crystal imperfections? Mention the different types of imperfection present in a crystal.
50. (i) Explain the Schottky and Frenkel defects. What is the difference between a Schottky defect and a Frenkel defect?  
(ii) Find an expression for the number of vacancies that can exist in a crystal at finite temperature.
51. What are Frenkel defects? *(Madras 2006)*
52. There are  $N$  atoms in a metal crystal and  $N_i$  interstitials in its structure. If  $\bar{E}_f$  and  $n$  are the average energy required to create one Frenkel defect and equilibrium number of defects, respectively, then show that

$$n = (NN_i)^{1/2} \exp \left\{ \frac{\vec{E}_f}{2kT} \right\}$$

where  $k$  is Boltzmann constant.

53. What are point and line defects?
54. Explain edge and screw dislocations.
55. Discuss the various types of dislocations in a crystal and indicate the manner in which the Burgers vector  $\vec{b}$  differs for different dislocations.
56. Explain the terms: grain boundary, Tilt boundary, Stacking fault.

## LATTICE VIBRATIONS AND PHONONS

**After reading this chapter, you should be able to**

- ◆ Discuss the features of the vibrations of monatomic and of diatomic linear chains and explain the significance of dispersion curves
- ◆ List the characteristics of phonons
- ◆ Discuss Einstein's and Debye's theory of heat capacity
- ◆ Explain why Einstein's theory is not capable of giving correct variation of specific heat of solids at low temperatures.

## PHONONS I. CRYSTAL VIBRATIONS

### 26.1 INTRODUCTION

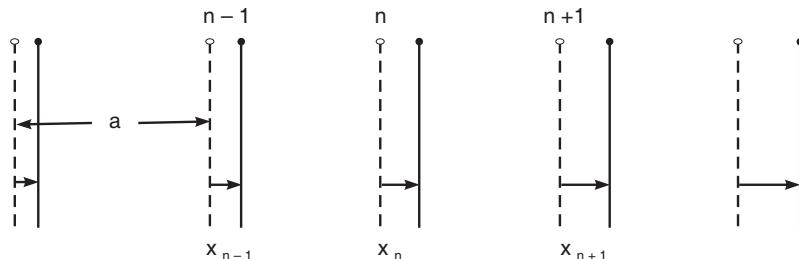
Our discussion so far in the previous chapters have been concerned with systems involving electrons. We now switch emphasis in this chapter to some properties of solids not dependent on the presence of mobile electrons but dependent rather on the motions of the much more massive ion cores. While these atomic structures do not generally diffuse throughout the crystal, they do vibrate about the lattice sites and create collective wave motions that are called *phonons*.

A lattice is usually regarded as an array of atoms connected with each other by elastic springs. With this model for the lattice in mind, it is quite easy to understand that the motion of the adjacent atoms (excited, say, by dynamical external forces) would be coupled, or in other words, we can say that the motion of every atom would be shared by all the atoms in the crystal and the crystal would thus vibrate as a whole. Our problem is to investigate the characteristics of this elastic vibrational motion of the crystal lattices.

### 26.2 VIBRATIONS OF MONATOMIC LINEAR LATTICE

A monatomic linear crystal lattice is an array of identical, equidistant atoms (or ions), each of mass  $m$ , with interatomic separation  $a$ . The atoms are presumed to be connected by springs and the displacements of atoms are within the Hooke's law limits. Further, only nearest neighbour interactions are considered. When an elastic wave propagates through the crystal, the atoms are set in vibrations either parallel or perpendicular to the direction of the wave vector. We shall consider longitudinal elastic vibrations of the atoms.

Let us consider an infinitely long linear chain of identical atoms (Fig. 26.1). The open circles represent the equilibrium positions and the black dots the displaced atoms.



**Fig. 26.1**

Let

$x_n$  = displacement of  $n$ th atom from equilibrium position

$x_{n-1}$  = displacement of  $(n-1)$ th atom from equilibrium position

$x_{n+1}$  = displacement of  $(n+1)$ th atom from equilibrium position

$F_n$  = force on  $n$ th atom, and

$f$  = force constant for the nearest-neighbour interaction.

According to Hooke's law, the force on an atom due to the displacement of its (nearest) neighbour is proportional to the difference of their displacements. So the net force on the  $n$ th atom is

$$\begin{aligned} F_n &= f(x_{n+1} - x_n) - f(x_n - x_{n-1}) \\ &= f(x_{n+1} + x_{n-1} - 2x_n) \end{aligned}$$

The equation of motion of the  $n$ th atom is

$$m \ddot{x}_n = f(x_{n+1} + x_{n-1} - 2x_n). \quad \dots(1)$$

Let a solution to this equation be of the form of a travelling wave :

$$x_n = e^{-i\omega(t-na/v)} \quad \dots(2)$$

Here,

$na$  = the equilibrium position (x-co-ordinate) of the  $n$ th particle relative to the origin, and

$v$  = the velocity of wave propagation and  $v = v\lambda$

$\omega$  = the angular frequency =  $2\pi v$ .

$$\text{Now, the wave vector } K = \frac{2\pi}{\lambda} = \frac{\omega}{v}.$$

Therefore, Eq. (2) can be written as

$$x_n = e^{-i(\omega t - Kna)} \quad \dots(3)$$

$$\therefore x_{n+1} = e^{-i\{\omega t - K(n+1)a\}} = e^{iKa} x_n$$

$$\text{and } x_{n-1} = e^{-i\{\omega t - K(n-1)a\}} = e^{-iKa} x_n.$$

Differentiating Eq. (3) twice with respect to  $t$ , we get

$$\ddot{x}_n = \frac{d^2 x_n}{dt^2} = (-i\omega)^2 e^{-i(\omega t - Kna)} = -\omega^2 x_n.$$

Substituting for  $\ddot{x}_n, x_{n+1}, x_{n-1}$  and  $x_n$  in Eq. (1), and dividing by  $x_n$ , we get

$$\begin{aligned} -m\omega^2 &= f(e^{iKa} + e^{-iKa} - 2) \\ &= f(e^{iKa/2} - e^{-iKa/2})^2. \end{aligned}$$

$$\text{Now, } \sin \frac{Ka}{2} = \frac{e^{iKa/2} - e^{-iKa/2}}{2i}$$

$$\text{or } \sin^2 \frac{Ka}{2} = -\frac{1}{4} (e^{iKa/2} - e^{-iKa/2})^2.$$

$$\therefore -m\omega^2 = -4f \sin^2 \frac{Ka}{2}$$

$$\text{or } \omega^2 = \frac{4f}{m} \sin^2 \frac{Ka}{2}$$

$$\text{or } \omega = 2 \sqrt{\frac{f}{m}} \left| \sin \frac{Ka}{2} \right|. \quad \dots(4)$$

The absolute value signs occur because a frequency is always a positive quantity by definition. Fig. 26.2 gives the plot of  $\omega$  versus  $K$ . It shows that frequency is a periodic function of  $K$ , with period  $2\pi/a$ . The dashed lines correspond to a continuous string for which  $\omega \propto K$ .

Eq. (4) is called the dispersion relation.

**Results**— We draw the following results from the above discussion :

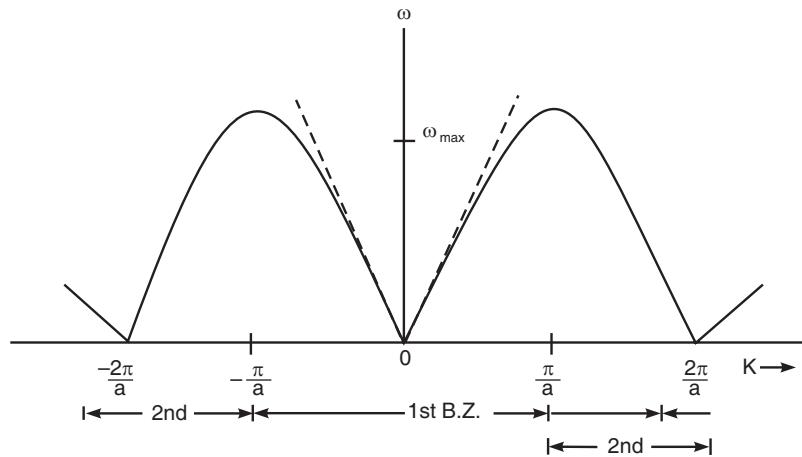


Fig. 26.2

- At low frequencies ( $K \rightarrow 0$ , the long wavelength limit)

$$\sin \frac{Ka}{2} \rightarrow \frac{Ka}{2}$$

$$\therefore \omega = Ka \sqrt{\frac{f}{m}} \quad \dots(5)$$

In this long wavelength approximation the phase velocity will be essentially constant, since

$$v_p = \frac{\omega}{K} = a \sqrt{\frac{f}{m}} = v_0 \quad \dots(6)$$

Group velocity

$$v_g = \frac{d\omega}{dK} = a \sqrt{\frac{f}{m}} = v_0 \quad \dots(7)$$

$$\therefore v_g = v_p = \text{constant}$$

For very long wavelengths then, the dispersion effects are negligible and the medium acts like a continuous and homogeneous elastic medium. That is what happens in a continuous string. Thus, *for long wavelengths*, the periodicity of the lattice is of little importance.

For shorter wavelengths the approximation of Eq. (5) is no longer valid and waves of different frequencies travel at different velocities. The dispersion arises because of the discrete character of this chain of particles.

- Another important result is obtained by comparing the solution Eq. (3) with another in which  $K$  is replaced by  $K_m$  where

$$K_m = K + \frac{2\pi m}{a} \text{ with } m = \pm 1, \pm 2, \dots$$

We have for the displacement of the  $n$ th atom

$$\begin{aligned} x_n &= e^{-i(\omega t - K_m n a)} \\ &= e^{-i\{\omega t - (K + 2\pi m/a)n a\}} \\ &= e^{-i(\omega t - K n a)} \quad [\because e^{2\pi i m} = 1] \end{aligned}$$

From Eq. (4), we have

$$\begin{aligned}\omega &= \sqrt{\frac{4f}{m}} \left| \sin \frac{1}{2} K_m a \right| \\ &= \sqrt{\frac{4f}{m}} \left| \sin \frac{1}{2} a \left( K + \frac{2\pi m}{a} \right) \right| \\ &= \sqrt{\frac{4f}{m}} \left| \sin \frac{1}{2} K a \right|.\end{aligned}$$

Thus the solutions Eq. (3) and the frequencies Eq. (4) corresponding to  $K$  and  $K_m$  are identical. It means that the state of vibration of the lattice corresponding to a wave vector  $K$  is the same as that for any of the wave vectors  $K + 2\pi m/a$ . Therefore, to find a unique relation between the state of vibration and the wave vector, the wave vector  $K$  must be restricted to a range of values  $2\pi/a$ . We specify the range of independent values of  $K$  by

$$-\frac{\pi}{a} \leq K \leq \frac{\pi}{a}.$$

We have taken both positive and negative values of  $K$  because waves can propagate to the right or to the left. This range of  $K$  values is referred to as the “first Brillouin zone” of the linear lattice. The extreme values of  $K$  in this zone are

$$K_{\max} = \pm \frac{\pi}{a}$$

where  $K_{\max}$  may be the order of  $10^{10} \text{ m}^{-1}$ .

At the boundaries  $K_{\max} = \pm \frac{\pi}{a}$  of the Brillouin zone the solution  $x_n = e^{-i(\omega t - Kna)}$  becomes

$$x_n = e^{-i\omega t} e^{\pm i\pi} = e^{-i\omega t} (-1)^x \quad \dots(8)$$

Eq. (8) represents a “standing” wave in which alternate atoms vibrate in opposite phases. This situation is equivalent to Bragg reflection of X-rays. When the Bragg condition is satisfied, a wave cannot propagate in a lattice, but through successive reflections back and forth, a standing wave is set up. Here the critical value  $K_{\max} = \pm \frac{\pi}{a}$  satisfies the Bragg condition  $2d \sin \theta = n\lambda$ . Here we have  $\theta = \frac{\pi}{2}$  (linear lattice),  $d = a$ ,  $n = 1$ , so that

$$\lambda = \lambda_{\min} = 2a$$

which corresponds to  $K_{\max} = \pm \pi/a$ .

3. The frequency corresponding to zone boundaries ( $K = K_{\max} = \pm \pi/a$ ) is according to Eq. (4),

$$\omega_{\max} = 2\sqrt{\frac{f}{m}}$$

$$\text{or} \quad v_{\max} = \frac{1}{\pi} \sqrt{\frac{f}{m}}. \quad [\because \omega = 2\pi v]$$

Thus there exists a maximum frequency  $v_{\max}$  which can be propagated through the chain. Hence a monatomic linear lattice acts as a low-pass filter which transmits only in the frequency range between zero and  $v_{\max}$ .

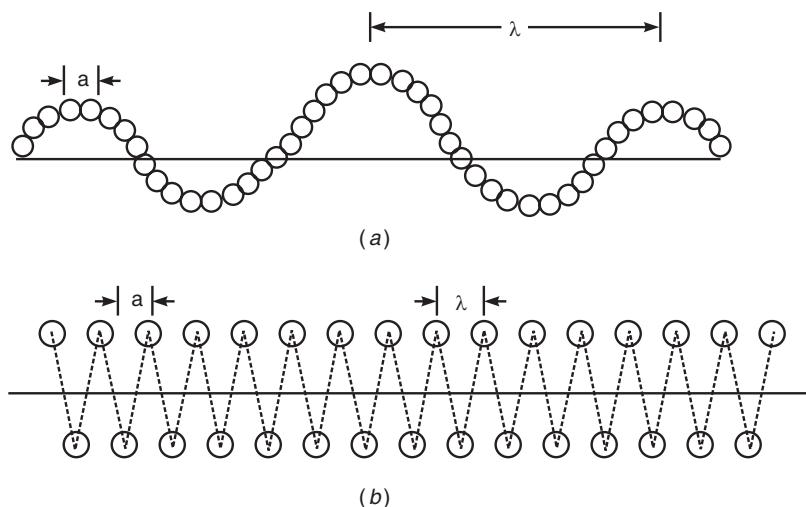
$$v_{max} = \frac{v}{\lambda_{min}} = \frac{v}{2a}.$$

Now,  $a = 10^{-10} \text{ m}$ . The velocity of sound in solids is  $\sim 10^3 - 10^4 \text{ ms}^{-1}$ .

$$\therefore v_{max} \approx 10^{13} \text{ Hz.}$$

Hence the cut off frequency lies in the infrared region.

The actual physical character of motion is illustrated by Fig. 26.3 (a) for the long wavelength case. Fig. 26.3 (b) illustrates the condition  $\lambda = 2a$  in which the stationary waves are set up.

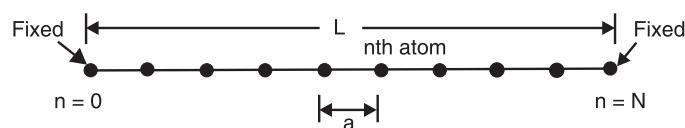


**Fig. 26.3**

The displacements are shown as transverse ones for the sake of clarity in representation, but it should be remembered that the actual displacements which are discussed in the text are longitudinal.

### 26.3 VIBRATIONAL MODES OF A FINITE ONE-DIMENSIONAL LATTICE OF IDENTICAL ATOMS

Let there be  $N + 1$  atoms in the array numbered from 0 to  $N$ . We further suppose that the two end atoms are rigidly fixed so that the number of mobile atoms is  $(N - 1)$  (Fig. 26.4). If the length of the chain be  $L$ , and each atom separated from its neighbour by a distance  $a$ , then  $a = L/N$ .



**Fig. 26.4**

Let  $a$  be the equilibrium interatomic spacing, and  $x_0, x_1, x_2, \dots, x_N$  the instantaneous displacements from the equilibrium positions. Let  $m$  be the mass of an atom and  $f$  the force constant for the nearest-neighbour interaction.

Assuming Hooke's law to be obeyed and considering only nearest-neighbour interaction, the equation of motion of the  $n$ th atom is

$$\begin{aligned} m \ddot{x}_n &= f(x_{n+1} - x_n) - f(x_n - x_{n-1}) \\ &= f(x_{n+1} + x_{n-1} - 2x_n) \end{aligned} \quad \dots(1)$$

The general solution of Eq. (1) can be written as a sum of two travelling waves, one propagating to the right and the other to the left :

$$x_n = [A_1 e^{i(Kna + \beta_1)} + A_2 e^{i(-Kna + \beta_2)}] e^{-i\omega t} \quad \dots(2)$$

where  $A_1$  and  $A_2$  are amplitudes,  $\beta_1$  and  $\beta_2$  are phase angles and  $K$  is wave vector ( $K = 2\pi/\lambda$ ). The boundary conditions are

$$\left. \begin{array}{l} x_0 = 0 \\ x_N = 0 \end{array} \right\} \text{for all } t$$

The first boundary condition, when substituted in Eq. (2) yeilds

$$\left. \begin{array}{l} A_1 = -A_2 \\ \beta_1 = \beta_2 = 0 \end{array} \right\} \text{and} \quad \dots(3)$$

(since phase angles are equal we choose them to be equal to zero)

Eq. (2) becomes

$$x_n = A_1 (e^{iKna} - e^{-iKna}) e^{-i\omega t} \quad \dots(4)$$

Taking real part of Eq. (4), we have

$$x_n = 2A_1 \sin Kna \sin \omega t. \quad \dots(5)$$

This represents a “standing” wave and leads to the dispersion relation

$$\omega^2 = \frac{4f}{m} \sin^2 \frac{Ka}{2} \quad \dots(6)$$

as that for travelling wave solution [Eq. (4) of Art. 26.2], with the difference that  $K$  is now limited to positive values ranging from 0 to  $\pi/a$ .

The second boundary condition imposed on Eq. (5) gives

$$0 = 2A_1 \sin KNa \sin \omega t$$

or

$$\sin KNa = 0$$

or

$$K = \frac{\pi}{Na} j, \quad \dots(7)$$

where  $j$  is an integer. However,  $j = 0$  must be excluded because it corresponds to  $K = 0$  which means that all atoms are at rest. The maximum value of  $K$  is  $\frac{\pi}{a}$  which gives  $j_{max} = N$ . This value

must also be excluded because it yields in Eq. (5)  $x_n = 0$  for all  $t$ , that is again all atoms are at rest. We, therefore, conclude that

$$j = 1, 2, 3, \dots, (N-1) \quad \dots(8)$$

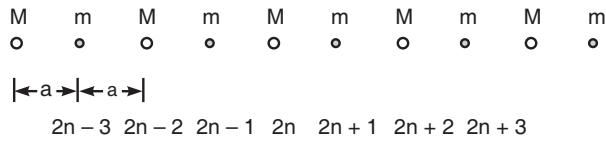
Thus the boundary conditions select a discrete set of  $K$  values given by Eqs. (7) and (8). We see that, *there are just as many modes of vibration (K-values) as there are mobile atoms*.

To each value of  $K$  there corresponds a frequency  $\omega_k$ . Hence the frequency spectrum consists of  $(N-1)$  discrete lines.

## 26.4 VIBRATIONAL MODES OF A DIATOMIC LINEAR LATTICE

For crystals with more than one atom per primitive cell, the vibrational spectrum shows new features. For each polarization mode in a given direction of propagation, the dispersion relation  $\omega - K$  gives rise to two branches, known as ‘acoustical branch’ and ‘optical branch’.

Consider a linear diatomic lattice, with atoms of two kinds arranged alternately (Fig. 26.5). The particles are numbered in such a way that the even numbered have mass  $M$  and odd ones  $m$  ( $M > m$ ).



**Fig. 26.5**

Let  $a$  be the distance between nearest neighbours, so that  $2a$  is the repeat distance.

We assume that each atom interacts only with its nearest neighbour and that the force constants are identical between all nearest-neighbour pairs. The equations of motion are

$$\begin{aligned} M \ddot{x}_{2n} &= f(x_{2n+1} + x_{2n-1} - 2x_{2n}) \\ \text{and } m \ddot{x}_{2n+1} &= f(x_{2n+2} + x_{2n} - 2x_{2n+1}) \end{aligned} \quad \dots(1)$$

where  $f$  is the force constant and  $x_{2n}$  is the displacement of the  $2n$ th particle from the equilibrium position.

Let the solutions to these equations be of the form of a travelling wave :

$$\begin{aligned} x_{2n} &= A e^{-i(\omega t - 2nKa)} \\ \text{and } x_{2n+1} &= B e^{-i[\omega t - (2n+1)Ka]} \end{aligned} \quad \dots(2)$$

where  $K$  is the wave vector of a particular mode of vibration, and  $A$  and  $B$  are the amplitudes corresponding to the atoms  $M$  and  $m$  respectively.

Substituting Eq. (2) in Eq. (1), we get

$$\begin{aligned} (M\omega^2 - 2f)A + 2Bf \cos Ka &= 0 \\ \text{and } (m\omega^2 - 2f)B + 2Af \cos Ka &= 0 \end{aligned} \quad \dots(3)$$

These equations have non-vanishing solutions for  $A$  and  $B$  only if the determinant of the coefficients of  $A$  and  $B$  vanishes, i.e.,

$$\begin{vmatrix} (M\omega^2 - 2f) & 2f \cos Ka \\ 2f \cos Ka & (m\omega^2 - 2f) \end{vmatrix} = 0$$

$$\text{or } (M\omega^2 - 2f)(m\omega^2 - 2f) - 4f^2 \cos^2 Ka = 0$$

$$\text{or } Mm\omega^4 - 2f(M+m)\omega^2 + 4f^2(1 - \cos^2 Ka) = 0$$

$$\text{or } Mm\omega^4 - 2f(M+m)\omega^2 + 4f^2 \sin^2 Ka = 0$$

This gives the dispersion relation

$$\omega^2 = f \left( \frac{1}{M} + \frac{1}{m} \right) \pm f \sqrt{\left( \frac{1}{M} + \frac{1}{m} \right)^2 - \frac{4 \sin^2 Ka}{Mm}}$$

This is the dispersion relation for the linear diatomic lattice. The form of the relation shows that, as in the monatomic case, the frequency  $\omega$  is a periodic function of the wave vector  $K$ .

Since  $\omega$  should be positive, each value of  $\omega^2$  leads to a single value for  $\omega$ . Thus, in contrast to the monatomic lattice, there are now two angular frequencies  $\omega_+$  and  $\omega_-$  corresponding to a single value of the wave vector  $K$ . In a  $\omega - K$  plot (Fig. 26.6) this leads to two "branches", the one corresponding to  $\omega_+$  is called the 'optical' branch and that corresponding to  $\omega_-$  is called the 'acoustical' branch.

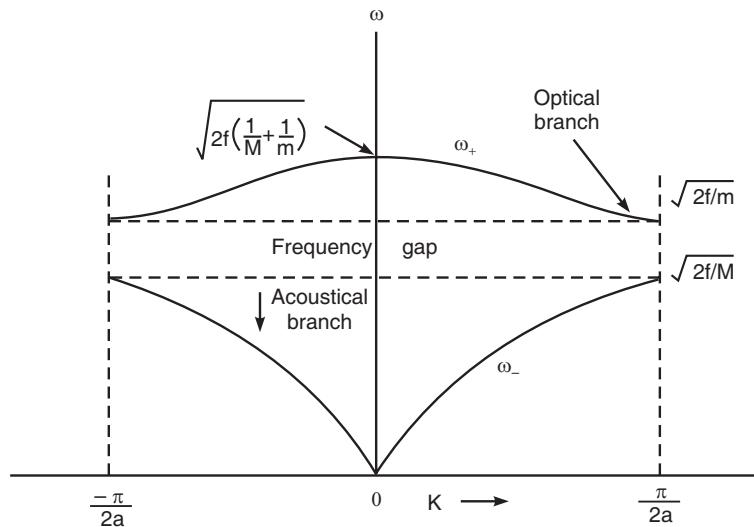


Fig. 26.6

It is further observed that as in the mono-atomic case, the frequency is a periodic function of the wave vector  $K$ . The first zone thus limits the values of  $K$  to the range between  $-\pi/2a$  and  $+\pi/2a$ . For  $K = \pm\pi/2a$ , the frequencies are

$$\omega_+ = \sqrt{\frac{2f}{m}}, \quad \omega_- = \sqrt{\frac{2f}{M}} \quad \dots(5)$$

It is of interest to investigate the physical difference between the acoustical and optical branches and why they are named so. To find out, we calculate  $\frac{A}{B}$  as  $K \rightarrow 0$  for both the branches.

As  $K \rightarrow 0$ ,  $\cos Ka \rightarrow 1$  and  $\sin Ka \rightarrow 0$

$$\omega_- = 0 \text{ and } \omega_+ = \sqrt{2f\left(\frac{1}{M} + \frac{1}{m}\right)} \quad \dots(6)$$

$$\left. \begin{array}{l} \text{For acoustical branch} \quad A = B \\ \text{and for optical branch} \quad -MA = mB \end{array} \right\} \quad \dots(7)$$

For the optical branch at  $K = 0$ , the vibrations of the atoms are in opposite directions and amplitudes are inversely in the ratio of the masses, so that the centre of mass of the unit cell remains fixed during the period of motion. For acoustical mode of vibration, the two types of atoms move in the same direction with same amplitude as shown in Fig. 26.7. Both waves are travelling waves. The optical branch can have  $K = 0$  mode but the acoustical cannot.

These characteristics are typical of the optical and acoustical branch in general. The acoustical branch vibrations can be excited by some kind of force that forces all the atoms in the crystal to go in the same direction. For example, we may direct a beam of sound waves at the surface of a crystal and this will produce the desired effect. It is for this reason, such vibrations are called the acoustical vibrations and constitute the acoustical branch in dispersion relation.

The optical mode of vibrations in ionic crystals where the two types of atoms are oppositely charged, can be excited by an electric field, which tends to move the ions in opposite directions. Specially in ionic crystals, this mode can be excited by the electric field associated with light wave, from which the term optical vibrations is derived.

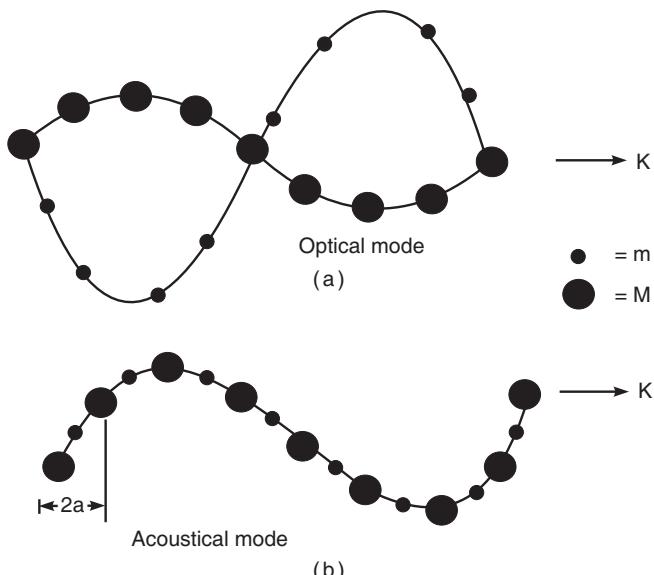


Fig. 26.7

## 26.5 QUANTISATION OF LATTICE VIBRATIONS

The energy in a lattice vibration or elastic wave is quantized. The quantum of energy in an elastic wave is called *phonon*. In fact, phonon represents the particle concept associated with the *quantization of lattice waves*. The situation is quite analogous to the electromagnetic waves where the quanta are called *photons*. It is quite possible to treat the interaction between two phonons or between a phonon and an electron as scattering ‘collision’ between two particles. Sound waves in crystals are composed of phonons. Thermal vibrations in crystals are thermally excited phonons.

It is convenient to call  $\hbar \omega_k$  ( $\omega_k$  is the angular frequency of the  $k$ th mode of vibration) the energy of a ‘phonon’. The vibrational mode at the temperature  $T$  corresponds to

$$n = \frac{1}{e^{\hbar \omega_k / kT} - 1} \quad \dots(1)$$

phonons. At any temperature  $T$  the crystal may be regarded as filled with a gas of phonons. More phonons are created with increase in temperature and their number is decreased by lowering temperature. As the phonons are associated with elastic waves in solid, they travel with the velocity of sound ( $C_s$ ) in the solid medium. The energy of a phonon is

$$\hbar \omega_k = \hbar C_s k \quad \dots(2)$$

The vibration spectrum of the phonon waves occupy a wide range of frequency from  $10^4$  to  $10^{13}$  Hz. The low frequency part of the spectrum is in the acoustic range and the higher frequency part is in infrared range.

The quantisation of lattice vibration energy has an important consequence when lattice vibration is involved in an interaction. The energy of a mode of vibration has to be  $(n + 1/2) \hbar \omega$  with  $n$  any positive integer. Further, to a first order approximation, transition occurs only between adjacent allowed states of the system. Therefore, the energy change  $\Delta \varepsilon$  of the oscillator corresponding to the energy of the quantum which is absorbed or emitted must be

$$\Delta \varepsilon = \pm \hbar \omega$$

i.e., in any such transition the quantum number of the oscillator may change only by

$$\Delta n = \pm 1.$$

The (+) sign refers to absorption and the (-) sign to emission.

There is no direct experimental evidence that the energy of an elastic wave is quantised. However, the following facts provide evidence for the existence of phonons :

- (i) The lattice contribution to the heat capacity of solids always approaches zero as the temperature approaches zero. This can be explained only if the lattice vibrations are quantised.
- (ii) X-rays and neutrons are scattered inelastically by crystals. The change of energy and momentum in this process corresponds to the creation or absorption of one or more phonons.

## 26.6 INELASTIC SCATTERING OF NEUTRONS BY PHONONS

A phonon of wave vector  $K$  interacts with other particles such as photons, neutrons and electrons as if it has a momentum  $\hbar K$ . Sometimes  $\hbar K$  is called the ‘crystal momentum’.

Neutron inelastic scattering is the most common method for the experimental determination of phonon dispersion relations. A neutron sees the crystal lattice chiefly by interaction with the nuclei of the atoms. The kinematics of the scattering of a neutron beam by a crystal lattice are described by the general wave vector selection rule :

$$\mathbf{k} + \mathbf{G} = \mathbf{k}' \pm \mathbf{K} \quad \dots(1)$$

and by the requirement of the conservation of energy. Here  $\mathbf{K}$  is the wave vector of the phonon created (+) or absorbed (-) in the process, and  $\mathbf{G}$  is any reciprocal lattice vector. We choose  $\mathbf{G}$  such that  $\mathbf{K}$  lies in the first Brillouin zone, as it must for a phonon.

The kinetic energy of the incident neutron is

$$K.E. = \frac{1}{2} M_n v^2 = \frac{1}{2} \frac{p^2}{M_n} = \frac{\hbar^2 k^2}{2M_n}$$

where  $M_n$  is the mass of the neutron. The momentum  $p$  is given by  $\hbar \mathbf{k}$  where  $\mathbf{k}$  is the wave vector of the neutron. If  $\mathbf{k}'$  is the wave vector of the scattered neutron, the energy of the scattered neutron is  $\frac{\hbar^2 k'^2}{2M_n}$ . Thus from the conservation of energy law, we have

$$\frac{\hbar^2 k^2}{2M_n} = \frac{\hbar^2 k'^2}{2M_n} \pm \hbar\omega \quad \dots(2)$$

where  $\hbar\omega$  is the energy of the phonon created (+) or absorbed (-) in the process.

To determine the dispersion relation using Eqs. (1) and (2), it is necessary in the experiment to find the energy gain or loss of the scattered neutron as a function of the scattering direction  $\mathbf{k} - \mathbf{k}'$ . We determine  $\mathbf{G}$  the reciprocal lattice vector, from the elastically scattered (Bragg reflected) neutrons. The elastic scattering of a neutron by crystals obeys the following wave vector conservation law :

$$\mathbf{k}' = \mathbf{k} + \mathbf{G} \quad \dots(3)$$

where  $\mathbf{k}'$  is wave vector of scattered neutron, and  $\mathbf{k}$  wave vector of incident neutron.

The reciprocal lattice vector  $\mathbf{G}$  can be measured through Eq. (3) by elastic scattering for a given  $\mathbf{k}$  and  $\mathbf{k}'$ . This, in turn, gives  $\mathbf{K}$  in Eq. (1). On the other hand,  $\omega$  is obtained from Eq. (2). Thus a pair  $(\mathbf{K}, \omega)$  could be obtained. This would locate a point on the dispersion curve. By changing  $\mathbf{k}$  (and hence  $\mathbf{k}'$ ), other points on the dispersion curve could be located to get the entire graph.

As an example of a widely used experimental method for the measurement of phonon dispersion relation, we briefly describe the triple axes (neutron) spectrometer (TAS), shown schematically in Fig.

26.8. The wave vector ( $\mathbf{k}$ ) of the incoming neutron beam is known from the reactor characteristics. The TAS method is designed to measure the magnitude and the direction of the scattered beam ( $\mathbf{k}'$ ) through a facility for changing the angle ( $\phi$ ) between  $\mathbf{k}$  and  $\mathbf{k}'$  and the angle ( $\psi$ ) between  $\mathbf{k}$  and some reference direction of the crystal.

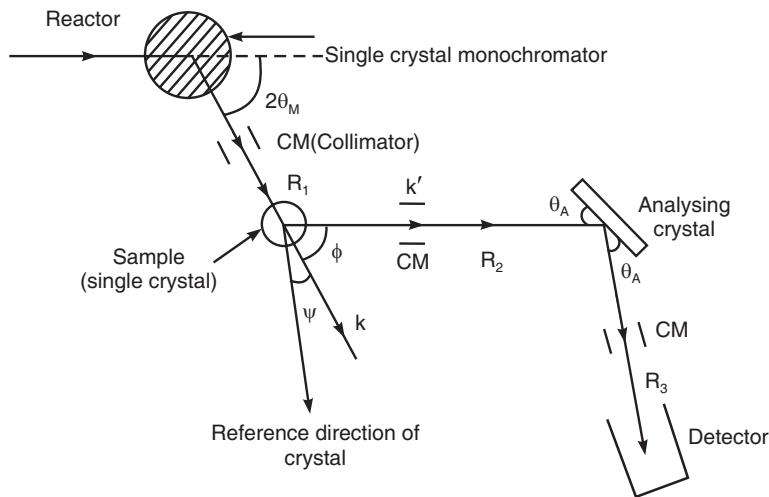


Fig. 26.8

By changing the angle  $\theta_M$ , it is possible to change  $\mathbf{k}$ . By a proper setting of the table of the crystal (angle  $\psi$ ), any desired angle between  $\mathbf{k}$  and a certain crystal plane (characterised by a set of Miller indices) corresponding to some reciprocal lattice vector may be obtained. By rotating the second arm, the angle  $\phi$  between  $\mathbf{k}$  and  $\mathbf{k}'$  can be changed. By rotating the third arm (with the crystal analyser synchronised to follow with half the angular velocity) the scattered beam in the direction  $\mathbf{k}'$  can be analysed.

The neutrons of desired energy are scattered towards the sample crystal by Bragg reflecting them from a suitably oriented single crystal (the monochromator in Fig. 26.8). The single crystal analyser is used to define the energy of those neutrons scattered from the sample and which reach the detector. The whole arrangement is called the triple axes spectrometer, since rotation about three axes is required to adjust the three scattering angles.

The dispersion curves of sodium for phonons propagating in [100], [110] and [111] directions at 90 K, as determined by inelastic scattering of neutrons are given in Fig. 26.9.

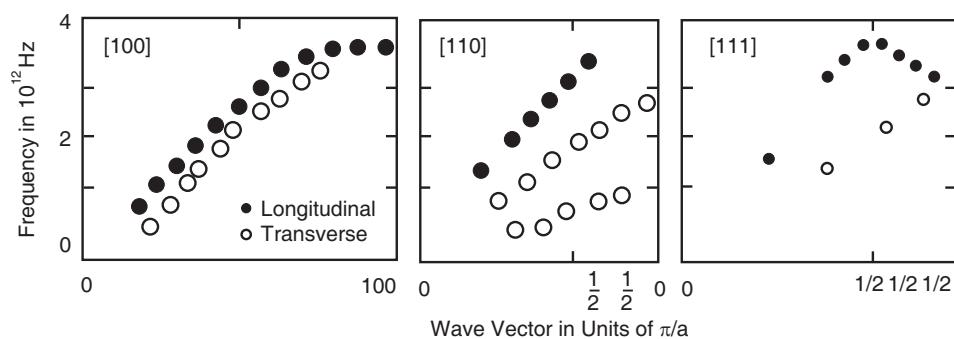


Fig. 26.9

### PHONONS II. THERMAL PROPERTIES (PHONON HEAT CAPACITY)

- By heat capacity we shall usually mean the heat capacity at constant volume.
- The heat capacity at constant volume is defined as  $C_v = (\partial U / \partial T)_v$ , where  $U$  is the energy and  $T$  the temperature.
- **Specific heat capacity.** *The specific heat capacity of a substance is defined as the quantity of heat required to raise the temperature of 1 kg of the substance through 1K.*  
The unit of specific heat capacity ( $C$ ) is  $\text{J kg}^{-1} \text{ K}^{-1}$ .
- **Atomic heat capacity.** Atomic heat capacity of a substance is the product of the specific heat capacity and the atomic weight.

### 26.7 DULONG AND PETIT'S LAW

Dulong and Petit studied the specific heat capacities of various substances in solid state and they formulated a law.

**Statement.** *The product of atomic weight and the specific heat capacity (i.e., atomic heat capacity) for all elements in solid state is a constant. The constant is nearly  $25 \text{ J mol}^{-1} \text{ K}^{-1}$ .*

**Explanation.** According to this law, the atomic heat capacity of a substance is constant at all temperatures.

- The law agrees with experiment at ordinary temperatures. But experiments show that the specific heat capacity of solids decreases rapidly with decrease in temperature at very low temperatures.

#### Derivation of Dulong and Petit's law

All solids are made up of atoms or molecules closely packed together. The atoms are arranged in a crystalline array by strong electro-magnetic forces exerted by every atom on each other. They are not free to move as the gas molecules. At absolute zero, the atoms are at rest and the energy of solid is assumed to be zero. When the temperature is raised, the atoms start vibrating about their mean positions and the vibrations are transmitted from one atom to another atom. At low temperatures, the displacement of atoms about their mean position is very small as compared to interatomic distances. Hence these are regarded as harmonic oscillators. Thus the vibrations of atoms are simple harmonic in nature.

We know that in case of S.H.M., the average K.E. is equal to average potential energy. So the average energy corresponding to one vibration is equal to twice the kinetic energy corresponding to one degree of freedom. From the law of equipartition of energy, the average kinetic energy per atom

in each degree of freedom is  $\frac{1}{2} kT$ . Thus, the total energy per atom in each degree of freedom is

$$\frac{1}{2} kT + \frac{1}{2} kT = kT.$$

Each atom has three degrees of freedom.

Total average energy of each atom =  $3 kT$ .

Total thermal energy of one mole of solid =  $3NkT = 3RT$

Here,  $N$  is Avogadro's number and  $Nk = R$  (gas constant)

∴

$$U = 3RT$$

The atomic heat capacity at constant volume is given by

$$C_v = \frac{dU}{dT} = 3R.$$

This is called Dulong and Petit's law.

### 26.7.1. Variation of Specific Heat Capacity and Atomic Heat Capacity with Temperature

It was experimentally found by Nernst and Lindemann that the specific heat capacity of a substance decreases with decrease in temperature and tends to zero at absolute zero. The specific heat capacity increases with increase in temperature and reaches a *maximum* value at higher temperatures. The variation of atomic heat capacity with temperature for diamond, aluminium and silver is shown in Fig. 26.10.

The variation of specific heat capacity with temperature could not be explained on the basis of the classical theories of heat. The first attempt to explain the variation of specific heat capacity with temperature was made by Einstein in 1907 on the basis of quantum theory.

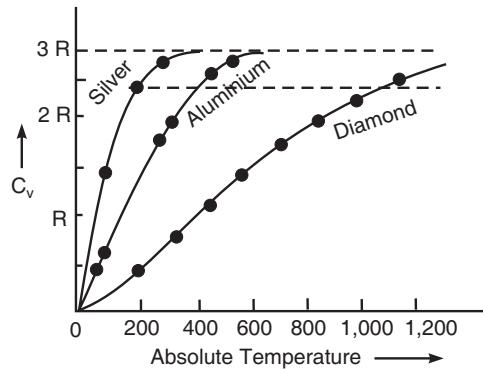


Fig. 26.10

## 26.8 EINSTEIN'S THEORY OF SPECIFIC HEAT CAPACITY

According to quantum theory, heat is radiated from a body in the form of discrete particles, called *photons*. The energy of a photon is equal to  $h\nu$  where  $h$  is the Planck's constant and  $\nu$  is the frequency of radiation.

**Assumptions of Einstein model.** Einstein assumed:

- (1) Every solid is composed of atoms. The atoms are at rest at 0K. The thermal energy is zero at 0K. When the solid is heated, the atoms vibrate simple harmonically about their mean equilibrium positions. All the atoms vibrate with the same frequency.
- (2) Each atom has three degrees of freedom.
- (3) The mean energy per degree of freedom is  $\frac{h\nu}{e^{\frac{h\nu}{kT}} - 1}$  as calculated by Planck by the application of quantum theory. [For Derivation, see Section 6.7, chapter 6]

**Derivation of Einstein's equation for  $C_v$ ,**

$$\text{The energy of each atom} = \frac{3h\nu}{e^{\frac{h\nu}{kT}} - 1}$$

∴ The internal energy of a solid consisting of  $N$  atoms

$$U = \frac{3N h\nu}{e^{\frac{h\nu}{kT}} - 1}$$

∴ Atomic heat capacity at constant volume

$$C_v = \frac{dU}{dT} = 3Nk \left( \frac{h\nu}{kT} \right)^2 \frac{e^{h\nu/kT}}{\left[ e^{\frac{h\nu}{kT}} - 1 \right]^2}$$

$$C_v = 3R \left\{ \left( \frac{\theta}{T} \right)^2 \frac{e^{\theta/T}}{(e^{\theta/T} - 1)^2} \right\} \text{ where } \theta = \frac{h\nu}{k} \quad \dots(1)$$

This is Einstein's equation for the atomic heat of a solid at constant volume.

**Merits.** (1) In the above equation, as  $T \rightarrow 0$ ,  $C_v \rightarrow 0$ . Also according to Eq. (1), atomic heat capacity decreases with decrease of temperature.

(2) At high temperatures, the atomic heat capacity approaches  $3R$ , which is in agreement with experiments.

**Demerit.** It was observed that at low temperature, there is some disagreement between experiment and theory. It was observed that in the case of some elements like copper, aluminium, iron etc., the atomic heat capacity at low temperature decreases more rapidly than that predicted by Einstein's theory. The disagreement is due to the assumption that all atoms of a solid vibrate at the same frequency. Debye considered the atoms, not as isolated oscillators, all vibrating at the same frequency, but as a system of coupled oscillators, having a continuous spectrum of natural frequencies. On this assumption, Debye modified Einstein's theory.

#### Einstein's Theory of Specific Heat (Einstein Model) [II Method]

The salient features of Einstein's model are:

(1) Einstein assumed that the atoms are identical independent harmonic oscillators with a single natural frequency but their energy spectrum is not continuous. The energies are quantised and they can have only discrete values of  $nhv$ ,  $2hv$ ,  $3hv$ , ...,  $nhv$ .

- The energy levels of a harmonic oscillator are given by

$$E_n = \left(n + \frac{1}{2}\right)hv = \left(n + \frac{1}{2}\right)\hbar\omega_0 \quad (n = 0, 1, 2, \dots) \quad \dots (1)$$

Here,  $\frac{1}{2}hv$  is called zero point energy.

(2) The atomic oscillators obey Maxwell-Boltzmann distribution law of energies.

The mean energy of the oscillator is

$$\begin{aligned} \bar{E} &= \frac{\sum_{n=0}^{\infty} \left(n + \frac{1}{2}\right)\hbar\omega_0 \exp\left[\frac{-\left(n + \frac{1}{2}\right)\hbar\omega_0}{k_B T}\right]}{\sum_{n=0}^{\infty} \exp\left[\frac{-\left(n + \frac{1}{2}\right)\hbar\omega_0}{k_B T}\right]} \quad \dots (2) \\ &= \frac{\hbar\omega_0 \sum_{n=0}^{\infty} \left(n + \frac{1}{2}\right) \exp\left[\left(n + \frac{1}{2}\right)x\right]}{\sum_{n=0}^{\infty} \exp\left[\left(n + \frac{1}{2}\right)x\right]} \quad \text{where } x = -\frac{\hbar\omega_0}{k_B T} \\ &= \frac{\hbar\omega_0 \left(\frac{1}{2}e^{x/2} + \frac{3}{2} \cdot e^{3x/2} + \frac{5}{2}e^{5x/2} + \dots\right)}{(e^{x/2} + e^{3x/2} + e^{5x/2} + \dots)} \\ &= \hbar\omega_0 \frac{d}{dx} \log\{e^{x/2} + e^{3x/2} + e^{5x/2} + \dots\} \\ &= \hbar\omega_0 \frac{d}{dx} \log\{e^{x/2}(1 + e^x + e^{2x} + \dots)\} \end{aligned}$$

$$\begin{aligned}
 &= \hbar\omega_0 \frac{d}{dx} \left[ \frac{1}{2}x - \log(1 - e^x) \right] \\
 &= \hbar\omega_0 \left[ \frac{1}{2} + \frac{1}{e^{-x} - 1} \right] \\
 &= \hbar\omega_0 \left[ \frac{1}{2} + \frac{1}{\exp\left(\frac{\hbar\omega_0}{k_B T}\right) - 1} \right]
 \end{aligned}$$

The internal energy is obtained by multiplying the average energy per oscillator by the number of oscillators ( $3N$ ).

The total internal energy  $U = 3N\bar{E}$

$$\therefore U = \frac{3N\hbar\omega_0}{2} + \frac{3N\hbar\omega_0}{\left\{ \exp\left(\frac{\hbar\omega_0}{k_B T}\right) - 1 \right\}} \quad \dots (3)$$

and  $C_v = \left( \frac{dU}{dT} \right)_v = 3Nk_B \left( \frac{\hbar\omega_0}{k_B T} \right)^2 \frac{\exp\left(\frac{\hbar\omega_0}{k_B T}\right)}{\left[ \exp\left(\frac{\hbar\omega_0}{k_B T}\right) - 1 \right]^2} \quad \dots (4)$

Put  $\hbar\omega_0 = k_B\Theta_E \quad \dots (5)$

Here  $\Theta_E$  is called "Einstein's temperature".

Substituting Eq. (5) in Eq. (4), we have

$$C_v = 3Nk_B \left( \frac{\Theta_E}{T} \right)^2 \frac{\exp\left(\frac{\Theta_E}{T}\right)}{\left[ \exp\left(\frac{\Theta_E}{T}\right) - 1 \right]^2} \quad \dots (6)$$

(i) For high temperatures  $T \gg \Theta_E$  and  $\exp\left(\frac{\Theta_E}{T}\right) \approx 1$ , then

$$C_v = 3Nk_B = 3R \quad \dots (7)$$

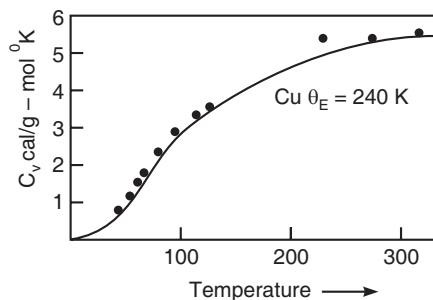
i.e.,  $C_v$  approaches the classical result.

(ii) At low temperatures  $T \ll \Theta_E$  and hence  $\exp\left(\frac{\Theta_E}{T}\right) \gg 1$ , then

$$C_v \approx 3Nk_B \left( \frac{\Theta_E}{T} \right)^2 \frac{\exp\left(\frac{\Theta_E}{T}\right)}{\left[ \exp\left(\frac{\Theta_E}{T}\right) \right]^2} = 3Nk_B \left( \frac{\Theta_E}{T} \right)^2 \exp\left(-\frac{\Theta_E}{T}\right) \quad \dots (8)$$

From equation (8) it is clear that the specific heat approaches zero at  $T$  approaches zero.

If we plot  $C_v$  versus  $T$  using Eq. (6), we obtain a curve of the general shape as Fig. 26.11.



**Fig. 26.11.** Specific heat of copper versus temperature. The dots represent experimental values and the Curve is given by the Einstein's expression.

**EXAMPLE 1.** Calculate the frequency of Einstein oscillator for \$\theta\_E = 236 \text{ K}\$. Given \$k\_B = 1.38 \times 10^{-23} \text{ JK}^{-1}\$ and \$h = 6.6 \times 10^{-34} \text{ J sec.}\$ (M.D.U. 2000)

**SOL.** Einstein temperature \$\theta\_E = 236 \text{ K}\$

$$\text{Now, } \theta_E = \frac{hv}{k}$$

\$\therefore\$ Frequency of Einstein oscillator

$$\begin{aligned} v &= \frac{k\theta_E}{h} = \frac{1.38 \times 10^{-23} \times 236}{6.6 \times 10^{-34}} \\ &= 49.34 \times 10^{11} \text{ Hz} \end{aligned}$$

**EXAMPLE 2.** Calculate the characteristic temperature \$\theta\_E\$ for copper if Einstein frequency \$v\_E = 4.8 \times 10^{12} \text{ Hz}\$. Given \$h = 6.6 \times 10^{-34} \text{ JS}\$ and \$k = 1.4 \times 10^{-23} \text{ JK}^{-1}\$. (M.D.U. 2008)

**SOL.** Characteristic Einstein temperature

$$\theta_E = \frac{hv_E}{k} = \frac{6.6 \times 10^{-34} \times 4.8 \times 10^{12}}{1.4 \times 10^{-23}} = 226 \text{ K}$$

## 26.9 DEBYE'S THEORY OF SPECIFIC HEAT CAPACITY OF A SOLID

### Assumptions.

Debye assumed that :

- (1) Any solid is capable of vibrating elastically in many different modes.
- (2) The frequency of vibration in one mode is different from that in another mode.
- (3) The number of modes of vibration of solids are limited in number.

### Number of independent modes of vibration per unit volume

A solid can be considered as an elastic body in which stationary waves, of both the longitudinal and transverse types, are set up. The frequencies range from 0 to a definite upper limit \$v\_m\$. This limit is chosen so that the total number of possible independent vibrations is equal to the number of vibrations of the separate atoms in the solid.

$$\left. \text{No. of independent modes of vibration per unit volume, with frequencies between } v \text{ and } v + dv \right\} = 4\pi \left( \frac{1}{C_l^3} + \frac{2}{C_t^3} \right) v^2 dv.$$

Here, \$C\_l\$ = velocity of longitudinal waves and \$C\_t\$ = velocity of transverse waves.

### Derivation of Debye's formula

If there are  $N$  atoms in volume  $V$ ,

$$\text{total No. of independent modes of vibration } \} = \int_0^{v_m} 4\pi V \left( \frac{1}{C_l^3} + \frac{2}{C_t^3} \right) v^2 dv.$$

Each atom has 3 degrees of freedom and hence the possible number of vibrations will be equal to  $3N$ . Therefore,

$$\int_0^{v_m} 4\pi V \left( \frac{1}{C_l^3} + \frac{2}{C_t^3} \right) v^2 dv = 3N \quad \dots(1)$$

$$\text{or} \quad 4\pi V \left( \frac{1}{C_l^3} + \frac{2}{C_t^3} \right) = \frac{9N}{v_m^3} \quad \dots(2)$$

According to quantum theory, the energy associated with each degree of freedom is  $\frac{hv}{e^{hv/kT} - 1}$ .

Hence, the energy in the solid for frequencies between  $v$  and  $v + dv$

$$\begin{aligned} &= 4\pi V \left( \frac{1}{C_l^3} + \frac{2}{C_t^3} \right) v^2 dv \times \frac{hv}{e^{hv/kT} - 1} \\ &= \frac{9N}{v_m^3} \frac{hv^3 dv}{(e^{hv/kT} - 1)} \end{aligned} \quad \dots(3)$$

$\therefore$  The total thermal energy associated with the solid

$$U = \frac{9N}{v_m^3} \int_0^{v_m} \frac{hv^3 dv}{(e^{hv/kT} - 1)} \quad \dots(4)$$

Let  $\frac{hv}{kT} = x$ ; then  $v = \frac{kT}{h}x$  and  $dv = \frac{kT}{h}dx$ .

Let us introduce a characteristic temperature  $\theta$  defined by  $\theta = hv_m/k$ .

When  $v = 0$ ,  $x = 0$  and when  $v = v_m$ ,  $x = \frac{hv_m}{kT} = \frac{\theta}{T}$ .

Substituting these values in equation (4),

$$U = 9Nk \frac{T^4}{\theta^3} \int_0^{\theta} \frac{x^3}{e^x - 1} dx \quad \dots(5)$$

$$\begin{aligned} C_v &= \frac{dU}{dT} \\ &= 9R \left[ 4 \left( \frac{T}{\theta} \right)^3 \int_0^{\theta} \frac{x^3 dx}{e^x - 1} - \frac{\theta}{T} \frac{1}{e^{\theta} - 1} \right] \end{aligned} \quad \dots(6)$$

This is Debye's formula for the atomic heat capacity.

The curve  $C_v$  against  $T$  is shown in Fig. 26.12.

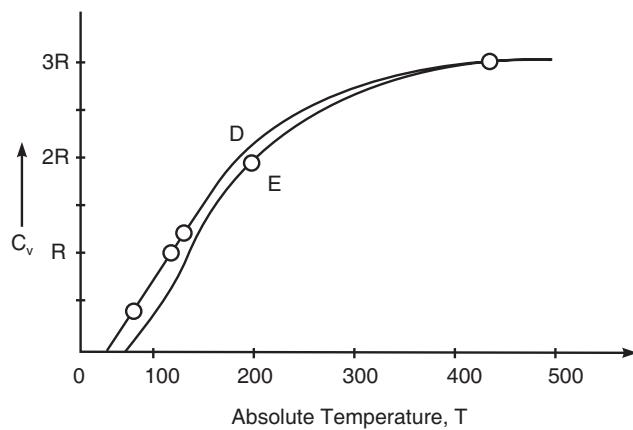


Fig. 26.12

**Test of Debye's Theory.** (1) At high temperatures,  $x = hv/kT$  is small and  $\theta/T$  also is small.

$$\text{Hence, } e^{\frac{\theta}{T}} = 1 + \frac{\theta}{T}. \text{ Thus, } \frac{\theta}{T} - \frac{1}{e^{\frac{\theta}{T}} - 1} = 1.$$

Also  $e^x - 1 = x$ . Hence,

$$4\left(\frac{T}{\theta}\right)^3 \int_0^{\frac{\theta}{T}} \frac{x^3 dx}{e^x - 1} = 4\left(\frac{T}{\theta}\right)^3 \int_0^{\frac{\theta}{T}} x^2 dx = \frac{4}{3}.$$

$$\text{Hence, } C_v = 9R\left(\frac{4}{3} - 1\right) = 3R.$$

This is in agreement with Dulong and Petit law and experiment.

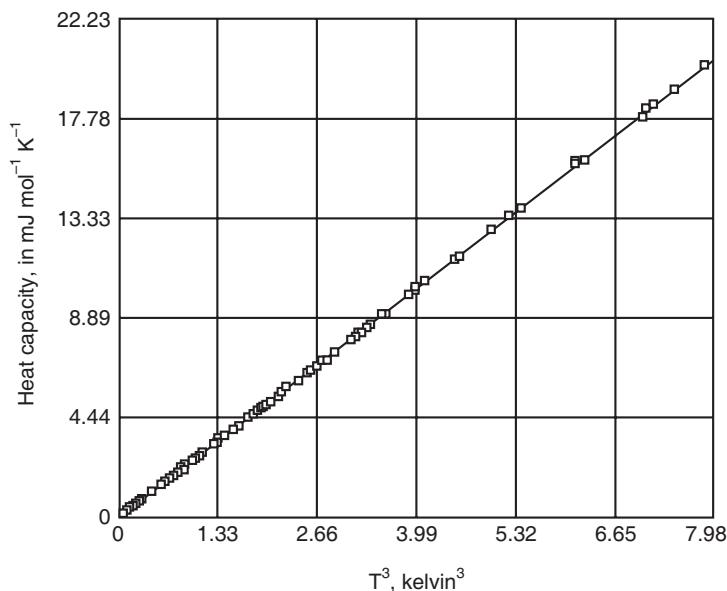
(2) At low temperatures,  $x$  and  $\theta/T$  are large. The exponential  $e^{\theta/T}$  becomes large, so that the last term in equation (6) vanishes. The upper limit for integration can be taken as  $\infty$  instead of  $\theta/T$ . The value of the definite integral is  $\frac{\pi^4}{15}$ . Hence, at low temperatures,

$$C_v = 9R \times 4\left(\frac{T}{\theta}\right)^3 \times \frac{\pi^4}{15} = \frac{12\pi^4}{5} R \left(\frac{T}{\theta}\right)^3$$

$$C_v = 234 R \left(\frac{T}{\theta}\right)^3$$

This is called Debye's  $T^3$  law. Hence, at low temperatures, the atomic heat is directly proportional to the cube of the absolute temperature. This is confirmed by experiment.

Fig. 26.13 shows low temperature heat capacity of solid argon, plotted against  $T^3$ . In this temperature region, the experimental results are in excellent agreement with the Debye  $T^3$  law with  $\theta = 92.0$  K.



**Fig. 26.13**

- According to Debye theory:
    - (1) The atomic heat capacity of all the elements tends to the maximum value of  $3R$  at high temperatures.
    - (2) The atomic heat capacity of all the elements decreases with the fall of temperature.
    - (3) The atomic heat capacity tends to the value zero, near about absolute zero.
    - (4) At low temperatures, the atomic heat capacity varies as the cube of the absolute temperature. Atomic heat capacity  $\propto T^3$ .

**Debye temperature.** The Debye temperature is given by

$$\theta_D = \frac{hv_D}{k}$$

Here,  $\nu_D$  is Debye frequency,  $h$  Planck's constant and  $k$  Boltzmann constant.

## EXERCISE



(B.U.M.Sc. 2006)



(R.U. 2012)

- ### 3. Einstein temperature $\theta_{\text{E}} \equiv$

$$(a) \frac{k}{\hbar\omega} \quad (b) \frac{\hbar\omega}{k} \quad (c) \frac{\omega}{\hbar k} \quad (d) \frac{\hbar k}{\omega}$$

(BU 2014)

4. Which of the following is Debye temperature?

$$(a) \theta_D = \frac{\hbar\omega_D}{k_B} \quad (b) \frac{\hbar\omega_D}{k_B} \quad (c) \frac{\hbar^2\omega_D}{k_B} \quad (d) \frac{\hbar\omega_D^2}{k_B}$$

(B.U. 2015)

5. If the Debye's temperature of metal is 450 K, the Debye's frequency is

$$(a) 10^{13} \text{ Hz} \quad (b) 10^2 \text{ Hz} \quad (c) 10^{23} \text{ Hz} \quad (d) 10 \text{ Hz}$$

(B.U. 2015)

6. At low temperature,  $C_v$  of solid is proportional to \_\_\_\_\_.

$$(a) T^2 \quad (b) T^3 \quad (c) T^5 \quad (d) T^{\frac{1}{2}}$$

(B.U. 2015)

[Ans. 1.(b), 2.(a), 3.(b), 4.(a), 5.(a), 6.(b)]

7. Discuss the vibrations of monatomic linear lattices and compare the results with the vibrations of a continuous string.

8. Discuss the modes of vibration of a finite one-dimensional lattice of identical atoms.

9. Discuss the vibrations of a diatomic linear lattice. Show that the vibrational spectrum consists of two branches. What are the main features of these branches?

10. Derive the vibrational modes of a diatomic linear lattice. What is the difference between the two branches and why are they named so? (Indore, 1980)

11. Discuss the vibrations of diatomic lattice and describe its optical and acoustic modes. (Delhi, 1983)

12. Give the basic assumptions of Dulong and Petit's law. (B.U. 2010)

13. State and explain Dulong and Petit's law. (B.U. 2014)

14. Using Einstein model, derive the expression for the specific heat of a solid. Discuss its merits and demerits. (B.U. 2012)

15. Obtain the expression for Debye's frequency. (B.U. 2010)

16. Discuss the salient features of Debye's theory of specific heat of solids.

# FREE ELECTRON THEORY AND BAND THEORY

**After reading this chapter, you should be able to**

- ◆ Utilize the free electron model to understand many physical properties of metals, that depend essentially on the kinetic properties of the conduction electrons
- ◆ Explain why condensed matter is so transparent to conduction electrons
- ◆ Discuss Drude's model, Sommerfeld's model, free electron Fermi gas, and Schottky effect
- ◆ Discuss the dynamics of electrons in the presence of a magnetic field (Hall effect)
- ◆ Explain qualitatively band theory of solids
- ◆ Explain how the band theory leads to the classification of solids into metals, semiconductors and insulators
- ◆ Explain how the band theory differs from the free electron model in explaining the properties of metals

## FREE ELECTRON THEORY OF METALS

### 27.1 DRUDE-LORENTZ THEORY OF ELECTRICAL CONDUCTION

The basic assumptions of Lorentz-Drude theory are:

1. There are a large number of free electrons in a metal. These electrons are free to move about the whole volume of the metal like the molecules of a perfect gas in a container. These electrons are called *conduction electrons*. The positive ions in the metal occupy fixed positions.
2. These electrons (conduction electrons) are treated as an *electron gas*.
3. They applied the Maxwell and Boltzmann statistics to the electron gas.
4. In the absence of an electric field, the free electrons move in *random directions*. They collide with other free electrons and positive ions fixed to the lattice (Fig. 27.1). All the collisions are elastic *i.e.*, there is no loss of energy.
5. The free electrons move in a completely *uniform potential field* due to ions fixed in the lattice.

6. When an electric field is applied between the ends of a metal, the free electrons are slowly drifting towards the positive potential. The free electrons move in a direction opposite to the direction of the electric field (Fig. 27.2). The *free electrons* acquire a constant average velocity called "*drift velocity*".
7. The total energy of free electrons is equal to its kinetic energy as its potential energy is zero.
8. The mutual repulsion between the free electrons is considered to be negligible.

#### Drude-Lorentz classical free electron theory

Drude and Lorentz developed this theory

**Principle.** A metal consists of a very large number of free electrons. These free electrons can move freely throughout the volume of the metal. They are fully responsible for the electrical conduction in the metal.

#### Basic Definitions

1. **Drift velocity ( $v_d$ ).** The average velocity acquired by the free electrons of a metal after steady state is reached in the presence of an electric field is defined as the drift velocity.
2. **Collision time ( $\tau_c$ ).** The average time taken by a free electron between any two successive collisions is defined as collision time.

$$\tau_c = \frac{\lambda}{v_d}$$

3. **Mean free path ( $\lambda$ ).** The average distance travelled by a free electron between any two successive collisions in the presence of an applied electric field is defined as mean free path.

It is the product of drift velocity ( $v_d$ ) of free electrons and collision time ( $\tau_c$ ).

$$\lambda = v_d \times \tau_c$$

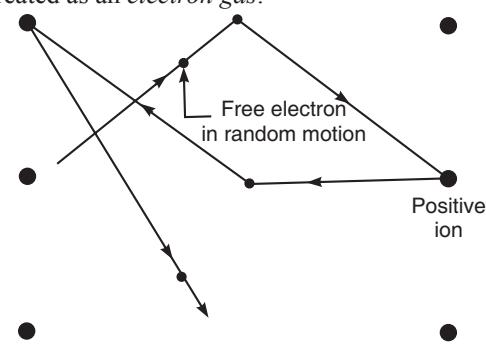


Fig. 27.1. Random Motion of Electron in the absence of an electric field

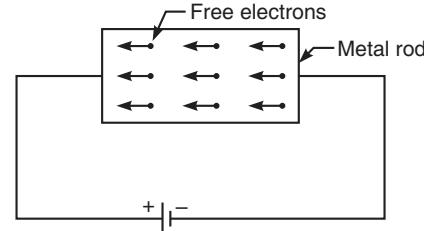


Fig. 27.2. Drift Velocity of Electron – Presence of electric field (E)

4. **Relaxation time ( $\tau$ ).** Relaxation time ' $\tau$ ' is defined as the time taken by the free electron to reach its equilibrium position from its disturbed position in the presence of the electric field.

For an isotropic collision, the relaxation time and collision time are equal.

For an isotropic solid, like a metal,  $\tau = \tau_c$ .

## 27.2 EXPRESSION FOR ELECTRICAL CONDUCTIVITY (DERIVATION)

An electric field ( $E$ ) is applied to a metallic rod (Fig. 27.3).

Let

$m$  = mass of the electron;

$v$  = mean velocity of the electron corresponding to temperature equilibrium with the metal and

$T$  = absolute temperature of the metal.

According to kinetic theory,

$$\frac{1}{2}mv^2 \propto T$$

or  $\frac{1}{2}mv^2 = \frac{3}{2}k_B T \quad \dots (1)$

Here,  $k_B$  is Botzmann constant.

The electron experiences a force  $Ee$ .

Due to this force, the electron gains acceleration  $a$ .

By Newton's second law of motion,

$$F = ma.$$

$$ma = Ee$$

$$\therefore a = \frac{Ee}{m} \quad \dots (2)$$

If  $\lambda$  is the mean free path of electron, then the time  $\tau$  between two successive collisions is given by

$$\tau = \frac{\lambda}{v}. \quad \dots (3)$$

The average distance travelled by electron in time  $\tau$  is

$$s = \frac{1}{2}a\tau^2$$

$$s = \frac{1}{2}\left(\frac{Ee}{m}\right)\frac{\lambda^2}{v^2}. \quad \dots (4)$$

The average drift velocity  $v_d$  of the free electron parallel to the direction of the field is

$$v_d = \frac{s}{\tau} = \frac{1}{2}\left(\frac{Ee}{m}\right)\frac{\lambda^2}{v^2} \cdot \frac{v}{\lambda} = \frac{1}{2}\left(\frac{Ee}{m}\right)\frac{\lambda}{v}$$

Substituting the value of  $m$  from Eq. (1),

$$v_d = \frac{1}{2} \frac{Eev^2}{3k_B T} \cdot \frac{\lambda}{v}$$

$$\therefore v_d = \frac{Ee\lambda v}{6k_B T} \quad \dots (5)$$

Let  $n$  be the number of electrons per unit volume of the metal.

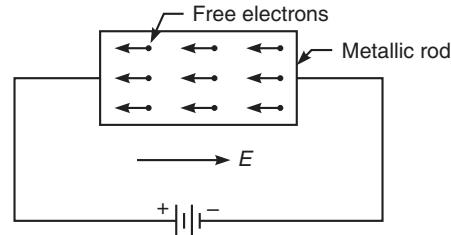


Fig. 27.3

The current density in the metal =  $J = nev_d$  ... (6)  
 Substituting the value of  $v_d$  from Eq. (5),

$$J = \frac{nEe^2\lambda v}{6k_B T} \quad \dots (7)$$

$$\therefore J \propto E \quad \dots (8)$$

This shows that the current is proportional to applied field which is *Ohm's law*.

Electrical conductivity of the metal =  $\sigma = \frac{\text{Current density}}{\text{Electric field}}$

$$\therefore \sigma = \frac{J}{E} = \frac{ne^2\lambda v}{6k_B T} \quad \dots (9)$$

$n$  is different for different metals. Hence the difference in the conductivity of different metals is due to the difference in the number of the free electrons in them.

#### Limitations of the Theory

- (1) It does not give a satisfactory explanation as to why certain metals like copper and silver have high conductivity.
- (2) It is unable to explain adequately the variation of resistance of pure metals at the temperature of liquid helium.

### 27.3 EXPRESSION FOR THERMAL CONDUCTIVITY

The conduction of heat in a metal is also mostly through free electrons. Consider any plane  $AB$  in a conductor normal to the direction of flow of heat (Fig. 27.4). Consider two parallel planes  $CD$  and  $EF$ , on either side of  $AB$  at a distance equal to the mean free path of the electrons  $\lambda$ .

- Let the temperatures of the planes  $CD$  and  $EF$  be  $T_1$  and  $T_2$  ( $T_1 > T_2$ ).
- The distance between the two planes =  $2\lambda$ .
- The K.E. of the molecules at the hotter end is greater than that of the molecules at the colder end.
- In collision, the electrons near  $CD$  lose their kinetic energy. The electrons near  $EF$  gain the energy.
- The conduction of heat through the element is due to this transference of energy.

Let  $v_1$  and  $v_2$  be the velocities of the electron corresponding to the temperature  $T_1$  and  $T_2$ . Then

$$\frac{1}{2}mv_1^2 = \frac{3}{2}k_B T_1 \text{ and } \frac{1}{2}mv_2^2 = \frac{3}{2}k_B T_2. \quad \dots (1)$$

If  $v$  is the mean velocity of each electron while crossing  $AB$ , the number of electrons crossing unit area of section  $AB$  per second is  $(n/6)v$ . ( $n$  = the number of electrons in unit volume).

$$\left. \begin{array}{l} \text{Heat energy transferred across unit area of} \\ \text{the plane } AB \text{ per second from left to right} \end{array} \right\} = \frac{n}{6}v \times \frac{1}{2}mv_1^2 = \frac{n}{6}v \cdot \frac{3}{2}k_B T_1.$$

$$\left. \begin{array}{l} \text{Heat energy transferred across unit area of} \\ \text{the plane } AB \text{ per second from right to left} \end{array} \right\} = \frac{n}{6}v \cdot \frac{3}{2}k_B T_2.$$

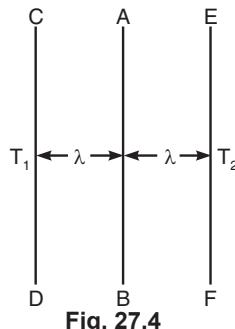


Fig. 27.4

$$\therefore \text{Net heat energy transferred per unit area of the plane } AB \text{ per second} = \frac{n}{6} v \cdot \frac{3}{2} k_B (T_1 - T_2) \quad \dots (2)$$

$$\text{Temperature gradient} = (T_1 - T_2)/2\lambda. \quad \dots (3)$$

The coefficient of thermal conductivity ( $K$ ) is the ratio of the quantity of heat transferred across until area in one second to the temperature gradient.

$$K = \frac{(n/6)v \cdot \frac{3}{2} k_B (T_1 - T_2)}{(T_1 - T_2)/2\lambda}$$

$$K = \frac{1}{2} nv \lambda k_B. \quad \dots (4)$$

The value of  $K$  obtained with the help of Eq. (4) is verified experimentally and the free electron theory is found to be successful to explain the thermal conductivity.

**Wiedemann-Franz Law: Statement.** *The ratio of thermal conductivity to electrical conductivity is a constant for all metals and is proportional to the absolute temperature.*

**Explanation.**  $\frac{\text{Thermal conductivity}}{\text{Electrical conductivity}} = \frac{K}{\sigma} = \frac{\frac{1}{2} nv \lambda k_B}{\frac{ne^2 \lambda v}{6k_B T}}$

$$\frac{K}{\sigma} = 3 \left( \frac{k_B}{e} \right)^2 T.$$

Since  $3 \left( \frac{k_B}{e} \right)^2 = \text{a constant}, \frac{K}{\sigma} \propto T.$

$$\frac{K}{\sigma T} = 3 \left( \frac{k_B}{e} \right)^2. \text{ This quantity is called Lorentz's number.}$$

This relation shows that the ratio  $K/\sigma$  is independent of the nature of the metal and is proportional to the absolute temperature.

Among the metals, the best electrical conductors ( $Ag$ ,  $Cu$ ,  $Al$ ,  $Au$ ) are also the best heat conductors.

#### *Merits of classical free electron theory*

- (i) It verifies Ohm's law.
- (ii) It explains the electrical and thermal conductivities of metals.
- (iii) It derives Wiedemann-Franz law.
- (iv) It explains the optical properties of metals.

#### *Drawbacks of classical free electron theory*

(1) According to the classical free electron theory, the value of electronic specific heat is  $\left(\frac{3}{2}\right)R$ . But experimental value is about  $0.01 R$  only.

**Failure of the theory.** The free electron theory completely fails to explain the heat capacity of conduction electrons. The free electron theory assumes that all the valence electrons in a metal can absorb thermal energy. Consider 1 kmol of a metal. There will be  $N_A$  free electrons, where  $N_A$  is Avogadro's number. Each atom contributes one valence electron to free electron gas.

Total energy of the electrons is given by

$$U = \frac{3}{2} N_A k_B T, \quad \dots (1)$$

Here,  $\left(\frac{3}{2}\right)k_B T$  is average kinetic energy of every free electron. This is in accordance with the law of equipartition of energy.

When heat is supplied to the material, the free electrons also absorb part of the heat. Now, the electronic specific heat is given by

$$(C_v)_{\text{electronic}} = \left( \frac{dU}{dT} \right) = \frac{3}{2} N_A k_B = \frac{3}{2} R \quad \dots (2)$$

where  $N_A k_B = R$ .

$$\begin{aligned} \therefore (C_v)_{\text{electronic}} &= \frac{3}{2} \times (6.02 \times 10^{26}) \times (1.38 \times 10^{-23}) \\ &= 12.5 \text{ kJ/mol/K} \end{aligned} \quad \dots (3)$$

The above molar electronic specific heat due to free electrons is about hundred times greater than experimentally predicted value.

(2) Let us consider the ratio of thermal conductivity to electrical conductivity, i.e., Wiedemann and Franz ratio.

$$\frac{K}{\sigma} = 3 \left( \frac{k_B}{e} \right)^2 T = LT, \quad \dots (4)$$

Here,  $L = 3 \left( \frac{k_B}{e} \right)^2$  is Lorentz number. Its value comes out to be  $1.12 \times 10^{-8}$  watt ohm/kelvin<sup>2</sup>.

For copper at 20°C, electrical conductivity and thermal conductivity are  $5.81 \times 10^7$  and 386 W/mK respectively. Hence,

$$L = \frac{386}{(5.81 \times 10^7)(293)} = 226 \times 10^{-8} \text{ W}\Omega\text{K}^{-2}$$

This does not agree with the value calculated from classical formula.

So the classical assumption that all the free electrons of metal participate in thermal conduction is not correct. The results of quantum theory with Fermi-Dirac statistics show that only those electrons which are near the Fermi-level participate in thermal and electrical conductivities.

## 27.4 ELECTRICAL CONDUCTIVITY: ALTERNATIVE ANALYSIS

The momentum of a free electron is related to the wavevector by

$$mv = \hbar k \quad \dots (1)$$

In an electric field  $\mathbf{E}$  and magnetic field  $\mathbf{B}$ , the force  $\mathbf{F}$  on an electron of charge  $-e$  is

$$-e[\mathbf{E} + \mathbf{v} \times \mathbf{B}] \quad \dots (2)$$

Newton's second law of motion becomes

$$\mathbf{F} = m \frac{d\mathbf{v}}{dt} = \hbar \frac{d\mathbf{k}}{dt} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

Let

$$\mathbf{B} = 0.$$

$$\hbar \frac{d\mathbf{k}}{dt} = -e\mathbf{E} \quad \dots (3)$$

$$\text{Integrating, } \mathbf{k}(t) - \mathbf{k}(0) = -\frac{e\mathbf{E}t}{\hbar} \quad \dots (4)$$

In the absence of collisions the Fermi sphere (Fig. 27.5) in  $\mathbf{k}$  space is displaced at a uniform rate by a constant applied electric field.

Let us consider electrical conductivity in  $\mathbf{k}$  space.

**(i) In the absence of an electric field.** In the absence of an electric field, the Fermi sphere is centered at the origin (Fig. 27.5 a).

The Fermi sphere encloses the occupied electron orbitals in  $\mathbf{k}$  space in the ground state of the electron gas. The net momentum is zero, because for every orbital  $\mathbf{k}$  there is an occupied orbital at  $-\mathbf{k}$ .

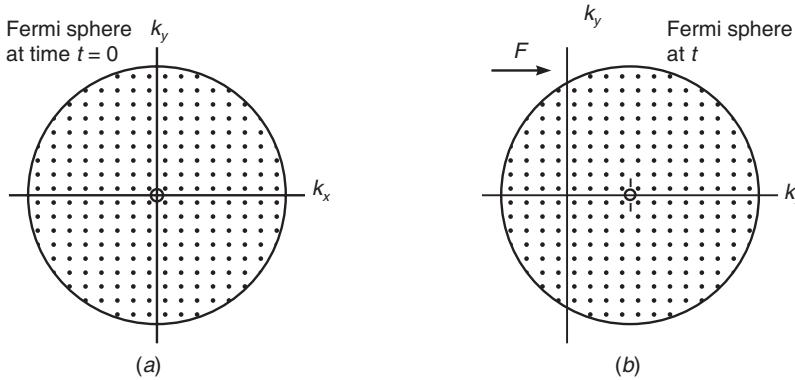


Fig. 27.5

**(ii) When an electric field is applied.** Suppose the electric field is applied at time \$t = 0\$ to an electron gas that fills the Fermi sphere centered at the origin of  $\mathbf{k}$  space. Then, at a later time \$t\$ the sphere will be displaced to a new centre at

$$\delta\mathbf{k} = -eEt/\hbar \quad \dots(5)$$

Notice that the Fermi sphere is displaced as a whole [Fig. 27.5(b)]. Thus in  $\mathbf{k}$  space electrical conduction is associated with a slight displacement of the Fermi sphere about the origin.

Because of collisions of electrons with impurities, lattice imperfections and phonons, the displaced sphere may be maintained in a steady state in an electric field.

Let  $\tau$  = collision time.

Then, the displacement of the Fermi sphere in the steady state is given by Eq. (5) with  $t = \tau$ .

The incremental velocity is

$$v = -eE\tau/m \quad \dots(6)$$

Let there be \$n\$ electrons of charge \$q = -e\$ per unit volume.

The electric current density is

$$\mathbf{j} = nqv = ne^2\tau\mathbf{E}/m \quad \dots(7)$$

This is Ohm's law.

The electrical conductivity \$\sigma\$ is defined by \$\mathbf{j} = \sigma\mathbf{E}\$.

$$\therefore \sigma = \frac{ne^2\tau}{m} \quad \dots(8)$$

The electrical resistivity \$\rho\$ is defined as the reciprocal of the conductivity.

$$\therefore \rho = \frac{m}{ne^2\tau} \quad \dots(9)$$

It is easy to understand the result Eq. (8) for the conductivity. We expect the charge transported to be proportional to the charge density \$ne\$. The factor \$e/m\$ enters because the acceleration in a given electric field is proportional to \$e\$ and inversely proportional to the mass \$m\$. The time \$\tau\$ describes the free time during which the field acts on the carrier.

## 27.5 EXPERIMENTAL ELECTRICAL RESISTIVITY OF METALS

The electron undergoes a collision only because the lattice is not perfectly regular. It is convenient to group the deviations from a perfect lattice into two classes :

(a) Lattice vibrations (phonons) of the ions around their equilibrium position due to thermal excitations of the ions [Fig. 27.6(a)].

(b) All static imperfections, such as foreign impurities and vacant lattice sites [Fig. 27.6(b)].

The electrical resistivity of most metals is dominated at room temperature (300 K) by collisions of the conduction electrons with lattice phonons and at liquid helium temperature (4K) by collisions with impurity atoms and mechanical imperfections in the lattice.

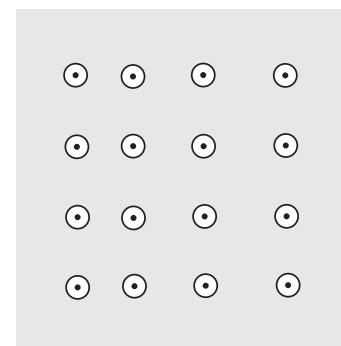
**To find the net resistivity  $\rho$  :**

(i) The probabilities of electrons being scattered by phonons and by impurities are additive, since these two mechanisms are assumed to act independently.

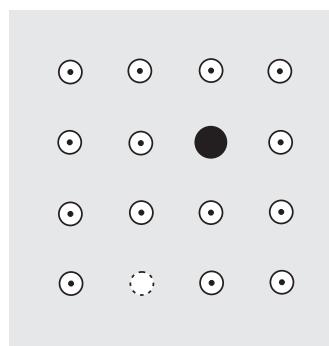
(ii) Since the scattering probability is inversely proportional to the relaxation time, the reciprocal of the effective relaxation time is the sum of the reciprocals of the individual relaxation times.

Therefore, we write

$$\frac{1}{\tau} = \frac{1}{\tau_L} + \frac{1}{\tau_i} \quad \dots(1)$$



(a)

(b)  
Fig. 27.6

Here, the first term on the right is due to phonons and the second is due to impurities.

(i)  $\tau_L$  is expected to depend on  $T$ .

(ii)  $\tau_i$  depends on impurities, but not on  $T$ .

The general formula for electrical resistivity is

$$\rho = 1/\sigma = m/ne^2 \tau \quad \dots(2)$$

The net resistivity is given by

$$\rho = \rho_L + \rho_i = \frac{m}{ne^2} \frac{1}{\tau_L} + \frac{m}{ne^2} \frac{1}{\tau_i} \quad \dots(3)$$

(i)  $\rho_L$  is the resistivity caused by the thermal phonons. Hence it is temperature dependent, and is called the *ideal resistivity*. It is the resistivity of a pure specimen.

The temperature-dependent part of the electrical resistivity is proportional to the rate at which an electron collides with thermal phonons and thermal electrons.

Often  $\rho_L$  is independent of the number of defects when their concentration is small.

(ii)  $\rho_i$  is the resistivity caused by scattering of the electron waves by static defects that disturb the periodicity of the lattice. Often  $\rho_i$  is independent of temperature. The residual resistivity,  $\rho_i(0)$ , is the extrapolated resistivity at 0K because  $\rho_L$  vanishes as  $T \rightarrow 0$ .

Fig. 27.7 shows resistance of potassium below 20K, as measured on two specimens. The different intercepts at 0K are attributed to different concentrations of impurities and static imperfections in the two specimens.

The lattice resistivity,  $\rho_L(T) = \rho - \rho(0)$ , is the same for different specimens of a metal, even though  $\rho_i(0)$  may itself vary widely. The resistivity ratio of a specimen is defined as the ratio of its resistivity at room temperature to its residual resistivity. It is a convenient approximate indicator of sample purity. In exceptionally pure specimens the resistivity ratio may be as high as 106, whereas in some alloys (e.g., manganin) it may be as low as 1.1.

**Matthiessen's rule.** Eq. (3) tells us that the resistivity in the presence of several distinct scattering mechanisms is simply the sum of the resistivities one would have if each alone were present. This is called *Matthiessen's rule*. In other words, this rule states that  $\rho$  can be split into two parts, one of which is independent of  $T$ . Matthiessen's rule is convenient in analyzing experimental data (Fig. 27.7).

#### Temperature Dependence of Electrical Resistivity of Metals

In a pure metal, we expect that  $\rho \rightarrow 0$  as  $T \rightarrow 0$ . However, experimental observations show that  $\rho$  for metals decreases to a small but finite value and remains practically constant and independent of temperature in the range 10 K to 0K (Fig.27.8).

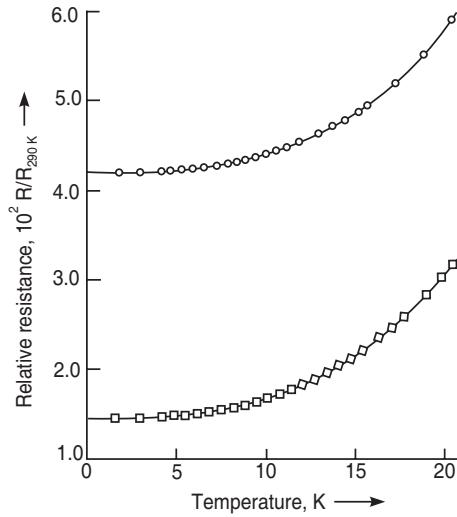


Fig. 27.7

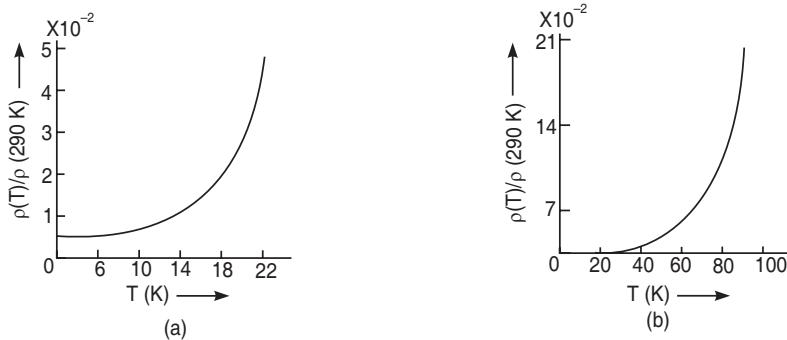


Fig. 27.8

The resistivity at 0K is called the *residual resistivity* of the metal. This resistivity may be attributed to

- (i) presence of impurity atoms in small concentrations, and
- (ii) geometrical imperfections like point defects, grain boundaries and stacking faults in the metal.

Fig. 27.8 shows the normalised resistivity  $\rho(T)/\rho(290 \text{ K})$  versus  $T$  for Na (a) in low temperature region and (b) at higher temperatures,  $\rho(290) \simeq 2.10 \times 10^{-8} \Omega\text{-m}$ .

(a) At very low temperatures, scattering by phonons is negligible, because the amplitudes of oscillations are very small in that region.

$$\tau_L \rightarrow \infty, \rho_L \rightarrow 0.$$

$$\therefore \rho = \rho_i = \text{a constant.}$$

This is in agreement with Fig. 27.8 (a).

(b) As  $T$  increases, scattering by phonons becomes more effective, and  $\rho_L(T)$  increases ; this is why  $\rho$  increases. When  $T$  becomes sufficiently large scattering by phonons dominates and  $\rho = \rho_L(T)$ . In the high temperature region,  $\rho_L(T)$  increases linearly with  $T$ . This is again in agreement with experimental results shown in Fig. 27.8 (b).

We expect that  $\rho_i$  should increase with impurity concentrations, and indeed it could be shown that for small concentrations  $\rho_i$  is proportional to the impurity concentration  $N_i$ . We also remark that, for small impurity concentration  $\rho_L \gg \rho_i$  except at very low temperature.

## 27.6 SOMMERFELD'S FREE ELECTRON MODEL OF METALS

Sommerfeld modified the free electron model by taking into account quantum statistics and the Pauli exclusion principle.

### Assumptions of Sommerfeld model

The assumptions of Sommerfeld model are:

1. The free electrons are the valence electrons of the composing atoms. Thus the alkali metals are assumed to contain one free electron per atom while aluminium has three free electrons per atom.
2. A valence electron in metal finds itself in the field of all ions and that of all other electrons. In this model, the mutual repulsion between the electrons is neglected and the potential field representing the attractive interaction of ions is assumed to be completely uniform everywhere inside the solid.
3. Since the electron emission can not be observed at room temperature, the potential energy of an electron at rest inside the metal is assumed to be lower than that of an electron outside the metal.

Thus the *potential energy for an electron is periodic*. The free electrons find themselves in a potential which is constant everywhere inside the metal but the potential outside the metal is very high which restrains the conduction electrons to remain inside the metal at room temperature. Thus the metal surfaces are the sites of a sharp potential barrier of height  $E_s$  as shown Fig. 27.9. This figure represents a *physical model for metal* where the interior is represented as consisting of a gas of electrons confined by potential barriers within a certain region of space called a '*potential energy box*' of depth  $E_s$ . Thus  $E_s$  represents the energy difference between an electron at rest inside the metal and one at rest in vacuum (*i.e.*, on the surface). At  $T=0$ , all energy levels upto  $E_F$  are filled while all higher ones are empty. Thus work function of metal  $\phi = E_s - E_F$ .

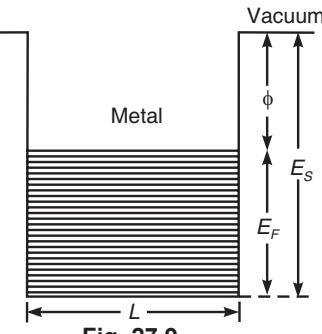


Fig. 27.9

$E_F$  is called the *Fermi energy* and is defined as the maximum kinetic energy that a free electron can have at the absolute zero of temperature.

4. The 'free electron gas' in a solid does not obey the classical Maxwell Boltzmann statistical law. The distribution of energy in an electron gas obeys the *Fermi-Dirac quantum statistics*.
  - FD statistics leads to the conclusion that the energy possessed by an electron is quantized.
  - 5. In a metal, the energy levels are filled in accordance with the *Pauli's exclusion principle*:
    - An energy level can accommodate at the most two electrons, one with spin up and the other with spin down.
    - Thus in filling the energy levels, two electrons occupy the lowest level, two more the next level and so forth until all the electrons in the metal have been accommodated.

### Wave function and energy eigenvalues

(i) **One dimensional box.** Assuming free electrons within a metal ( $V_0 = 0$ ) and using solution of Schrodinger equation for this case, we get the allowed energy values of electrons within the metal as

$$E_n = \frac{n^2 \pi^2 \hbar^2}{2m L^2} \quad (n = 1, 2, 3\dots)$$

These are known as *energy eigenvalues*.

The eigenfunction is

$$\psi_n(x) = \sqrt{\left(\frac{2}{L}\right)} \sin \frac{n\pi x}{L} \quad (\text{See Section 8.3 in chapter 8})$$

**Density of states.** It is defined as the number of electronic states (or quantum states) per unit energy range.

As electron can occupy two spin states, there are two quantum states for a single electron. Therefore, density of states for this case would be given by

$$D(E_n) = \frac{dn}{dE_n} = 2 \cdot \frac{L}{\pi \hbar} \left( \frac{m}{2E_n} \right)^{1/2}$$

This result is plotted in Fig. 27.10.

### Three dimensional box.

If we consider the free electron gas in three dimensions and extend the above mathematical formulation, we have Schrodinger equation in three dimensions as

$$\nabla^2 \psi + \frac{2m}{\hbar^2} E \psi = 0 \quad (\text{Since } V_0 = 0)$$

Consider the electrons to be confined in a cube of length  $L$ . Then, energy eigenvalues

$$E(n_x, n_y, n_z) = \frac{\pi^2 \hbar^2}{2m L^2} (n_x^2 + n_y^2 + n_z^2)$$

$$\text{eigenfunction } \psi(x, y, z) = \sqrt{\left(\frac{8}{L^3}\right)} \sin \frac{n_x \pi x}{L} \sin \frac{n_y \pi y}{L} \sin \left(\frac{n_z \pi z}{L}\right)$$

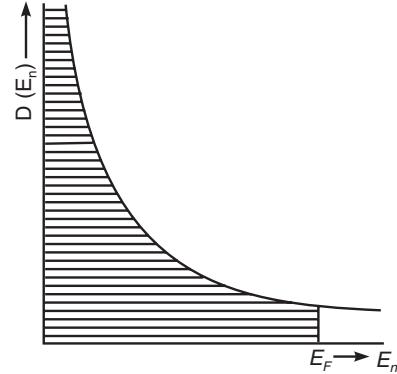
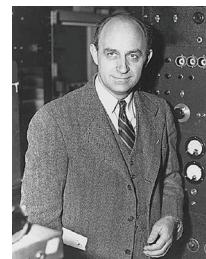


Fig. 27.10

## 27.7 FERMI-ENERGY

Let us consider  $N$ -non-interacting electrons contained in a box at absolute zero temperature. Using Pauli's exclusion principle on the energy distribution of conduction electrons, it can be shown that all the energy levels below a certain level will be filled with electrons and all the levels above it will be empty. The level which divides the filled and empty levels is called the *Fermi-level at absolute zero* and the corresponding energy of that level is called the *Fermi-energy* and is denoted by  $\epsilon_F(0)$ .

Let  $N$  non-interacting electrons be distributed among quantum states such that there are  $n_1, n_2, \dots, n_i \dots$  number of particles in quantum states possessing constant energies  $\epsilon_1, \epsilon_2, \dots, \epsilon_i \dots$  respectively. Since electrons are Fermi particles, they obey Pauli's exclusion principle and Fermi-Dirac distribution law.



Enrico Fermi (1901 – 1954)

According to this law, the number of electrons  $n_i$  with the energy  $\varepsilon_i$  is given by

$$n_i = \frac{g_i}{e^\alpha e^{\varepsilon_i/kT} + 1} \quad \dots(1)$$

where  $g_i$  is the number of quantum states having the same energy  $\varepsilon_i$ .

The number of particle states lying between momentum  $p$  and  $p + dp$  is given by

$$g(p) dp = g_s \cdot \frac{4\pi p^2}{h^3/V} dp = g_s \cdot \frac{4\pi V p^2 dp}{h^3} \quad \dots(2)$$

where  $g_s$  is the *spin degeneracy factor* caused by the particle spin  $s$ .

$$g_s = (2s + 1) = 2 \text{ for electron gas since } s = 1/2 \quad \dots(3)$$

This means that each level can be occupied by two particles one with spin up and the other with spin down ( $\uparrow\downarrow$ ).

Now

$$\varepsilon = p^2/2m \text{ or } dp = (m/2\varepsilon)^{1/2} d\varepsilon.$$

The number of states between energy  $\varepsilon$  and  $\varepsilon + d\varepsilon$  is given by

$$g(\varepsilon) d\varepsilon = 2 \frac{4\pi V (2m\varepsilon) \left(\frac{m}{2\varepsilon}\right)^{1/2} d\varepsilon}{h^3} \quad \dots(4)$$

There is a continuous distribution of electron energies in the solid. In this situation the F-D distribution law becomes

$$n(\varepsilon) d\varepsilon = \frac{g(\varepsilon) d\varepsilon}{e^\alpha e^{\varepsilon/kT} + 1}, \quad \dots(5)$$

Here  $n(\varepsilon) d\varepsilon$  is the number of electrons having energies between  $\varepsilon$  and  $\varepsilon + d\varepsilon$ .

For a gas of fermions at temperature  $T$  in a volume  $V$ , the mean number of fermions in the energy range between  $\varepsilon$  and  $\varepsilon + d\varepsilon$  is given by

$$n(\varepsilon) d\varepsilon = \frac{8\pi m V}{h^3} \sqrt{2m} \cdot \frac{\varepsilon^{1/2} d\varepsilon}{e^{\alpha+(\varepsilon/kT)} + 1} \quad \dots(6)$$

We introduce the *Fermi-distribution function*  $f(\varepsilon)$  defined by

$$f(\varepsilon) = \frac{n(\varepsilon)}{g(\varepsilon)} = \frac{1}{e^{\alpha+(\varepsilon/kT)} + 1} = \frac{1}{e^{(\varepsilon - \varepsilon_F)/kT} + 1} \quad \dots(7)$$

where  $\alpha$  in terms of *Fermi-energy*  $\varepsilon_F(T)$  at temperature  $T$  is given by

$$\alpha = -\frac{\varepsilon_F(T)}{kT} \quad \dots(8)$$

Let us consider the situation at absolute zero of temperature.

At  $T = 0$ ,  $(\varepsilon - \varepsilon_F)/kT \rightarrow -\infty$  (when  $\varepsilon < \varepsilon_F$ ) or  $+\infty$  (when  $\varepsilon > \varepsilon_F$ ). Thus at  $T = 0$

$$\begin{aligned} f(\varepsilon) &= \frac{1}{e^{-\infty} + 1} = 1 \text{ for } \varepsilon < \varepsilon_F(0) \\ &= \frac{1}{e^{\infty} + 1} = 0 \text{ for } \varepsilon > \varepsilon_F(0) \end{aligned} \quad \dots(9)$$

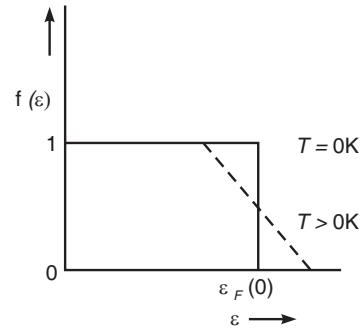


Fig. 27.11

Fig. 27.11 shows plot of Fermi-Dirac distribution function at  $T = 0\text{K}$  (solid line) and at low temperature ( $kT/\varepsilon_F \ll 1$ ) (dotted line).

At any temperature  $T$ ,

$$f(\varepsilon) = 1/2 \text{ for } \varepsilon = \varepsilon_F.$$

The value of  $\varepsilon_F$  is determined by the condition that the total number of particles is constant at a given temperature  $T$ .

The expression for the number of fermions with energy between  $\varepsilon$  and  $\varepsilon + d\varepsilon$  is written as

$$n(\varepsilon) d\varepsilon = f(\varepsilon) g(\varepsilon) d\varepsilon = \frac{8\pi m V}{h^3} \sqrt{2m} \frac{\varepsilon^{1/2} d\varepsilon}{e^{(\varepsilon-\varepsilon_F)/kT} + 1} \quad \dots(10)$$

This expression is called *Fermi-Dirac energy distribution law*.

Thus at absolute zero all states with  $0 < \varepsilon \leq \varepsilon_F(0)$  are completely filled and states with  $\varepsilon > \varepsilon_F(0)$  are empty. The value  $\varepsilon_F(0)$  of  $\varepsilon_F$  at  $T = 0\text{K}$  is determined by

$$\begin{aligned} N &= \int_0^{\varepsilon_F(0)} n(\varepsilon) d\varepsilon = \int_0^{\varepsilon_F(0)} f(\varepsilon) g(\varepsilon) d\varepsilon = \int_0^{\varepsilon_F(0)} g(\varepsilon) d\varepsilon \\ &= \int_0^{\varepsilon_F(0)} \frac{8\pi m V}{h^3} (2m)^{1/2} \varepsilon^{1/2} d\varepsilon = \frac{8\pi m V}{h^3} (2m)^{1/2} \frac{2}{3} [\varepsilon_F(0)]^{3/2} \\ i.e., \quad \varepsilon_F(0) &= \frac{h^2}{2m} \left[ \frac{3N}{8\pi V} \right]^{2/3} = \frac{h^2}{8m} \left[ \frac{3N}{\pi V} \right]^{2/3} \quad \dots(11) \\ \therefore \quad \varepsilon_F(0) &= (3\pi^2)^{2/3} \frac{\hbar^2}{2m} \left( \frac{N}{V} \right)^{2/3} \end{aligned}$$

**Fermi energy.** The Fermi energy is defined as the energy corresponding to the topmost filled level. Fermi energy is determined by the concentration of electrons  $\left( \frac{N}{V} \right)$  and mass ( $m$ ).

**Fermi level.** The energy spectrum of three dimensional free electron gas is discrete. At absolute zero temperature, all levels below a certain level will be filled and all the levels above it will be empty. The level dividing filled and vacant levels is called the Fermi level [ $\varepsilon_F(0)$ ].

The plot of  $n(\varepsilon)$  against  $\varepsilon$  at  $T = 0$  is shown in Fig. 27.12. After rising along a parabola from the origin, it drops abruptly to zero at  $\varepsilon = \varepsilon_F$ .

A typical plot of  $n(\varepsilon)$  against  $\varepsilon$  at a much higher temperature  $T = 1200\text{ K}$  is also shown in Fig. 27.12. Since at any temperature the area under the corresponding curve gives the total number of electrons in the metal, the two shaded areas must be equal. It is seen that even a considerable rise in temperature changes the energy state of only a very small fraction of the electrons.

Let us now define a temperature  $T_F$  called *Fermi-temperature*, by

$$T_F = \frac{\varepsilon_F(0)}{k} = (3\pi^2)^{2/3} \frac{\hbar^2}{2mk} \left( \frac{N}{V} \right)^{2/3}$$

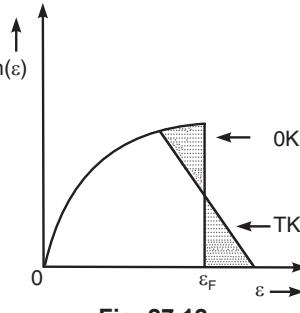


Fig. 27.12

**EXAMPLE 1.** Show that the mean energy of an electron in an electron gas at absolute zero temperature is  $\frac{3}{5} \varepsilon_F(0)$  where  $\varepsilon_F(0)$  is Fermi energy at absolute zero. (Rohilkhand, 1993)

**SOLUTION.** The number of electrons in an electron gas having energies between  $\varepsilon$  and  $\varepsilon + d\varepsilon$  is given by

$$n(\varepsilon) d\varepsilon = \frac{8\pi m V}{h^3} \sqrt{(2m)} \cdot \frac{\varepsilon^{1/2} d\varepsilon}{e^{(\varepsilon - \varepsilon_F)/kT} + 1} \quad \dots(1)$$

where  $m$  is the mass of electron and  $V$  is the volume of the electron gas.

If  $N$  is the total number of electrons, then the Fermi energy  $\varepsilon_F$  is given by

$$\begin{aligned} \varepsilon_F(0) &= \frac{h^2}{8m} \left( \frac{3N}{\pi V} \right)^{2/3} \\ \therefore \frac{V}{h^3} &= \frac{3N}{8\pi m \sqrt{(2m)}} [\varepsilon_F(0)]^{-3/2} \end{aligned}$$

Substituting this in Eq. (1), we get

$$n(\varepsilon) d\varepsilon = \frac{3}{2} N [\varepsilon_F(0)]^{-3/2} \frac{\varepsilon^{1/2} d\varepsilon}{e^{(\varepsilon - \varepsilon_F)/kT} + 1}$$

At  $T = 0$ , all of the electrons have energies less than or equal to  $\varepsilon_F(0)$ , (i.e.,  $\varepsilon \leq \varepsilon_F$ ) so that at  $T = 0$ , we have

$$e^{(\varepsilon - \varepsilon_F)/kT} = e^{-\infty} = 0.$$

$\therefore$  At absolute zero

$$n(\varepsilon) d\varepsilon = \frac{3N}{2} [\varepsilon_F(0)]^{-3/2} \varepsilon^{1/2} d\varepsilon \quad \dots(2)$$

The total energy at absolute zero

$$\begin{aligned} E_0 &= \int_0^{\varepsilon_F} \varepsilon n(\varepsilon) d\varepsilon \\ &= \frac{3N}{2} [\varepsilon_F(0)]^{-3/2} \int_0^{\varepsilon_F(0)} \varepsilon^{3/2} d\varepsilon = \frac{3}{5} N \varepsilon_F(0) \end{aligned}$$

$\therefore$  Average energy  $\bar{\varepsilon}_0$  at absolute zero is

$$\bar{\varepsilon}_0 = \frac{E_0}{N} = \frac{3}{5} \varepsilon_F(0).$$

**EXAMPLE 2.** Find the Fermi energy in copper on the assumption that each copper atom contributes one free electron to the electron gas. The density of copper is  $8.94 \times 10^3 \text{ kg/m}^3$  and its atomic mass is 63.5 u. (B.U. 2011)

**SOLUTION.** Mass of one copper atom

$$\begin{aligned} &= (63.5 \text{ u}) (1.66 \times 10^{-27} \text{ kg/u}) \\ &= 1.054 \times 10^{-25} \text{ kg.} \end{aligned}$$

$$\begin{aligned} \frac{N}{V} &= \frac{8.94 \times 10^3 \text{ kg/m}^3}{1.054 \times 10^{-25} \text{ kg}} = 8.48 \times 10^{28} \text{ atoms/m}^3. \\ &= 8.48 \times 10^{28} \text{ electrons/m}^3. \end{aligned}$$

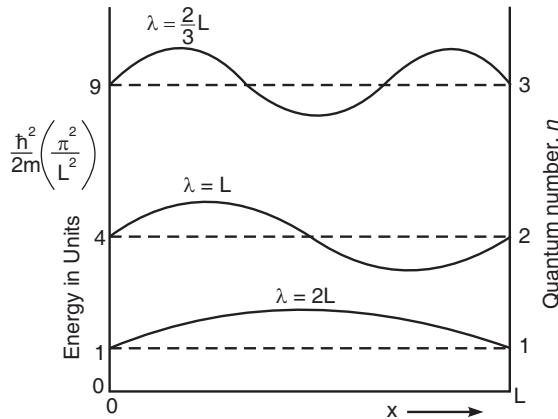
The Fermi energy in copper is

$$\begin{aligned} \varepsilon_F &= (3\pi^2)^{2/3} \frac{\hbar^2}{2m} \left( \frac{N}{V} \right)^{2/3} = (3\pi^2)^{2/3} \frac{(1.05 \times 10^{-34} \text{ Js})^2}{2(9.11 \times 10^{-31} \text{ kg})} (8.48 \times 10^{28} / \text{m}^3)^{2/3} \\ &= 1.13 \times 10^{-18} \text{ J} = 7.04 \text{ eV.} \end{aligned}$$

### FREE ELECTRON FERMI GAS

#### 27.8 ENERGY LEVELS IN ONE DIMENSION

By a free electron Fermi gas, we mean a gas of free electrons subject to the Pauli principle. Consider a free electron gas in one dimension. An electron of mass  $m$  is confined to a length  $L$  by infinite barriers (Fig. 27.13).



**Fig. 27.13**

The wave function  $\psi_n(x)$  of the electron is a solution of the Schrodinger equation

$$H\psi = \epsilon\psi \quad \dots(1)$$

Hamiltonian  $H =$  Kinetic energy + potential energy.

The potential energy is neglected. So we have

$$H = p^2/2m \quad \dots(2)$$

Here,  $p$  is the momentum. In quantum theory  $p$  is represented by  $-i\hbar \frac{d}{dx}$ .

$$\therefore H\psi_n = -\frac{\hbar^2}{2m} \frac{d^2\psi_n}{dx^2} = \epsilon_n \psi_n, \quad \dots(3)$$

Here,  $\epsilon_n$  is the energy of the electron in the orbital.

The boundary conditions are  $\psi_n(0) = 0$ ;  $\psi_n(L) = 0$ , as imposed by the infinite potential energy barriers.

They are satisfied if the wavefunction is sinelike with an integral number  $n$  of half-wavelengths between 0 and  $L$ .

$$\therefore \psi_n = A \sin\left(\frac{2\pi}{\lambda_n} x\right); \quad \frac{1}{2}n\lambda_n = L, \quad \dots(4)$$

or  $\psi_n = A \sin\left(\frac{n\pi x}{L}\right)$  where  $A$  is a constant.

Thus, only for integral values of  $n$  there are allowed wave functions  $\psi_n(x)$  and the corresponding energy values  $\epsilon_n$ .

The energy  $\varepsilon_n$  is given by

$$\varepsilon_n = \frac{\hbar^2}{2m} \left( \frac{n\pi}{L} \right)^2 \quad \dots(5)$$

Fig. 27.13 shows first three energy levels and wave functions of a free electron of mass  $m$  confined to a line of length  $L$ .

We now discuss the distribution of electrons among the allowed energy levels and determine the Fermi energy, total energy and the density of states.

### (i) Fermi Energy

We want to accommodate  $N$  electrons on the line. According to the *Pauli exclusion principle*, no two electrons can have all their quantum numbers identical. In a linear solid, the quantum numbers of a conduction electron orbital are  $n$  and  $m_s$ .  $n$  is any positive integer and  $m_s = \pm \frac{1}{2}$ ,

according to the spin orientation. Each energy level with quantum number  $n$  can accommodate two electrons, one with spin up and one with spin down.

Suppose there are six electrons. Then in the ground state of the system the filled orbitals are those given in the table.

$n$	$m_s$	electron occupancy	$n$	$m_s$	electron occupancy
1	$\uparrow$	1	3	$\uparrow$	1
1	$\downarrow$	1	3	$\downarrow$	1
2	$\uparrow$	1	4	$\uparrow$	0
2	$\downarrow$	1	4	$\downarrow$	0

More than one orbital may have the same energy. The number of orbitals with the same energy is called the *degeneracy*.

Let  $n_F$  denote the topmost filled energy level. Let  $N$  be the total number of electrons to be accommodated. Then for even  $N$ ,  $2n_F = N$  determines  $n_F$ .

The *Fermi energy*  $\varepsilon_F$  is defined as the energy of the topmost filled level in the ground state of the  $N$  electron system. For  $n = n_F$ , Eq. (5) gives

$$\varepsilon_F = \frac{\hbar^2}{2m} \left( \frac{n_F \pi}{L} \right)^2 = \frac{\hbar^2}{2m} \left( \frac{N\pi}{2L} \right)^2 \quad \dots(6)$$

### (ii) Total Energy

The total energy,  $\varepsilon_0$ , of all the  $N$  electrons in the ground state is determined by summing up the energies of the individual electrons. For  $N$  electrons, the number of filled energy levels is  $N/2$ .

$$\therefore \varepsilon_0 = 2 \sum_{n=1}^{N/2} \varepsilon_n = 2 \frac{\hbar^2}{2m} \left( \frac{\pi}{L} \right)^2 \sum_{n=1}^{N/2} n^2$$

Here the factor of 2 appears because each level contains two electrons with equal energies.

$$\begin{aligned} \text{Now, } \sum_{n=1}^{N/2} n^2 &= 1^2 + 2^2 + 3^2 + \dots + S^2 \quad \text{where } S = N/2. \\ &= \frac{1}{6} S (2S^2 + 3S + 1) \approx \frac{1}{3} S^3 \text{ for } S \gg 1 \end{aligned}$$

$$\begin{aligned} &\approx \frac{1}{3}(N/2)^3 \\ \varepsilon_0 &= 2 \frac{\hbar^2}{2m} \left(\frac{\pi}{L}\right)^2 \frac{1}{3} \left(\frac{N}{2}\right)^3 = \frac{1}{3} \frac{\hbar^2}{2m} \left(\frac{N\pi}{2L}\right)^2 N = \frac{1}{3} N \epsilon_F \quad \dots(7) \end{aligned}$$

### (iii) Density of States

The density of states,  $D(\varepsilon)$ , is defined as the number of electronic states present in a unit energy range.

$$D(\varepsilon) = \frac{dn}{d\varepsilon} \quad \dots(8)$$

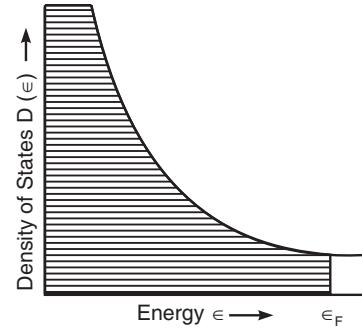
For a free electron gas, since each energy level contains two electronic states, one with spin up and the other with spin down, the actual density of states is twice the value given by Eq. (8).

$$D(\varepsilon) = 2 \frac{dn}{d\varepsilon}$$

Differentiating Eq. (5), we get

$$\begin{aligned} d\varepsilon &= \frac{\hbar^2}{2m} \left(\frac{\pi}{L}\right)^2 2n dn. \\ D(\varepsilon) &= \frac{2dn}{d\varepsilon} = 2 \cdot \frac{2m}{\hbar^2} \left(\frac{L}{\pi}\right)^2 \frac{1}{2n} = \frac{4L}{\hbar} \left(\frac{m}{2\varepsilon_n}\right)^{1/2} \end{aligned}$$

The plot of  $D(\varepsilon)$  versus  $\varepsilon$  is shown in Fig. 27.14.



**Fig. 27.14**

## 27.9 FREE ELECTRON GAS IN THREE DIMENSIONS

The free-particle Schrodinger equation in three dimensions is

$$-\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi_k(\mathbf{r}) = \epsilon_k \psi_k(\mathbf{r}) \quad \dots(1)$$

Suppose the electrons are confined to a cube of edge  $L$ .

The wavefunction is the standing wave

$$\psi_n(\mathbf{r}) = A \sin(\pi n_x x / L) \sin(\pi n_y y / L) \sin(\pi n_z z / L), \quad \dots(2)$$

Here,  $n_x, n_y, n_z$  are positive integers. The origin is at one corner of the cube.

We now require the wavefunctions to be periodic in  $x, y, z$  with period  $L$ .

Thus

$$\begin{aligned} \psi(x+L, y, z) &= \psi(x, y, z), \\ \psi(x, y+L, z) &= \psi(x, y, z), \\ \psi(x, y, z+L) &= \psi(x, y, z). \end{aligned} \quad \dots(3)$$

Wavefunctions satisfying the free-particle Schrodinger equation and the periodicity condition are of the form of a travelling plane wave.

$$\psi_k(\mathbf{r}) = \exp(i \mathbf{k} \cdot \mathbf{r}) = \exp(i(k_x x + k_y y + k_z z)) \quad \dots(4)$$

The components of the wavevector  $\mathbf{k}$  ( $k_x, k_y, k_z$ ) must satisfy

$$k_x = 0, \pm \frac{2\pi}{L}, \pm \frac{4\pi}{L}, \dots \quad \dots(5)$$

Similar results are obtained for  $k_y$  and  $k_z$ .

Any component of  $\mathbf{k}$  is of the form  $2n\pi / L$ , where  $n$  is a positive or negative integer.

The quantum number  $m_s$  represents the spin direction. Thus the state of an electron is specified completely by a set of four quantum numbers  $k_x$ ,  $k_y$ ,  $k_z$  and  $m_s$ .

On substituting Eq. (4) in Eq. (1), we have the energy  $\varepsilon_k$  of the orbital with wave vector  $\mathbf{k}$ .

$$\varepsilon_k = \frac{\hbar^2}{2m} k^2 = \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2). \quad \dots(6)$$

The magnitude of the wavevector  $\mathbf{k}$  is related to the wavelength  $\lambda$  by

$$k = \frac{2\pi}{\lambda} \quad \dots(7)$$

Thus the energy spectrum consists of discrete energy levels.

### (a) The Fermi Energy

In quantum mechanics, the linear momentum  $\mathbf{p}$  is represented by the operator  $\mathbf{p} = -i\hbar\nabla$ . When operated on Eq. (4),

$$\mathbf{p} \psi_k(\mathbf{r}) = -i\hbar\nabla \psi_k(\mathbf{r}) = \hbar\mathbf{k} \psi_k(\mathbf{r}). \quad \dots(8)$$

This indicates that the plane wave  $\psi_k$  is an eigenfunction of the linear momentum with the eigenvalue  $\hbar\mathbf{k}$ . The particle velocity in the orbital  $\mathbf{k}$  is given by  $\mathbf{v} = \hbar\mathbf{k} / m$ .

In the ground state of a system of  $N$  free electrons, the occupied orbitals may be represented as points inside a sphere in  $\mathbf{k}$  space. The energy at the surface of the sphere is the *Fermi energy*.

Let  $\mathbf{k}_F$  be the wave vector from the origin of the  $\mathbf{k}$ -space to the surface of the sphere (Fig. 27.15). Then, using Eq. (6), the Fermi energy is written as

$$\varepsilon_F = \frac{\hbar^2}{2m} k_F^2 \quad \dots(9)$$

From Eq. (5) we see that there is one allowed wavevector or one distinct triplet of quantum numbers  $k_x$ ,  $k_y$ ,  $k_z$  which corresponds to the volume element  $(2\pi/L)^3$  of  $\mathbf{k}$  space. Thus, in the sphere called the *Fermi sphere* of volume  $(4\pi/3)k_F^3$ , the total number of orbitals is

$$2 \cdot \frac{4\pi k_F^3 / 3}{(2\pi/L)^3} = \frac{V}{3\pi^2} k_F^3 = N. \quad \dots(10)$$

Here, the factor 2 appears because there are two allowed values of  $m_s$  for each allowed value of  $\mathbf{k}$ .

$$\text{From Eq. (10),} \quad k_F = \left( \frac{3\pi^2 N}{V} \right)^{1/3}. \quad \dots(11)$$

This shows that the value of  $k_F$  depends on the concentration of electrons,  $N/V$ , and is independent of the mass of the electron.

$$\text{Using Eq. (9),} \quad \varepsilon_F = \frac{\hbar^2}{2m} \left( \frac{3\pi^2 N}{V} \right)^{2/3} \quad \dots(12)$$

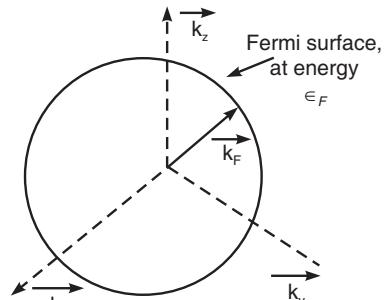


Fig. 27.15

The Fermi energy depends on both the electronic concentration and mass.

The electron velocity,  $v_F$ , at the Fermi surface is

$$v_F = \frac{\hbar k_F}{m} = \frac{\hbar}{m} \left( \frac{3\pi^2 N}{V} \right)^{1/3} \quad \dots(13)$$

### (b) Density of States

The number of orbitals per unit energy range,  $D(\epsilon)$ , is called the *density of states*. We use Eq. (12) for the total number of orbitals of energy  $\leq \epsilon$ .

$$N = \frac{V}{3\pi^2} \left( \frac{2m\epsilon}{\hbar^2} \right)^{3/2} \quad \dots(14)$$

The density of states is

$$D(\epsilon) = \frac{dN}{d\epsilon} = \frac{V}{2\pi^2} \cdot \left( \frac{2m}{\hbar^2} \right)^{3/2} \cdot \epsilon^{1/2} \quad \dots(15)$$

## 27.10 SCHOTTKY EFFECT

**Thermionic emission.** The phenomenon of emission of electrons from the surface of a metal when thermal energy is supplied to it is called thermionic emission.

**Work function.** The minimum amount of work that has to be done to eject an electron from a material without imparting it any velocity is called the **work function of the material**. It is denoted by  $\phi$ .

**Richardson-Dushman equation.** The number of electrons emitted by heated surface depends upon the temperature of the surface and work function. It is given by the following relation:

$$j = AT^2 \exp \left( \frac{-\phi}{k_B T} \right).$$

$k_B$  = Boltzman constant

$T$  = Temperature in Kelvin

$A$  = Emission constant

$\phi$  = Work function of the metal in eV

$j$  = emission current density.

By the action of electric field alone (if the field is strong enough), it is possible to remove electrons from a metal surface. We know that an accelerating electric field is usually applied to the emitter. This field removes the emitted electrons from the vicinity of the surface. Moreover, it accelerates the electrons to the collector electrode where they constitute the thermionic current. Besides this, the field also lowers the height of potential energy barrier at the surface and hence increases the emission. This increase of emission and its dependence on the external electric field is called *Schottky effect*.

**Explanation of Schottky effect.** When a positive potential is applied to the anode with respect to the cathode, it produces an electric field at the cathode and tends to pull the electrons out of the cathode (Fig. 27.16). This externally applied field reduces the height of the potential barrier at the surface of the cathode, thereby decreasing the work function.

In Fig. 27.17, curve 1 represents the potential energy barrier of the metal surface in the absence of an external field.

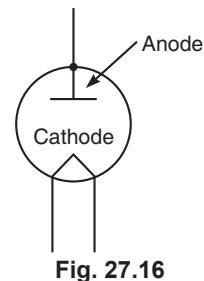


Fig. 27.16

- Curve 2 represents the potential energy barrier in the presence of an external field.

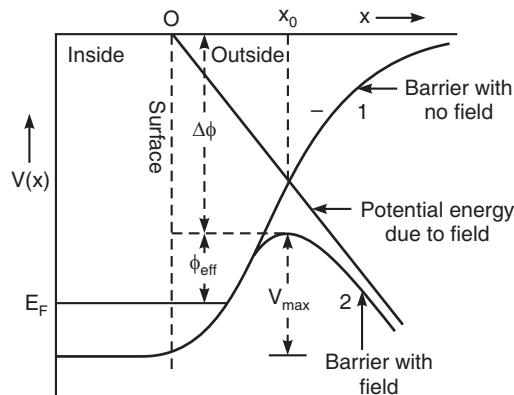


Fig. 27.17

- The effective work function reduces from  $\phi$  to  $\phi_{\text{eff}}$  i.e.,

$$\phi_{\text{eff}} = \phi - e \left[ \frac{eE_x}{4\pi\epsilon_0} \right]^{1/2}.$$

Here,  $E_x$  = external electric field between the emitting metal surface and the collecting electrode. So the new formula for thermionic emission becomes

$$j = AT^2 \exp \left[ \frac{-\left\{ \phi - e \left( \frac{eE_x}{4\pi\epsilon_0} \right)^{1/2} \right\}}{k_B T} \right]$$

- This observed increase in the emission current from a heated cathode with increase in anode voltage is called *schottky effect*.
- The greater the anode potential, the greater is the reduction in work function and higher is the electron emission. In fact, here thermionic emission is aided by field emission. The emission current, therefore, rises slowly with the increase in anode potential.

## 27.11 PARAMAGNETIC SUSCEPTIBILITY OF CONDUCTION ELECTRONS

The *magnetic susceptibility*  $\chi$  of a specimen is defined as magnetic moment per unit magnetising field i.e.,

$$\chi = \frac{M}{H} \quad \dots(1)$$

Here,  $M$  is magnetic moment and  $H$  is magnetising field.

Classical free electron theory gives an unsatisfactory account of the paramagnetic susceptibility of the conduction electrons. Classically, the *paramagnetic susceptibility* is given by

$$\chi_p = \frac{n\mu_B^2}{k_B T} m^{-3} \quad \dots(1)$$

Here,  $n$  = number of conduction electrons per unit volume,

$\mu_B$  = Bohr magneton,  $k_B$  = Boltzmann constant and  $T$  = temperature.

This shows that  $\chi_p \propto (1/T)$ .

It is observed that the paramagnetic susceptibility of most normal nonferromagnetic metals is independent of temperature.

Pauli used Fermi-Dirac statistics to explain this disagreement between theory and experiment.

We first give a qualitative explanation of the situation.

(i) In the absence of any external field ( $H = 0$ ) and at absolute zero ( $T = 0$ ), all energy levels below Fermi level  $E_F(0)$  are occupied and all those above  $E_F(0)$  are empty [Fig. 27.18(a)]. The electrons are divided into two groups with spin up and spin down.

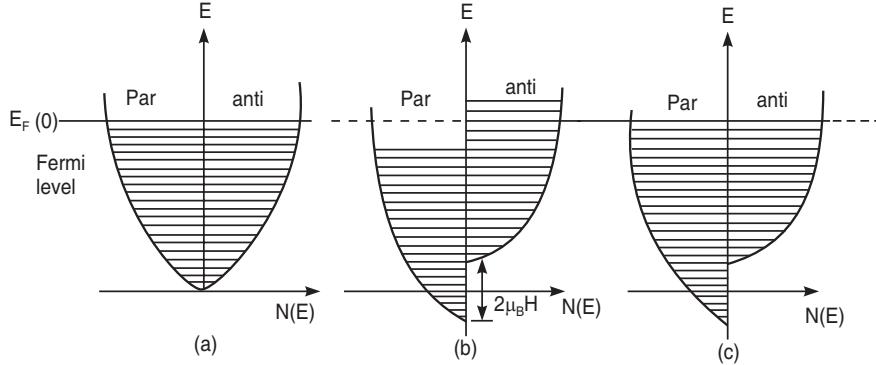


Fig. 27.18

(ii) Now, when a magnetic field is applied, the electrons with their spins parallel to the field have their energy lowered by an amount  $\mu_B H$  and the electrons with their spins antiparallel raised by the same amount. Fig. 27.18(b) shows instantaneous shift of states when field is applied. It may be noted that  $\mu_B H \ll E_{F0}$ . This situation is unstable because for equilibrium the two halves must be filled upto the same level.

(iii) Equilibrium is established when a certain number of electrons with antiparallel spins has entered the group of parallel ones till both the halves are filled upto the same level [Fig. 27.18(c)].

The number of such electrons is very small at 0K, and is of the order of  $n\left(\frac{T}{T_F}\right)$  per unit volume at temperatures such that  $k_B T \ll E_F, T_F$  being the Fermi temperature. It is clear that these electrons would lie near the Fermi level. Now, each of these electrons behaves, qualitatively, as if it were to contribute the susceptibility of  $\mu_B^2/k_B T$ .

The susceptibility of the metal is

$$\chi_p \approx \frac{n\mu_B^2}{k_B T_F} \quad \dots(2)$$

This is independent of temperature and is of the observed order of magnitude.

Thus quantum theory is capable of explaining the paramagnetism of metals. We now calculate  $\chi_p$  quantitatively.

We suppose that the concentration of electrons with their spins, and hence magnetic moments, parallel to the magnetic field is  $n_p$  and that of antiparallel ones is  $n_a$ . Then expression for  $n_p$  may be written as

$$n_p = \int_{-\mu_B H}^{E_F} f(E) \frac{1}{2} D_P(E + \mu_B H) dE. \quad \dots(3)$$

Here,  $\frac{1}{2}D(E+\mu_B H)$  is the density of states of parallel spin orientation with allowance for the downward shift of energy by  $-\mu_B H$ .

$f(E) = 1$  throughout the interval of integration at  $T = 0$  and

$$\begin{aligned} D_p(E + \mu_B H) &= \frac{1}{2\pi^2} \left( \frac{2m}{\hbar^2} \right)^{3/2} (E + \mu_B H)^{1/2} / \text{unit volume} \\ \therefore n_p &= \frac{1}{2} \left( \frac{1}{2\pi^2} \right) \left( \frac{2m}{\hbar^2} \right)^{3/2} \int_{-\mu_B H}^{E_F} (E + \mu_B H)^{1/2} dE \end{aligned} \quad \dots(4)$$

Similarly,

$$n_a = \frac{1}{2} \left( \frac{1}{2\pi^2} \right) \left( \frac{2m}{\hbar^2} \right)^{3/2} \int_{\mu_B H}^{E_F} (E - \mu_B H)^{1/2} dE \quad \dots(5)$$

The magnetisation  $M$  is given by

$$\begin{aligned} M &= \mu_B(n_p - n_a) \\ &= \mu_B \left( \frac{1}{2} \right) \left( \frac{1}{2\pi^2} \right) \left( \frac{2m}{\hbar^2} \right)^{3/2} \left[ \int_{-\mu_B H}^{E_F} (E + \mu_B H)^{1/2} dE - \int_{\mu_B H}^{E_F} (E - \mu_B H)^{1/2} dE \right]. \end{aligned}$$

Since  $\mu_B H \ll E_F$  and there are very few states near  $E = 0$ , we can set the lower limit in both the integrals equal to zero. Hence

$$M = \left( \frac{\mu_B}{2} \right) \left( \frac{1}{2\pi^2} \right) \left( \frac{2m}{\hbar^2} \right)^{3/2} \left[ \int_0^{E_F} (E + \mu_B H)^{1/2} dE - \int_0^{E_F} (E - \mu_B H)^{1/2} dE \right] \quad \dots(6)$$

Now,  $(E + \mu_B H)^{1/2} = \sqrt{E} \left( 1 + \frac{\mu_B H}{E} \right)^{1/2} = \sqrt{E} + \frac{1}{2} \frac{\mu_B H}{\sqrt{E}}$

and  $(E - \mu_B H)^{1/2} = \sqrt{E} \left( 1 - \frac{\mu_B H}{E} \right)^{1/2} = \sqrt{E} - \frac{1}{2} \frac{\mu_B H}{\sqrt{E}}$

$$(E + \mu_B H)^{1/2} - (E - \mu_B H)^{1/2} = \frac{\mu_B H}{\sqrt{E}}.$$

Eq.(6) becomes,

$$M = \left( \frac{\mu_B}{2} \right) \left( \frac{1}{2\pi^2} \right) \left( \frac{2m}{\hbar^2} \right)^{3/2} \mu_B H \int_0^{E_F} \frac{dE}{\sqrt{E}} = \left( \frac{\mu_B^2}{2\pi^2} \right) \left( \frac{2m}{\hbar^2} \right)^{3/2} E_F^{1/2} \cdot H$$

Paramagnetic susceptibility,

$$\chi_p = \frac{M}{H} = \left( \frac{1}{2\pi^2} \right) \left( \frac{2m}{\hbar^2} \right)^{3/2} \mu_B^2 E_F^{1/2}$$

Substituting the value of  $E_F$ , we finally have

$$\chi_p = \frac{3\pi^2 n \mu_B^2}{2\pi^2 E_F} = \frac{3}{2} \cdot \frac{n \mu_B^2}{k_B T_F} \quad \dots(7)$$

This is the Pauli paramagnetism formula. This formula is independent of temperature in agreement with experiment. This formula is satisfactory, so long as  $k_B T \ll k_B T_F$ .

### HALL EFFECT

#### 27.12 HALL EFFECT

When a magnetic field is applied perpendicular to a current carrying conductor, a voltage is developed across the specimen in a direction perpendicular to both the current and the magnetic field. This phenomenon is called the '*Hall effect*'. The voltage so developed is called the '*Hall voltage*'.

##### Hall voltage and Hall coefficient

Consider a slab of material subjected to an external electric field  $E_x$  along  $x$ -axis and a magnetic field  $H_z$  along  $z$ -axis (Fig. 27.19). As a result of the electric field, a current density  $J_x$  will flow in the direction of  $E_x$ . Let the current be carried by an electron of charge  $-e$ . The external magnetic field  $H_z$  applied perpendicular to current will exert a transverse magnetic deflecting force on the electron which causes the electron to drift downward to the lower edge of the specimen. Consequently, lower surface

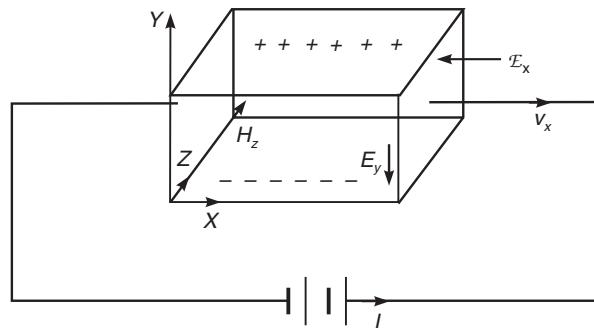


Fig. 27.19

of the slab collects a negative charge and upper surface a positive charge. This excess of negative charge on the lower edge and the corresponding excess of positive charge at the upper edge creates a transverse electric field  $E_y$  known as '*Hall electric field*' which opposes the transverse drifting of electrons. Ultimately an equilibrium is reached in which the force due to accumulation of electrons becomes equal to the magnetic force and so the flow of electrons stops. Thus the net force on electrons becomes zero, i.e.,  $F_y = 0$ .

The electric force on the electron having a charge  $-e = -eE_y$

The force due to the magnetic field  $= ev_x H_z$ .

In the steady state

$$F_y = 0.$$

$$F_y = -eE_y + e(v_x H_z) = 0$$

or

$$E_y = v_x H_z \quad \dots(1)$$

$E_y$  is called the *Hall voltage* or *Hall field*.

Let  $n$  be the free electron density.

The current density is given by

$$J_x = -nev_x \quad \dots(2)$$

$$v_x = -\frac{J_x}{ne}$$

$$\therefore E_y = v_x H_z = -\frac{J_x H_z}{ne} = R_H J_x H_z \quad \dots(3)$$

Here,  $R_H = -\frac{1}{ne}$  is called the '*Hall coefficient*' for the substance of the bar.

The negative value of  $R_H$  indicates that current carriers are electrons. If the charge carriers are holes,  $R_H$  is positive.

### Relation between $R_H$ and $\mu$

The mobility ' $\mu$ ' is defined as the velocity acquired by the current carrying particle per unit electric field, *i.e.*,

$$\mu = \frac{v_x}{E_x} \quad \therefore v_x = \mu E_x$$

Eq. (3) becomes  $E_y = \mu E_x H_z$  ... (4)  
Comparing Eqs. (3) and (4),

$$R_H J_x H_z = \mu E_x H_z$$

$$R_H = \mu \frac{E_x}{J_x}$$

$$\therefore R_H = \mu \cdot \frac{1}{\sigma}$$

Here,  $\sigma$  is the electrical conductivity of the metal.

$$\therefore \mu = R_H \sigma.$$

### Relation between Mobility and Hall angle

$$\mu = R_H \sigma$$

From Eq. (3),

$$R_H = \frac{E_y}{J_x H_z}$$

$$\mu = \frac{E_y \sigma}{J_x \cdot H_z}$$

Now,

$$J_x = \sigma E_x$$

$$\mu = \frac{E_y \sigma}{\sigma E_x H_z} = \phi \frac{1}{H_z}$$

Here,

$$\phi = \frac{E_y}{E_x} \text{ is called Hall angle.}$$

$$\therefore \boxed{\phi = \mu H_z}$$

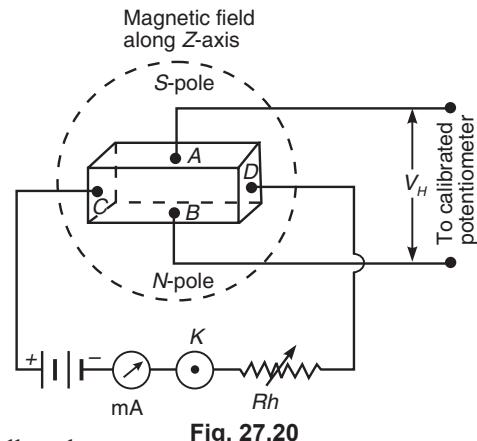
### Experimental Determination of Hall Coefficient

A thin metallic strip, several mm wide and several cm long, is placed in  $x$ -direction and a magnetic field  $H_z$  is applied in  $z$ -direction (Fig. 27.20).

A suitable current is passed in the specimen which can be adjusted by the rheostat. Two potential leads are placed between the points  $A$  and  $B$  which are connected to a sensitive calibrated potentiometer to measure the developed Hall voltage  $V_H$ . By measuring the Hall voltage, Hall-coefficient can be calculated as given below :

The value of Hall electric field is given by

$$E_y = -\frac{J_x H_z}{ne}$$



**Fig. 27.20**

Let  $I$  be the current flowing through the specimen.

Let  $b$  and  $d$  be respectively the breadth and thickness of the specimen.

The current density is given by

$$J_x = \frac{I}{b \times d}$$

$$E_y = -\frac{I}{nebd} \cdot H_z$$

or  $E_y \cdot d = V_H = -\frac{I}{neb} \cdot H_z = R_H \cdot \frac{IH_z}{b}$

Hall coefficient,  $R_H = \frac{V_H \cdot b}{IH_z}$ .

Thus, the value of Hall coefficient ( $R_H$ ) can be calculated by knowing  $I$  and  $H_z$  and by measuring experimentally the Hall voltage  $V_H$  developed across the points  $A$  and  $B$  by means of a calibrated potentiometer.

#### Importance of Hall effect

The Hall effect measurements provide the following information about the solid.

1. The sign (electrons or holes) of charge carrier is determined.
2. From the magnitude of Hall coefficient, the number of charge carriers per unit volume can be calculated.
3. The mobility of charge carriers is measured directly.
4. We can decide whether a material is a metal, semiconductor or insulator.
5. From the knowledge of measured Hall voltage, the unknown magnetic field can be measured provided the value of Hall constant for the slab is known.

### BAND THEORY OF SOLIDS

#### 27.13 BAND THEORY OF SOLIDS

A solid contains an enormous number of atoms packed closely together. Each atom, when isolated, has a discrete set of electron energy levels,  $1s$ ,  $2s$ ,  $2p$ , ..... If we imagine all the  $N$  atoms of the solid to be isolated from one another, they would have completely coinciding schemes of their energy levels. The energies of electrons within any one isolated atom obey the following conditions.

- (i) There are specific electronic energy levels around each atom [(Fig. 27.21 (a)]. Electrons cannot occupy spaces between these levels.

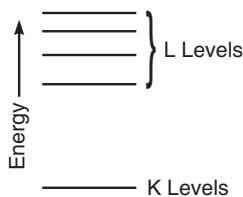


Fig. 27.21 (a)

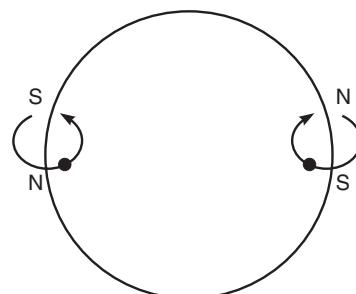


Fig. 27.21 (b)

(ii) Electrons fill the lowest energy levels first. A specific quantity of energy, called a *quantum of energy*, must be supplied to move an electron to the next higher level.

(iii) According to the Pauli exclusion principle, no two electrons can occupy the same quantum state. Not more than two electrons may occupy any one energy level. Two electrons may occupy the same energy level because they have opposite magnetic characteristics i.e., *electron spins* are in opposite directions [(Fig. 27.21 (b)].

Now, let us look into what happens to the energy levels of isolated atoms, as they are brought closer and closer together, to form a solid. If the atoms are brought in close proximity, the valence electrons of adjacent atoms interact. Hence the valence electrons constitute a single system of electrons common to the entire crystal, their outermost electronic orbits overlapping. Therefore, the  $N$  electrons will now have to occupy different energy levels. This is brought about by the electric forces exerted on each electron by all the  $N$  nuclei. As a result of these forces, each atomic energy level is split up into a large number of closely spaced energy levels. A set of such closely spaced energy levels is called an *energy band*.

As an example, the 11 electrons in a neutral sodium atom, each occupy a specific energy level as indicated in Fig. 27.22. The energy levels of sodium become bands when the atoms are close together. In the figure,  $r_0$  represents the spacing between atoms in solid sodium. When the atoms are part of a solid, they interact with each other and the electrons have slightly different energies.

In an energy band, the allowed energies are almost continuous. These energy bands are separated by regions, which have no allowed energy levels. These regions are known as “*forbidden bands*” or “*energy gaps*” [Fig. 27.23].

The amount of splitting is not the same for different levels. The levels filled by the valence electrons in an atom are disturbed to a greater extent, while those filled by inner electrons are disturbed only slightly. Fig. 27.22 shows the splitting of the levels as a function of the distance  $r$  between the atoms.

If there are  $N$  atoms in a solid, there will be  $N$  allowed quantum states in each band. Each quantum state can be occupied by a maximum of two electrons with opposite spins. Thus each energy band can be occupied by  $2N$  electrons.

The *valence band* consists of a group of states containing the outermost or the *valence* electrons of an atom. The band formed from the atomic energy levels containing valence electrons is called ‘*valence band*’. These electrons have the highest energy. The band is, obviously, the highest occupied band. Above the valence band, there is the band of next higher permitted energies called the *conduction band*. Conduction band is separated by a *gap* from the valence band (Fig. 27.24). The interval between them represents the energies which their electrons cannot possess. The conduction band corresponds to the first excited states and is normally the lowest *unfilled* energy band. In this band, the electrons can move freely and the electrons are generally called “*conduction electrons*”.

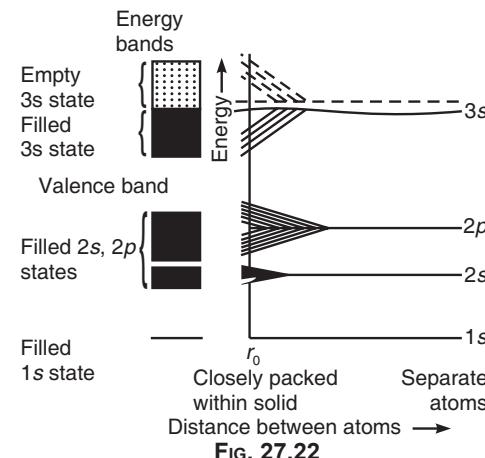


FIG. 27.22

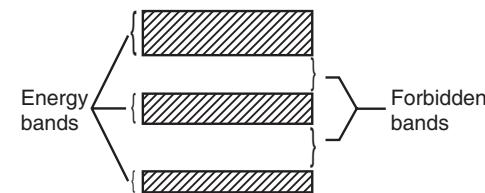


FIG. 27.23

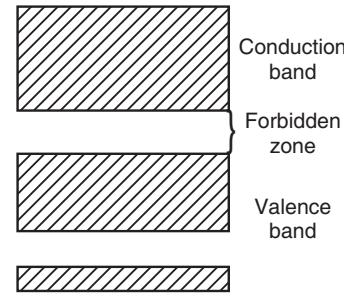


FIG. 27.24

## 27.14 CLASSIFICATION OF SOLIDS ON THE BASIS OF BAND THEORY

Whether a solid is an insulator, conductor or semiconductor, is determined by the energy band structure, the arrangement of electrons with respect to these bands and the width of forbidden bands.

**(i) Conductors.** Fig. 27.25 is a simplified diagram of the energy levels of a sodium atom ( $1s^2 2s^2 2p^6 3s^1$ ) and the energy bands of solid sodium. A sodium atom has a single  $3s$  electron in its valence shell. Thus if there are  $N$  atoms in a solid piece of sodium, its  $3s$  valence band will contain  $N$  energy levels. The  $N$  valence electrons will fill in pairs only lower half of the energy levels of the band. Now suppose an electric field is applied across a piece of solid sodium. Then electrons in the valence band easily acquire additional energy to move to the higher unoccupied energy levels *within the same band* without crossing any energy gap. The additional energy is in the form of kinetic energy, and the moving electrons constitute an electric current. Sodium metal is, therefore, a good conductor of electricity. An unoccupied band into which electrons can pass is termed as ‘conduction band’. In metals, the valence band itself is also the conduction band.

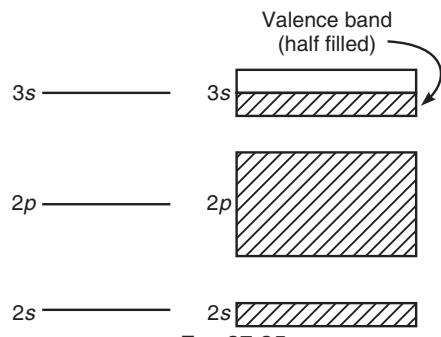


FIG. 27.25

**(ii) Insulators.** Fig. 27.26 is a diagram of the energy bands of diamond. The valence band is completely filled and the conduction band is empty at the absolute zero of temperature. There exists a large energy gap  $E_g$  of 6 eV between the valence and conduction bands. This means that at least 6 eV of additional energy must be provided to an electron in a diamond crystal if it is to have any kinetic energy, since it cannot have an energy lying in the forbidden band. If an electric field is applied across diamond, the electrons in the valence band would not move within the band because there are no unoccupied levels in this band. They can, however, move to the higher empty band provided they get the required energy  $E_g$  to cross the forbidden band. An electric field cannot give this amount of energy to an electron in the solid. Thus no electron acquires additional kinetic energy in the electric field. Therefore, there will be no electric current and the solid will be an insulator.

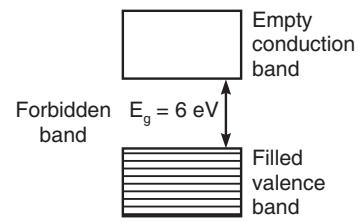


FIG. 27.26

**(iii) Semiconductors.** Silicon has a crystal structure resembling that of diamond. As in diamond, an energy gap separates the top of the filled valence energy band from a vacant higher band (conduction band). But the width of the forbidden band in silicon is only 1.1 eV (Fig. 27.27). At low temperatures, conductivity of silicon will be very low. Since the width of the energy gap is small, at ordinary room temperatures, a small proportion of electrons in the valence band have sufficient kinetic energy of thermal motion to cross the narrow energy gap and enter the conduction band above it. These electrons are sufficient to permit a limited amount of current to flow when an electric field is applied. Thus silicon has an electrical resistivity lying between that of conductors and insulators. Hence it is termed a semiconductor. Thus solids such as silicon and germanium are conductors at higher temperatures even though they are insulators at very low temperatures.

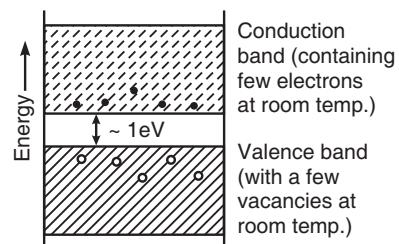


FIG. 27.27

In a metallic conductor, increase in temperature impedes the flow of electrons due to the resulting greater thermal agitation of the lattice ions of the solid and hence the conductivity decreases. In

semiconductors, this effect is masked by the ability of more and more electrons to overcome the energy gap as a result of their increased thermal energies at higher temperatures. The conductivity of a semiconductor, therefore, increases with temperature.

## 27.15 OPTICAL PROPERTIES OF SOLIDS

The optical properties of solids and their energy band structures are closely connected. We can explain the following optical properties of solids on the basis of energy bands.

**(i) All metals are opaque to light of all wavelengths.** In a metal, the valence energy band is only partly filled. Therefore, the valence electrons can acquire additional energy, however small, to move to the unoccupied energy levels of the same band. Photons of visible light have energies between about 1 and 3 eV. Therefore, when light of any wavelength falls on a metal, the electrons absorb the light photons (energy) and are excited to the higher unoccupied energy levels. Hence metals are opaque to visible light.

**(ii) Insulators are transparent to visible light.** The energy gap in insulators is large. The electrons in an insulator need more than 3 eV of energy to jump across the forbidden band to the next allowed band. Insulators, therefore, cannot absorb photons of visible light. Hence insulators are transparent.

**(iii) Semiconductors are transparent to infrared light but opaque to visible light.** The energy gap in semiconductors are about the same in width as the energies of photons of visible light. The photons of visible light can excite the valence electrons to the upper empty conduction band. Hence visible light is absorbed in the semiconductor. Infrared light has lower frequencies and hence lower photon energies. So photons of infrared light cannot excite the valence electrons to the upper empty conduction band. Hence infrared light is not absorbed in the semiconductor.

### Optical constants

- Some solids are transparent and some opaque to the light. Some solid surfaces are highly reflecting. But some solid surfaces bend the electromagnetic waves incident on them. These properties are investigated by measuring certain physical quantities that have a close relationship with the dielectric constant,  $\epsilon(\omega)$ , of the solid. This dielectric constant represents the behaviour of all electrons, those in filled and those in unfilled bands.

The optical response of the solid material to electromagnetic waves is characterized by the *complex refractive index*

$$\bar{n}(\omega) = n + ik = \sqrt{\epsilon(\omega)} \quad \dots (1)$$

Here,  $\bar{n}(\omega)$  is the complex index of refraction. It is given by the ratio of the velocity of light in vacuum to the velocity of light in the given medium as

$$\bar{n}(\omega) = \frac{c}{v} \quad \dots (2)$$

Complex dielectric constant is

$$\epsilon(\omega) = \epsilon'(\omega) + i\epsilon''(\omega) \quad \dots (3)$$

Using Eq. (1), we get

$$\epsilon'(\omega) = n^2 - k^2 \quad \dots (4)$$

and

$$\epsilon''(\omega) = 2nk$$

Here  $k$  is called the extinction coefficient.

We generally quote the two optical constants  $n$  and  $k$ .

## 27.16 ENERGY BANDS: ALTERNATIVE ANALYSIS

We shall discuss qualitatively the motion of electrons in a periodic lattice and explain how it leads to the origin of the energy bands and the forbidden bands in solids.

In general, an electron in a solid moves in a region of *periodically-varying potential* (with the periodicity of the lattice) caused by the ion-cores situated at the lattice points, plus the average energy of all the other free electrons. Fig. 27.28 shows the potential energy ( $U$ ) of an electron in a periodic array of positive ions. When the de Broglie wavelength of the electron ( $= h/p$ ) corresponds to a periodicity in the spacing of the ions, the electron interacts strongly with the lattice and undergoes Bragg "reflection". This limits the electron to certain ranges of momenta and, correspondingly, to certain ranges of energy (energy bands).

The de Broglie wavelength of a free electron of momentum  $p$  is

$$\lambda = \frac{h}{p} \quad \dots(1)$$

Let  $a$  be the spacing between the ions of the lattice. Unbound low-energy electrons can travel freely through a crystal since their wave lengths are long relative to the lattice spacing  $a$ . More energetic electrons have wavelengths comparable with  $a$ . Such electrons are diffracted. The Bragg's condition for the reflection of waves is

$$2a \sin \theta = n \lambda, \quad n = 1, 2, 3, \dots \dots \quad \dots(2)$$

Here, the electron approaches the atomic plane at an angle  $\theta$ .

The wave number  $k$  is

$$k = \frac{2\pi}{\lambda} = \frac{p}{h} \quad \dots(3)$$

Bragg's formula in terms of  $k$  is

$$k = \frac{n\pi}{a \sin \theta}, \quad n = 1, 2, 3, \dots \dots \quad \dots(4)$$

Fig. 27.29 shows Bragg reflection in a two-dimensional square lattice. Bragg reflection from the vertical rows of ions occurs when  $k_x = n \pi/a$ . Similarly, reflection from the horizontal rows occurs when  $k_y = n \pi/a$ .

### Brillouin Zones

The region in  $k$ -space (here an imaginary plane whose rectangular coordinates are  $k_x$  and  $k_y$ ) that low- $k$  electrons can occupy without being diffracted is called the *first Brillouin Zone*. The second Brillouin zone contains electrons with  $k > \pi/a$  that do not fit into the first zone, yet which have sufficiently small wave numbers to avoid diffraction by the diagonal sets of atomic planes in Fig. 27.29. The second zone contains electrons with  $k$  values from  $\pi/a$  to  $2\pi/a$  for electrons moving in the  $\pm x$  and  $\pm y$  directions. The first and second Brillouin zones of a two-dimensional square lattice are shown in Fig. 27.30. The range of  $k$  values between  $-\pi/a$  and  $+\pi/a$  defines the first Brillouin zone, those of  $k$  values between  $-2\pi/a$  and  $-\pi/a$  and between  $+\pi/a$  and  $+2\pi/a$  define the second Brillouin zone.

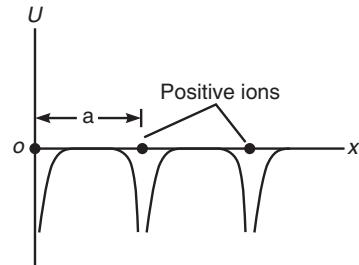


FIG. 27.28

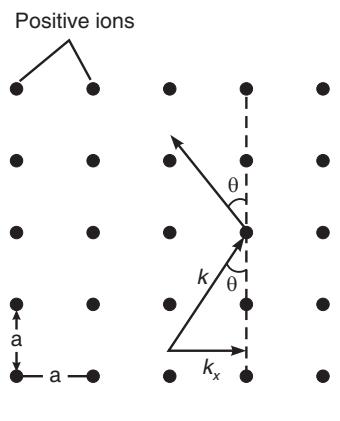


Fig. 27.29

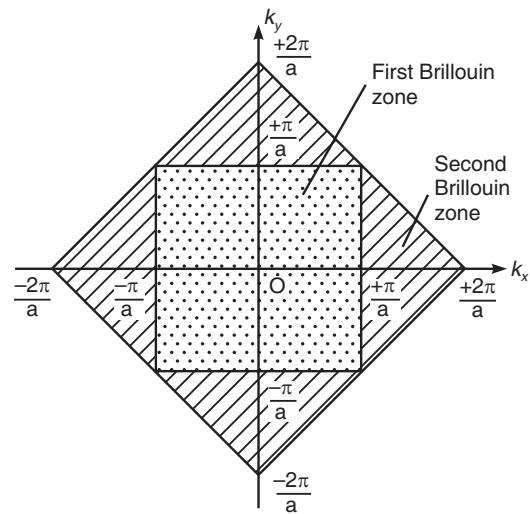


Fig. 27.30

### Origin of Forbidden Bands

The energy of a free electron is related to its wave number  $k$  by

$$E = \frac{\hbar^2 k^2}{2m} \quad \dots(5)$$

Fig. 27.31 shows how  $E$  varies with  $k$  in the  $x$  direction. The dashed curve is the free-electron parabola.

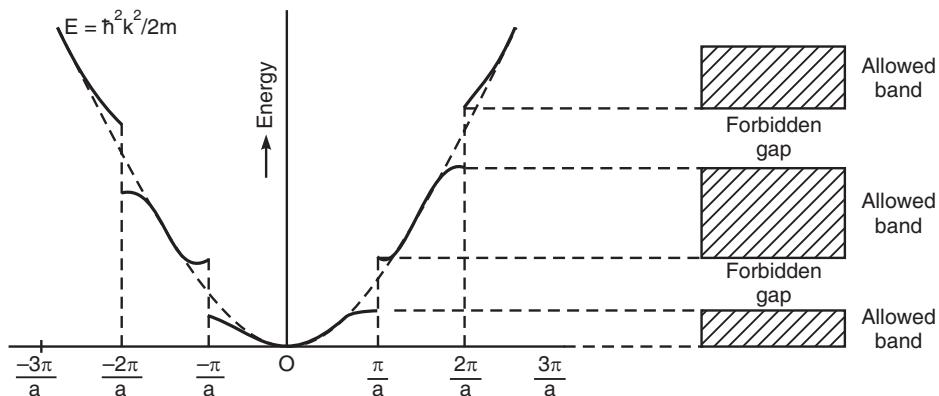


Fig. 27.31

Gaps in energy occur at values given by

$$k = \pm \frac{\pi}{a}, \pm \frac{2\pi}{a}, \pm \frac{3\pi}{a}, \dots$$

At the above values of  $k$  we get energy gaps, whereas for values of  $k$  not near these values the energies are much like that of a free electron. At  $k = \pi/a$ ,  $E$  has two values, the lower belonging to the first Brillouin zone and the higher to the second zone. There is a definite gap between the possible energies in the first and second Brillouin zones. This energy gap corresponds to a forbidden band.

The waves corresponding to values of  $k$  not satisfying the above condition travel almost freely. Those satisfying the condition, however, are reflected resulting in "standing" waves.

Let us consider electrons moving in the  $x$  direction. When  $k = \pm \pi/a$ , the waves are Bragg-reflected back and forth. So the only solutions of Schrodinger's equation consist of standing waves whose wavelength is equal to the periodicity of the lattice. There are two possibilities for these standing waves for  $n = 1$ :

$$\psi_1 = A \sin \frac{\pi x}{a}$$

$$\psi_2 = A \cos \frac{\pi x}{a}$$

Thus we get two *different* possible standing waves.

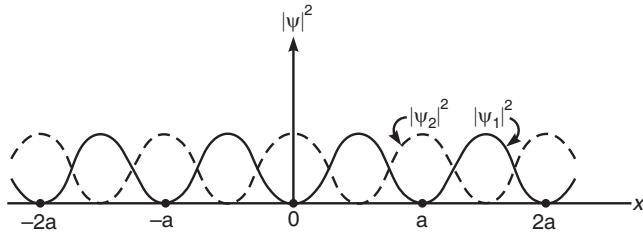


Fig. 27.32

The probability densities  $|\psi_1|^2$  and  $|\psi_2|^2$  for these waves are plotted in Fig. 27.32. Evidently  $|\psi_1|^2$  has its minima at the lattice points occupied by the positive ions ( $x = 0, \pm a, \pm 2a, \dots$ ).  $|\psi_2|^2$  has its maxima at the lattice points. The charge density corresponding to an electron wave function  $\psi$  is  $e|\psi|^2$ . The charge density in the case of  $\psi_1$  is concentrated *between* the positive ions. In the case of  $\psi_2$ , it is concentrated *at* the positive ions. The potential energy of an electron in a lattice of positive ions is greatest midway *between* each pair of ions and least *at* the ions themselves. So an electron for  $k = \pi/a$  can have two *different* energies  $E_1$  and  $E_2$  associated with the standing waves  $\psi_1$  and  $\psi_2$ . No other solutions are possible when  $k = \pm \pi/a$ , and accordingly no electron can have an energy between  $E_1$  and  $E_2$ . The difference  $E_1 - E_2$  is the energy gap. This is the origin of all the energy gaps.

## 27.17 BLOCH THEOREM

**Statement.** *The solutions of the Schrodinger equation for a periodic potential must be of the form*

$$\psi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad \dots(1)$$

Here,  $u_{\mathbf{k}}(\mathbf{r})$  has the period of the crystal lattice with  $u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r} + \mathbf{T})$ .

$\mathbf{T}$  is the crystal translation vector.

**Explanation.** The eigenfunctions of the wave equation for a periodic potential are the product of a plane wave  $\exp(i\mathbf{k} \cdot \mathbf{r})$  times a function  $u_{\mathbf{k}}(\mathbf{r})$  with the periodicity of the crystal lattice. The function  $\psi_{\mathbf{k}}(\mathbf{r})$  in Eq. (1) is called the *Bloch function*.

**Proof.** We assume that  $\psi_{\mathbf{k}}$  is nondegenerate. Consider  $N$  identical lattice points on a ring of length  $Na$ . The potential energy is periodic in  $a$ , with  $U(x) = U(x + sa)$ .

Here,  $s$  is an integer.

The ring has symmetry. We look for solutions of the wave equation such that

$$\psi(x + a) = C\psi(x), \quad \dots(2)$$

Here,  $C$  is a constant. Then, on going once around the ring,

$$\psi(x + Na) = \psi(x) = C^N \psi(x),$$

because  $\psi(x)$  must be single-valued.

So,  $C$  is one of the  $N$  roots of unity.

$$C = \exp(i2\pi s/N); \quad s = 0, 1, 2, \dots, N-1. \quad \dots(3)$$

Suppose  $u_k(x)$  has the periodicity  $a$ , so that  $u_k(x) = u_k(x+a)$ . Then

$$\psi(x) = u_k(x) \exp(i2\pi sx/Na)$$

with  $k = 2\pi s/Na$ , satisfies Eq. (2)

Thus the theorem is proved.

## 27.18 KRONIG-PENNEY MODEL

This model illustrates the behaviour of electrons in a periodic potential. It assumes a one-dimensional model of periodic potential (Fig. 27.33).

The wave equation can be solved in terms of elementary functions for the square-well array.

The wave functions associated with this model are calculated by solving the Schrodinger equation

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + U(x)\psi = \varepsilon\psi \quad \dots(1)$$

Here,  $U(x)$  is the potential energy and  $\varepsilon$  is the energy eigenvalue.

In the region  $0 < x < a$ ,  $U = 0$ . Eq. (1) becomes

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2} \varepsilon\psi = 0 \quad \dots(2)$$

The eigenfunction is

$$\psi = Ae^{ikx} + Be^{-ikx}, \quad \dots(3)$$

It represents plane waves travelling to the right and to the left, with energy

$$\varepsilon = \hbar^2 k^2/2m \quad \dots(4)$$

In the region  $-b < x < 0$ , the Schrodinger wave equation is

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2} [\varepsilon - U_0] \psi = 0 \quad \dots(5)$$

The eigenfunction is

$$\psi = Ce^{Qx} + De^{-Qx}, \quad \dots(6)$$

and  $U_0 - \varepsilon = \hbar^2 Q^2/2m \quad \dots(7)$

Since the potential is periodic, the wave functions must be of the form of Bloch function, i.e.,

$$\psi(x) = u_k(x) e^{ikx} \quad \dots(8)$$

Thus the solution in the region  $a < x < a+b$  must be related to the solution (6) in the region  $-b < x < 0$  by the Bloch theorem:

$$\psi(a < x < a+b) = \psi(-b < x < 0) e^{ik(a+b)}, \quad \dots(9)$$

The constants  $A, B, C, D$  are chosen so that  $\psi$  and  $d\psi/dx$  are continuous at  $x=0$  and  $x=a$ .

At  $x=0$ ,

$$A + B = C + D; \quad \dots(10)$$

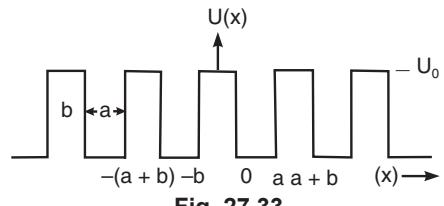


Fig. 27.33

$$iK(A - B) = Q(C - D). \quad \dots(11)$$

At  $x = a$ , with the use of (9) for  $\psi(a)$  under the barrier in terms of  $\psi(-b)$ ,

$$Ae^{ika} + Be^{-ika} = (Ce^{-Qb} + De^{Qb}) e^{ik(a+b)}; \quad \dots(12)$$

$$iK(Ae^{ika} - Be^{-ika}) = Q(Ce^{-Qb} - De^{Qb}) e^{ik(a+b)}. \quad \dots(13)$$

The four equations (10), (11), (12), (13) have a solution only if the determinant of the coefficients of  $A, B, C, D$  vanishes. This leads to the following equation:

$$[(Q^2 - K^2) / 2QK] \sinh Qb \sin Ka + \cosh Qb \cos Ka = \cos k(a + b) \quad \dots(14)$$

The result is simplified if we represent the potential by the periodic delta function obtained when we pass to the limit  $b = 0$  and  $U_0 = \infty$  in such a way that  $Q^2 ba / 2 = P$ , a finite quantity. In this limit  $Q \gg K$  and  $Qb \ll 1$ . Then (14) reduces to

$$(P/Ka) \sin Ka + \cos Ka = \cos ka \quad \dots(15)$$

*This is the condition for the solutions of the wave equation to exist.*

Figure 27.34 shows a plot of the function  $(P/Ka) \sin Ka + \cos Ka$  versus  $Ka$  for  $P = 3\pi/2$ .

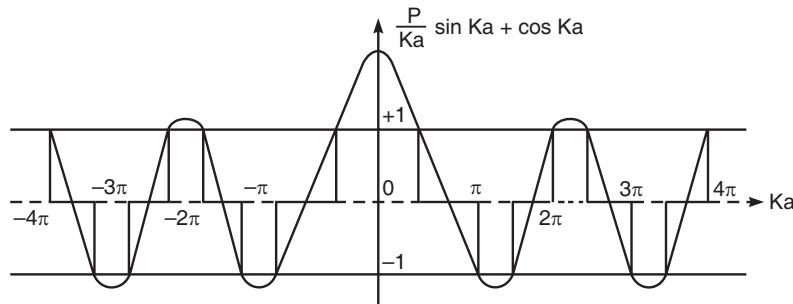


Fig. 27.34

The allowed values of the energy  $\epsilon$  are given by those ranges of

$$Ka = \left( \frac{2m\epsilon}{\hbar^2} \right)^{1/2} a \quad \dots(16)$$

for which the function lies between +1 and -1.

Clearly there are regions for  $Ka$  where the value of  $\left[ \frac{P \sin Ka}{Ka} + \cos Ka \right]$  does not lie between -1 and +1. For these values of  $Ka$  and hence of energy  $\epsilon$ , no solutions exist. Such regions of energy are prohibited and are called *forbidden bands*.

Thus, the *energy spectrum of the electron consists of alternate regions of allowed energy (continuous) and forbidden energy (dotted)*. These regions are usually referred as the allowed and forbidden energy bands.

Fig. 27.35 shows plot of energy vs. wavenumber for the Kronig-Penney potential, with  $P = 3\pi/2$ . Notice the energy gaps at  $ka = \pi, 2\pi, 3\pi, \dots$ . The figure shows the periodic dependence of  $\epsilon$  on  $k$  and demonstrates the existence of forbidden energy gaps.

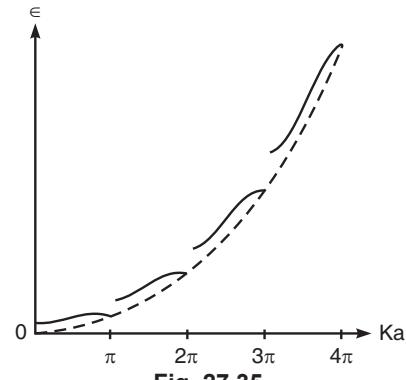


Fig. 27.35

## 27.19 EFFECTIVE MASS OF ELECTRON

We derive the equation of motion of an electron in an energy band. We look at the motion of a wave packet in an applied electric field  $E$ . Suppose that the wave packet is made up of wave functions near a particular wave vector  $k$ .

The frequency associated with a wave function of energy  $\epsilon$  is

$$\omega = \epsilon/\hbar \quad \dots(1)$$

The group velocity is

$$v_g = \frac{d\omega}{dk} = \frac{1}{\hbar} \frac{d\epsilon}{dk} \quad \dots(2)$$

The effects of the crystal on the electron motion are contained in the dispersion relation  $\epsilon(k)$ .

The work  $\delta\epsilon$  done on the electron by the electric field  $E$  in the interval  $\delta t$  is

$$\delta\epsilon = -eEv_g \delta t \quad \dots(3)$$

Using Eq. (2), we observe that

$$\delta\epsilon = (d\epsilon / dk) \delta k = \hbar v_g \delta k, \quad \dots(4)$$

On comparing Eq.(3) with Eq. (4), we have

$$\delta k = -(eE / \hbar)\delta t, \quad \dots(5)$$

or  $\hbar \frac{dk}{dt} = -eE \quad \dots(6)$

We may write Eq.(6) in terms of the external force  $F$  as

$$\hbar \frac{dk}{dt} = F \quad \dots(7)$$

In a crystal  $\hbar dk / dt$  is equal to the external force on the electron.

By differentiating the result of Eq. (2), we get the acceleration  $\frac{dv_g}{dt}$  of the wave packet.

$$\frac{dv_g}{dt} = \frac{l}{\hbar} \frac{d^2\epsilon}{dk dt} = \frac{l}{\hbar} \left( \frac{d^2\epsilon}{dk^2} \frac{dk}{dt} \right) \quad \dots(8)$$

From Eq. (7) we have

$$dk/dt = F/\hbar$$

Eq.(8) becomes

$$\frac{dv_g}{dt} = \left( \frac{1}{\hbar^2} \frac{d^2\epsilon}{dk^2} \right) F$$

or  $F = \frac{\hbar^2}{d^2\epsilon / dk^2} \frac{dv_g}{dt} \quad \dots(9)$

If we identify  $\hbar^2/(d^2\epsilon / dk^2)$  as a mass, then Eq. (9) assumes the form of Newton's second law. We define the effective mass  $m^*$  by

$$\frac{1}{m^*} = \frac{1}{\hbar^2} \frac{d^2\epsilon}{dk^2} \quad \dots(10)$$

An electron in a periodic potential is accelerated relative to the lattice in an applied electric or magnetic field as if the mass of the electron were equal to an effective mass  $m^*$ , defined by Eq.(10).

It is easy to generalize Eq. (10) to take account of an anisotropic energy surface, as for electrons in Si or Ge. We introduce the components of the reciprocal *effective mass tensor*:

$$\left(\frac{1}{m^*}\right)_{\mu\nu} = \frac{1}{\hbar^2} \frac{d^2\epsilon_k}{dk_\mu dk_\nu}; \quad \frac{dv_\mu}{dt} = \left(\frac{1}{m^*}\right)_{\mu\nu} F_\nu \dots (11)$$

Here,  $\mu, \nu$  are Cartesian coordinates.

**Effective Mass Tensor.** In a crystal, for an electron the ratio of force and acceleration is generally a tensor of second order. It is called the *effective mass tensor*. Eq. (11) is the formula of effective mass tensor for nearly free electron model. The formula would be valid for real potentials for an electron in a crystal. In simplest situations this tensor is transformed into a scalar.

**Nature of Effective Mass.** For the electron in a crystal the value of  $m^*$  depends on its energy (Fig. 27.36). Fig. 27.36 shows the variation of energy  $\epsilon$  with wave vector  $k$ . The effective mass  $m^*$  is determined by the curvature  $d^2\epsilon / dk^2$ .

(i) In region 1, the parabolic relation is good, and behaviour is therefore classical.

(ii) In region 2, the curvature is negative. Therefore, the effective mass is *negative*.

(iii) In region 4, the curvature is positive but larger than that of region 1. Therefore, the effective mass is *positive*. The effective mass is smaller, according to Eq. (10).

#### Physical Interpretation of the Effective Mass

A single electron in an energy band may have positive or negative effective mass  $m^*$ .

(i) The states of positive effective mass occur near the bottom of a band because positive effective mass means that the band has upward curvature ( $d^2\epsilon / dk^2$  is positive).

(ii) States of negative effective mass occur near the top of the band. A negative effective mass means that on going from state  $k$  to state  $k + \Delta k$ , the momentum transfer to the lattice from the electron is larger than the momentum transfer from the applied force to the electron.

(iii) If the energy in a band depends only slightly on  $k$ , then the effective mass will be very large. That is,  $m^*/m > 1$  when  $d^2\epsilon / dk^2$  is very small.

#### Explanation of negative masses which occur near a Brillouin zone boundary :

(i) In Fig. 27.37 (a), the energy of the electron beam incident on a thin crystal is slightly too low to satisfy the condition for Bragg reflection and the beam is transmitted through the crystal.

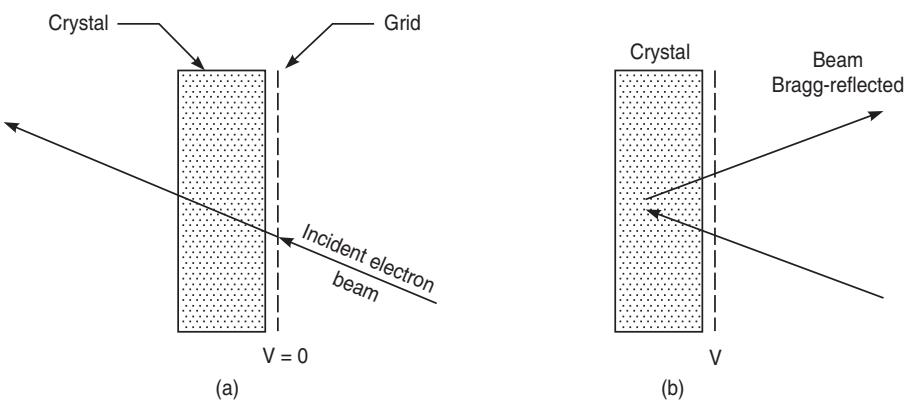


Fig. 27.37

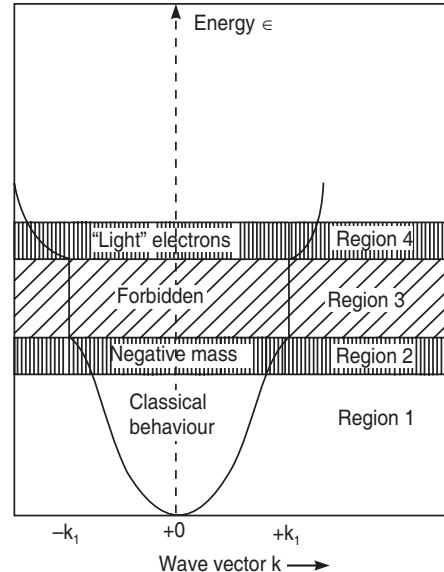


Fig. 27.36

(ii) In Fig. 27.37 (b), a small accelerating potential applied on the grid leads to the required condition for Bragg reflection. Then the electron beam will be reflected by appropriate set of crystal planes. Thus, a small change in energy of electron beam would induce a large change in the momentum associated with the small effective mass. If the grid potential accelerates an electron beam of small energy, the effective mass is negative (Because the acceleration occurs in a direction opposite to change in momentum).

### EXERCISE

1. The current density ( $J$ ) and electric field intensity ( $E$ ) can be related by  
 (a)  $J = \sigma + E$       (b)  $J = \sigma - E$       (c)  $J = \sigma E$       (d)  $J = \sigma E^2$   
 (B.U. 2012)
2. The electrical conductivity  $\sigma = \underline{\hspace{2cm}}$   
 (a)  $\frac{ne^2\lambda v}{2k_B T}$       (b)  $\frac{ne^2\lambda v}{4k_B T}$       (c)  $\frac{ne^2\lambda v}{6k_B T}$       (d)  $\frac{ne^2\lambda v}{8k_B T}$   
 (B.U. 2013)
3. Wideman and Franz ratio is  $\frac{k}{\sigma} = \underline{\hspace{2cm}}$   
 (a)  $3\left(\frac{k_B}{e}\right)^2 T$       (b)  $2\left(\frac{k_B}{e}\right)^2 T$       (c)  $4\left(\frac{k_B}{e}\right)^2 T$       (d)  $\left(\frac{k_B}{e}\right)^2 T$   
 (B.U. 2013)
4. At  $T = 0K$ , if  $E > E_{F_0}$  then  $F(E) = \underline{\hspace{2cm}}$   
 (a) 0      (b)  $\frac{1}{2}$       (c)  $\frac{1}{4}$       (d) 1  
 (B.U. 2013)
5. The Hall coefficient  $R_H$  is  
 (a)  $R_H = ne$       (b)  $R_H = n + e$       (c)  $R_H = 1/ne$       (d)  $R_H = ne/2$   
 (B.U. 2012)
6. If Hall coefficient is positive, then the material is called  
 (a) insulator      (b)  $p$ -type semiconductors  
 (c)  $n$ -type semiconductors      (d) conductors  
 (B.U. 2011)
7. The unit of hall coefficient is  
 (a)  $Vm^3 A^{-1} Wb^{-1}$       (b)  $Vm^3 AWb^{-1}$       (c)  $Vm^3 A^{+1} Wb^{-1}$       (d)  $Vm^2 A^{-2} Wb$   
 (B.U. 2010)
8. Mobility ' $\mu$ ', Hall coefficient ' $R_H$ ' and conductivity ' $\sigma$ ' are related by  
 (a)  $\mu = \frac{R_H}{\sigma}$       (b)  $\mu = R_H \cdot \sigma$       (c)  $R_H = \mu \sigma^2$       (d)  $R_H = \frac{\sigma^2}{\mu}$   
 (B.U. M.Sc. 2006)
9. The Hall angle is  $\underline{\hspace{2cm}}$   
 (a)  $\phi = -H_z$       (b)  $\phi = -\mu$       (c)  $\phi = \mu H_z$       (d)  $\phi = H_z$   
 (B.U. 2014)  
 [Ans. 1. (c), 2. (c), 3. (a), 4. (a), 5. (c), 6. (b), 7. (a), 8. (b), 9. (c)]
10. Briefly describe Drude's free-electron theory of metals.      (B.U. 2010)
11. Explain Drift Velocity and relaxation time of free electrons in metals.      (B.U. 2012)
12. (i) Deduce a mathematical expression for electrical conductivity and thermal conductivity of a conducting material and hence obtain Wiedman-Franz law.      (B.U. 2012)  
 (ii) State and explain Wide-Mann and Franz law.      (B.U. 2015)

- 
- 13. Describe the free electron model of a metal given by Sommerfield. (B.U. 2012)
  - 14. Obtain expression for the Fermi energy, the total energy and the density of states for a free electron gas in one dimension.
  - 15. Derive an expression for the Fermi energy of a free electron gas in three dimensions. (Raj., 1975)
  - 16. State and explain Schotcky effect. (B.U. 2014)
  - 17. State and explain Hall effect. (B.U. 2014)
  - 18. Explain Hall voltage and Hall coefficient. (B.U. 2013)
  - 19. What is Hall effect? Find an expression for Hall coefficient of a metal and describe the experimental set up to measure it. (B.U. 2015)
  - 20. Derive an expression for Mobility and Hall angle. (B.U. 2014)
  - 21. Mention the applications of Hall effect. (B.U. 2011)
  - 22. Describe the experimental method for the determination of Hall voltage on the basis of the elementary theory of Hall effect. (B.U. 2011)
  - 23. Explain how the semiconducting material can be classified into p-type and n-type semiconductors using Hall constant. (B.U. 2011)
  - 24. Classify the solids based on band theory of solids. (B.U. 2015)
  - 25. Based on the band theory of solids, explain nature of conductors, semiconductors and insulators. (B.U. 2011)
  - 26. Explain optical properties of solids. (B.U. 2014)
  - 27. Discuss Kronig Penney model. Using the model show that the energy spectrum of electron consists of a number of allowed energy bands separated by forbidden regions. (Meerut U, 2006)
  - 28. Give a qualitative discussion of energy bands in solids. Explain the concept of effective mass of electrons. (Meerut M.Sc. 1988)

## SUPERCONDUCTIVITY

**After reading this chapter, you should be able to**

- ◆ Explain the practical and theoretical aspects of superconductivity
- ◆ List the properties of type-I and type-II superconductors
- ◆ Explain the behaviour of superconductors in external electric and magnetic fields
- ◆ Discuss qualitative ideas of BCS theory
- ◆ List the properties of high- $T_C$  superconductors
- ◆ Discuss some important applications of superconductors

## 28.1 INTRODUCTION

- Superconductivity was first observed in 1911 by Dutch physicist H.K. Onnes in the course of his experiments on the electrical conductivities of metals at low temperatures. He observed that as purified mercury is cooled, its resistivity vanished abruptly at 4.2 K (Fig. 28.1). Above this temperature, the resistivity is small, but finite. The resistivity below this point is so small that it is essentially zero.
- The temperature at which the transition takes place is called the critical temperature ( $T_c$ ).
- The temperature ( $T_c$ ), which marks the transition of a normal conductor to the superconducting state, is defined as the transition temperature.
- Above the critical temperature ( $T_c$ ), the substance is in the familiar *normal state*, but below ( $T_c$ ) it enters an entirely different *superconducting state*.
- The superconducting state is marked by a sudden fall of the electrical resistivity of the material to nearly zero, when it is cooled to a sufficiently low temperature.

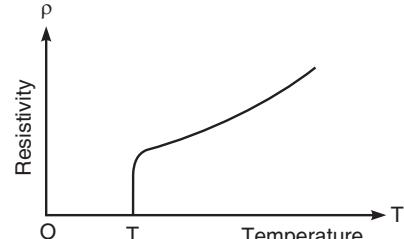


Fig. 28.1

### Definitions

(1) **Superconductivity.** Certain metals and alloys exhibit almost zero resistivity (i.e., infinite conductivity) when they are cooled to sufficiently low temperatures. This phenomenon is called superconductivity.

- Figure 28.2 shows the curve between the resistivity ( $\rho$ ) and temperature ( $T$ ) of normal metal and superconductor at very low temperatures.

The superconducting transition temperature  $T_c$  of a material is defined as a critical temperature at which the resistivity of the material is suddenly changed to zero. Thus at that temperature, a material is changed from normal material to superconductor.

The transition region is of finite width  $AB$ .

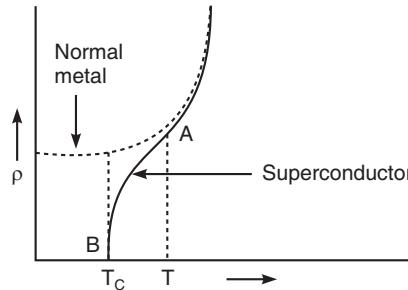


Fig. 28.2

- When impurities are added to superconducting elements, the superconducting property is not lost but the transition-temperature is lowered.

(2) **Superconductor.** A material which exhibits superconductivity is called superconductor or superconducting material.

(3) **Transition temperature.** The temperature at which the transition from normal state to superconducting state takes place on cooling in the absence of magnetic field is called the critical temperature ( $T_c$ ) or the transition temperature.

### Nature of Superconductivity

We can gain some insight into the nature of superconductivity using the free electron model.

$$\text{The resistivity of a metal is } \rho = \frac{m}{ne^2 \tau}. \quad \dots (1)$$

Here,  $m$  = mass of the electron,  $e$  = charge of the electron,  $n$  = No. of electrons per unit volume,  $\tau$  = collision time.

- The vibration of the ions in a crystal decreases with decreasing temperature. This has the effect of decreasing the probability of an electron-ion collision or of increasing the collision time  $\tau$ . Thus  $\rho$  decreases as the temperature is lowered.

- According to Eq. (1), this implies that the resistivity of a metal should tend toward zero, as the temperature approaches zero. If  $\tau$  becomes infinite at sufficiently low temperatures, then the resistivity vanishes entirely, which is what is observed in superconductivity. As the temperature is lowered below  $T_c$ , a fraction of the electrons become superconducting, in the sense that they have infinite collision times. These electrons undergo no scattering whatsoever, even though the substance may contain some impurities and defects. It is these electrons which are responsible for superconductivity.
- Onnes found that the superconducting transition is *reversible*: When he heated the superconducting sample, it recovered its normal resistivity at the temperature  $T_c$ .
- Superconductivity occurs in metallic elements (silver, lead etc.) of the periodic system and also in alloys and semiconductors. The range of transition temperatures at present extends from 23.2 K for the alloy  $Nb_3 Ge$  to 0.01 K for some semiconductors.

## 28.2 PROPERTIES OF SUPERCONDUCTORS

### 1. Zero electrical resistance

The first characteristic property of a superconductor is its electrical resistance. The electrical resistance of the superconductor is zero below the transition temperature ( $T_c$ ).

Superconductors have virtually zero electrical resistance and they can conduct electricity without resistance.

### 2. Effect of magnetic field

Below the transition temperature of a material ( $T_c$ ), its superconductivity can be destroyed by the application of a strong magnetic field.

**At temperatures** below  $T_c$ , as the magnetic field strength reaches a critical value  $H_c$ , the superconductivity disappears.

**Critical magnetic field.** The minimum magnetic field strength required to destroy the superconducting property is known as *critical field* ( $H_c$ ).

#### Relation between Critical Magnetic Field and Temperature

The critical magnetic field ( $H_c$ ) depends upon the temperature of the superconducting material. The relation between critical magnetic field and temperature is given by

$$H_c = H_0 \left[ 1 - \left( \frac{T}{T_c} \right)^2 \right]$$

Here,  $H_0 \rightarrow$  critical magnetic field at absolute zero temperature (0 K) of the material.

$T_c \rightarrow$  superconducting transition temperature of the material.

$T$  is the temperature below  $T_c$  of the superconducting material.

$H_c \rightarrow$  Critical magnetic field of the material.

(i) At  $T = T_c$ ,  $H_c = 0$ .

The critical magnetic field is zero at superconducting transition temperature.

(ii)  $H_c$  increases as the temperature decreases below  $T_c$ .

(iii) At  $T = 0K$ ,  $H_c = H_0$ .

Figure 28.3 shows the variation of critical magnetic field  $H_c$  with temperature  $T(K)$  in a superconductor.

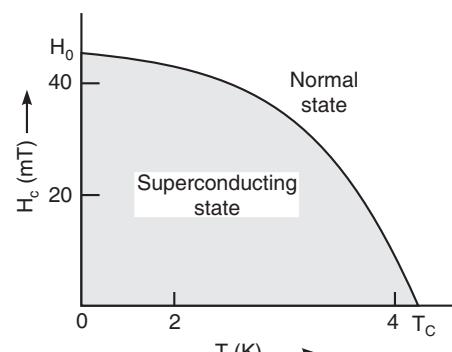


Fig. 28.3

**EXAMPLE 1.** Superconducting tin has a critical temperature of 3.7 K at zero magnetic field and a critical field of 0.0306 tesla at 0 K. Find the critical field at 2K. (Bharathiar Univ. 2010)

**SOL. Formula:**  $H_c = H_0 \left[ 1 - \left( \frac{T}{T_c} \right)^2 \right]$

**Given:** Critical temperature  $T_c = 3.7$  K

Critical magnetic field at 0 K,  $H_0 = 0.0306$  tesla

Temperature,  $T = 2$  K

Critical field at temperature  $T$ ,  $H_c = ?$

Substituting the given values, we have

$$H_c = 0.0306 \left[ 1 - \left( \frac{2}{3.7} \right)^2 \right]$$

$$H_c = 0.02166 \text{ tesla}$$

### 3. Effect of electric current

The critical magnetic field required to destroy superconductivity need not necessarily be applied externally. An electric current flowing through the superconducting material itself may give rise to necessary magnetic field. For example, when a current flows through a superconducting wire, it gives rise to its own magnetic field. As the current is increased to a critical value  $i_c$ , the associated magnetic field becomes  $H_c$  and the superconductivity disappears.

*The minimum current which can destroy the superconducting state of a superconductor is defined as critical current.*

Consider a superconducting wire of radius  $r$  (Fig. 28.4)

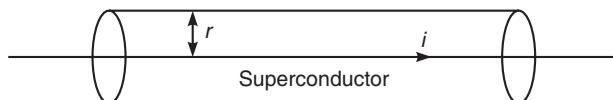


Fig. 28.4

Let  $i$  be the current flowing through the wire. The flow of high current induces a magnetic field. This induced magnetic field in the conductor destroys the superconducting property.

- The critical current ( $i_c$ ) required to destroy the superconducting property is given by

$$i_c = 2\pi r H_c \quad \dots (1)$$

Here,  $H_c \rightarrow$  critical magnetic field.

$r \rightarrow$  radius of the superconducting wire.

Thus, Eq. (1) limits the maximum possible current that flows through superconductor. Hence this is the main hurdle in producing high field superconducting magnets.

### 4. Persistent current

When the superconductor in the form of a ring is placed in a magnetic field, then a current is induced in it by electromagnetic induction (Fig. 28.5).

If the ring is in normal conducting state, the current decreases quickly because of the resistance of the ring.

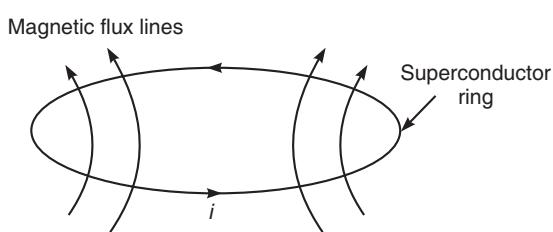


Fig. 28.5

- Since the ring is in superconducting state, the ring has *zero resistance*. Once the current is set up, it flows indefinitely without any decrease in its value.  
The current persists even after the removal of the magnetic field.
- Such a steady current which flows through a superconducting ring without any decrease in its strength, as long as the material is in superconducting state, is called persistent current.*

### 5. Meissner effect

- When a superconducting material in its normal conducting state is placed in a uniform magnetic field of flux density  $B$ , the magnetic lines of force penetrate through the material [Fig. 28.6 (a)].
- When the material placed in a uniform magnetic field ( $H < H_c$ ) is cooled below  $T_c$ , the magnetic flux inside the material is excluded from the material [Fig. 28.6 (b)]. Thus, inside superconducting specimen, magnetic induction ( $B$ ) is zero. This phenomenon is called *Meissner effect*.

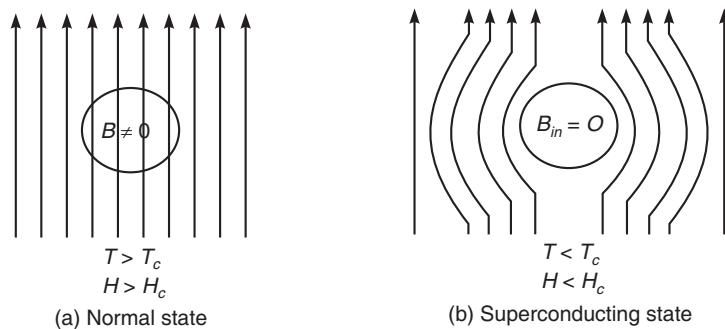


Fig. 28.6

- A superconductor is a perfect diamagnetic material, with the magnetic susceptibility  $\chi$  equal to  $-1$ .
  - When a weak magnetic field is applied to a superconducting specimen at a temperature below transition temperature  $T_c$ , the magnetic flux lines are expelled as shown in Fig. 28.6(b).
- In the superconducting state, the flux lines of a magnetic field are ejected out of the superconductor. The specimen acts as an ideal diamagnet. This effect is called *Meissner effect*.
- Superconductor is a perfect diamagnet.** The magnetic induction ( $B$ ) inside the specimen is given by

$$B = \mu_0(H + M) \quad \dots (1)$$

$H \rightarrow$  external applied magnetic field

$M \rightarrow$  magnetisation produced inside the specimen.

When the specimen is superconducting, according to Meissner effect, inside the bulk superconductor  $B = 0$ . Hence

$$\mu_0(H + M) = 0$$

Since  $\mu_0 \neq 0, H + M = 0$

or  $M = -H$

$$\text{Magnetic susceptibility, } \chi = \frac{M}{H} = -1 \quad \dots (3)$$

Thus the material is perfectly diamagnetic.

**EXAMPLE 1.** Prove that susceptibility of a superconductor is  $-1$  and relative permeability is zero.

**SOL.** For a magnetic material, the flux density is given by

$$B = \mu_0(M + H) \quad \dots (1)$$

Here,  $H$  is the applied field and  $M$  is the induced magnetisation.

But  $B = 0$  inside the superconductor.

$$\therefore M = -H \quad \dots (2)$$

$$\text{Magnetic susceptibility, } \chi = \frac{M}{H} = -1 \quad \dots (3)$$

Further  $B = \mu H = \mu_0 \mu_r H = \mu_0(H + M) = 0$

$$\mu_r H = H + M = 0 \quad \dots (4)$$

$$\therefore \mu_r = 0$$

Equations (3) and (4) indicate that a superconductor behaves as a perfect diamagnetic material.

## 6. Isotopic effect

- The critical temperature of superconductors varies with isotopic mass.

**Example.** In mercury the critical temperature  $T_c$  varies from  $4.185\text{ K}$  to  $4.146\text{ K}$  as the average atomic mass  $M$  varies from  $199.5$  to  $203.4$  atomic mass units.

- The transition temperature changes smoothly when we mix different isotopes of the same element.
- The experimental results within each series of isotopes may be fitted by a relation of the form

$$M^\alpha T_c = \text{constant}$$

- Maxwell found that the transition temperatures ( $T_c$ ) are inversely proportional to the square roots of the atomic masses ( $M$ ) of the isotopes of a single superconductor. Thus,

$$M^\alpha T_c = \text{constant.}$$

Here,  $\alpha$  is a constant and it is approximately equal to  $0.5$ .

$$\therefore M^{\frac{1}{2}} T_c = \text{constant i.e., } T_c \propto M^{-\frac{1}{2}} \quad (\text{Fig. 28.7})$$

- From the dependence of  $T_c$  on the isotopic mass, we learn that lattice vibrations and hence electron-lattice interactions are deeply involved in superconductivity.

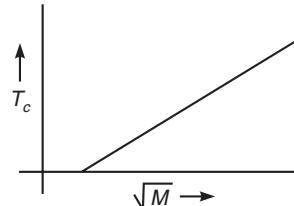


Fig. 28.7

## 7. Thermodynamic effects

- The thermal properties such as entropy, thermal conductivity and specific heat capacity of a substance change abruptly, when it passes over into the superconducting state.

(i) **Entropy.** In all superconductors, the entropy decreases markedly on cooling below the critical temperature  $T_c$ . Measurements for aluminium are plotted in Fig. 28.8.

- The decrease in entropy between the normal state and the superconducting state tells us that the superconducting state is more ordered than the normal state, because the entropy is a measure of the disorder of a system.

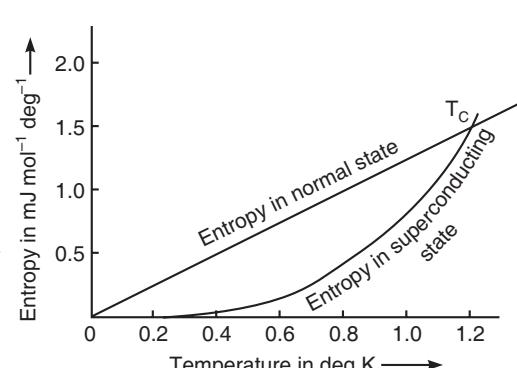


Fig. 28.8. Entropy  $S$  of aluminum in the normal and superconducting states as a function of the temperature.

- Some or all of the electrons thermally excited in the normal state are ordered in the superconducting state. The change in entropy is small, in aluminum of the order of  $10^{-4} k_B$  per atom.
- The small entropy change must mean that only a small fraction (of the order of  $10^{-4}$ ) of the conduction electrons participate in the transition to the ordered superconducting state.

(ii) **Thermal conductivity.** The thermal conductivity of superconductors undergoes a continuous change between the two phases. It is usually lower in the superconducting phase. This suggests that the electronic contribution drops and possibly the superconducting electrons playing no part in the heat transfer.

- Variation of thermal conductivity with temperature for tin is shown in Fig. 28.9.

It can be seen that thermal conductivity is lower in superconducting state.

(iii) **Specific heat.** The specific heat  $C_v$  in a normal conductor at low temperatures, consists of two contributions,  $C_v^e$  from the electrons in the conduction band and  $C_v^l$  from the lattice. Thus,

$$C_v = C_v^e + C_v^l = \gamma T + \beta T^3$$

$C_v^e$  is linearly proportional to  $T$  while  $C_v^l$  is proportional to  $T^3$ , and is explained by Debye theory. The contribution to the specific heat by the electrons ( $C_v^e$ ) in the superconducting state is plotted in Fig. 28.10 and 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  $C_v^e$  is obtained by subtracting the lattice contribution  $C_v^l$ . When the field is applied the metal is in the normal state, and when it is not applied 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.

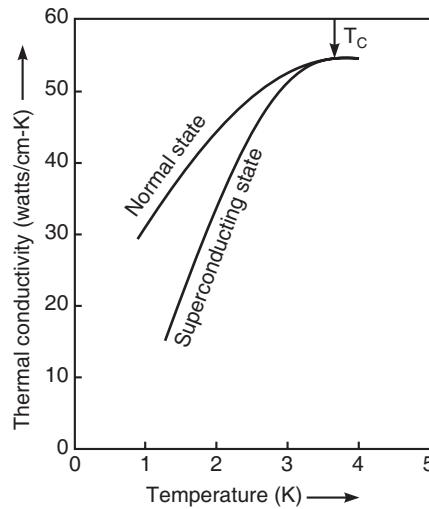


Fig. 28.9

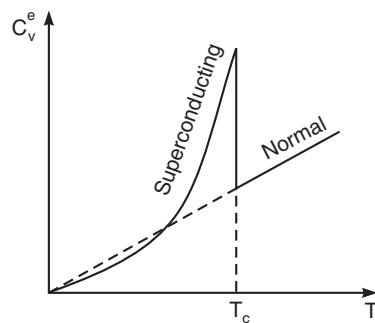


Fig. 28.10. The temperature dependence of electronic specific heat of a conductor in the normal and superconducting states

### 28.3 CLASSIFICATION OF SUPERCONDUCTORS

The superconductors can be classified into two distinct groups according to their behaviour in external magnetic field:

#### 1. Type-I Superconductors

The superconductors, in which the magnetic field is totally excluded from the interior of superconductors below a certain magnetising field  $H_C$  and at  $H_C$  the material loses superconductivity and the magnetic field penetrates fully are termed as type I or soft superconductors.

**Examples :** Tin, Lead, Aluminium, Mercury, Zinc, Magnesium etc. The magnetization curve for type I superconducting materials is shown in Fig. 28.11.

The important characteristics of type I superconductors are:

- (i) They exhibit complete Meissner effect.
- (ii) The critical values of magnetic field  $H_C$  at which magnetisation drops are very low. The maximum known critical field for type I superconductor is of the order of 0.1 T. The low value of  $H_C$  makes these materials unsuitable for use in high field superconducting magnets.

The magnetisation curve shows that transition at  $H_C$  is reversible. This means that if the magnetic field is reduced below  $H_C$ , the material again acquires superconducting property and the field is expelled.

- (iii) Below  $H_C$  the material is superconductor and above  $H_C$  it becomes a conductor.

## 2. Type II (or hard) Superconductors

*The superconductors in which the material loses magnetisation gradually rather than suddenly are termed as type II or hard superconductors.*

The magnetisation curve for these superconductors is shown in Fig. 28.12.

The important characteristics are :

- (i) They do not show complete Meissner effect.
- (ii) These superconductors have two critical fields:  $H_{C1}$  (the lower critical field) and  $H_{C2}$  (the upper critical field). The specimen is diamagnetic below  $H_{C1}$ , i.e., the magnetic field is completely excluded below  $H_{C1}$ . At  $H_{C1}$  the flux begins to penetrate the specimen, and the penetration of flux increases until the upper critical field  $H_{C2}$  is reached. At  $H_{C2}$  the magnetisation vanishes and specimen returns to normal conducting state. The value of  $H_{C2}$  is greater than  $H_C$ . The value of critical field for type II materials may be 100 times or more higher than the value of  $H_C$  for type I superconductors. Critical fields  $H_{C2}$  up-to 30 T have been observed.

The materials which display type II behaviour are essentially inhomogeneous and include Nb-Zr, Nb-Ti alloys and Va-Ga and Nb-Sn intermetallic compounds. These are technically more useful than type I superconductors due to tolerating high magnetic fields.

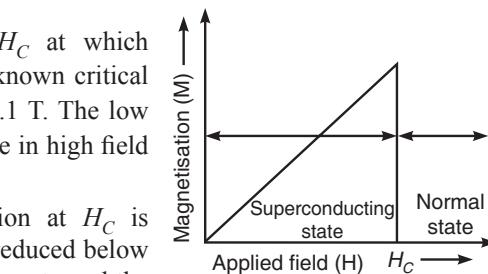


Fig. 28.11

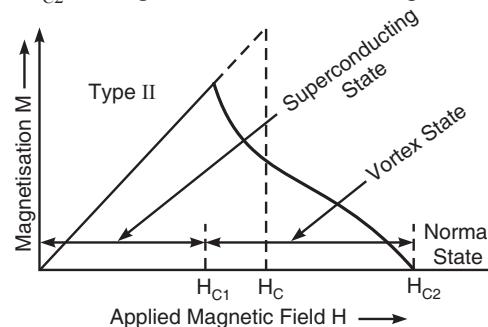


Fig. 28.12

## 28.4 BCS THEORY OF SUPERCONDUCTIVITY

Bardeen, Cooper and Schrieffer put forward the microscopic theory of superconductivity in 1957. The two important features of BCS theory are:

- (1) Electrons from pairs (called *Cooper pairs*) which propagate throughout the lattice.
- (2) The propagation of Cooper pairs is without resistance because the electrons move in resonance with phonons. Therefore, the interaction described by the BCS theory is called *electron-phonon-electron interaction*.

### 1. Electron-electron interaction via lattice deformation

Consider an electron passing through the lattice of positive ions. (Fig. 28.13).

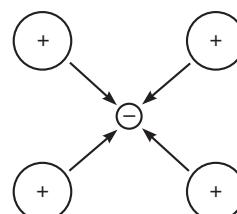


Fig. 28.13

The electron is attracted by the neighbouring positive ions. It forms a positive ion core and gets screened by them. The screening greatly reduces the effective charge of this electron. Due to the attraction between the electron and the ion core, the lattice gets deformed on local scale.

Now if another electron passes by the side of the assembly of the said electron and the ion core, it gets attracted towards the assembly. The second electron interacts with the first electron via lattice deformation.

#### Exchange of virtual phonon between two electrons

The interaction is due to the exchange of a virtual phonon,  $q$ , between the two electrons (Fig. 28.14).

- An electron with wave vector  $k_1$  polarizes the lattice, creating a phonon with wave vector  $q$ .
- Another electron with wave vector  $k_2$  absorbs the phonon.
- The end result is two electrons with wave vectors  $k_2 + q$  and  $k_1 - q$ .

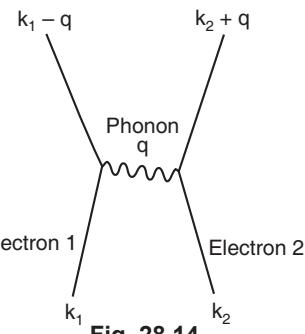


Fig. 28.14

These two electrons together form a *Cooper pair* and is known as *Cooper electron*.

#### 2. What is Cooper pair of electrons?

Cooper showed how two electrons in a superconductor could form a bound state despite their coulomb repulsion. If the attraction is stronger than the repulsion between the electrons, the electrons are effectively coupled together into a *Cooper pair* with the deformed lattice as the intermediary.

*Cooper showed that if there is an attraction between the two electrons, they are able to form a bound state so that their total energy is less than  $2E_F$ . These electrons are paired to form a single system called **Cooper pair**. Their motions are correlated. The binding is strongest when the electrons forming the pair have opposite momenta and opposite spins, i.e.,  $k\uparrow, -k\downarrow$ . All electron pairs with attraction among them and lying in the neighbourhood of the Fermi surface form Cooper pairs.*

In superconducting material, when a pair of electrons lying just below the Fermi surface is taken just above it, they form a Cooper pair and their total energy is reduced. This continues until the system can gain no additional energy by pair formation. Thus the total energy of the system is further reduced.

At absolute zero in superconductors, there is no abrupt discontinuity in the distribution of electrons across the Fermi surface. These super electrons are responsible for the resistivity at low temperatures.

#### 3. Formation of Cooper pair.

Figure 28.15 shows the movement of electrons in a one-dimensional lattice.

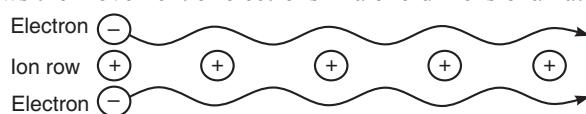


Fig. 28.15

(i) Two electrons propagate along a single lattice row. Each electron experiences an attraction towards its nearest positive ion.

(ii) When the electrons get very close to each other in the region between ions, they repel each other due to their mutual coulomb force.

(iii) In an equilibrium condition, a balance between attraction and repulsion is established. The two electrons (fermions) combine to form Cooper pair (bosons).

The collection of such bosons in a bulk sample condenses to form the superconducting state.

**Coherence length.** It is defined as the distance over which two electrons combine to form a Cooper pair. In other words, it is the smallest dimension over which superconductivity can be established or destroyed.

#### 4. Zero resistivity of a superconductor

BCS theory successfully explains the zero resistivity of a superconductor.

Figure 28.16 shows the movement of electrons between rows of a two-dimensional lattice.

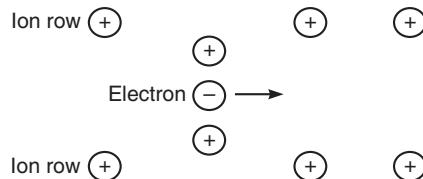


Fig. 28.16

Consider one of the electrons of the Cooper pair propagating through the lattice. The coulomb attraction between the electron and ions deforms the lattice which is propagated along with the electron. This propagating wave is associated with phonon transmission. The electron-phonon resonance allows the electron along with its pair elsewhere in the lattice to move without resistance.

#### 5. Flux quantization in a superconducting ring

Magnetic flux  $\phi$  through a superconducting ring is quantized i.e.,

$$\phi = n\phi_0 = n \frac{h}{2e}.$$

Here,  $n$  is an integer and

$\phi_0 = h/2e = 2.068 \times 10^{-15} \text{ Tm}^2$  is referred to as a *quantum fluxoid*.

The BCS ground state involves pair of electrons; thus flux quantization in terms of the pair charge  $2e$  is a consequence of the theory.

#### 6. Energy gap at 0 K.

The BCS theory relates the energy gap of a superconductor at 0 K to its critical temperature  $T_c$  by the formula

$$\text{Energy gap at } 0 \text{ K} \quad E_g(0) = 3.53 kT_c \quad \dots (1)$$

Equation (1) agrees fairly well with the observed values of  $E_g$  and  $T_c$ . At temperatures above 0K, some Cooper pairs break up. The resulting individual electrons interact with the remaining Cooper pairs and reduce the energy gap. Finally, at the critical temperature  $T_c$ , the energy gap disappears, there are no more Cooper pairs, and the material is no longer superconducting.

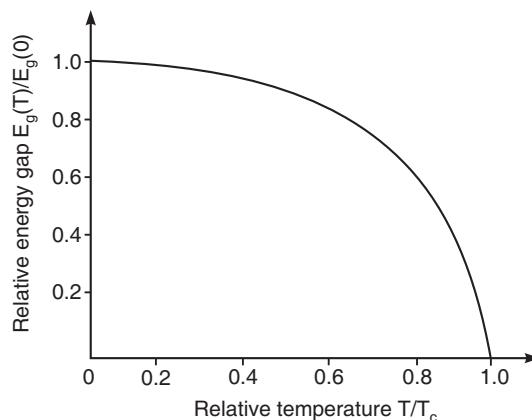


Fig. 28.17

Figure 28.17 shows the variation of the superconducting energy gap with temperature.

Here  $E_g(T)$  is the energy gap at the temperature  $T$  and  $E_g(0)$  is the gap at  $T=0$ .  $T_c$  is the critical temperature of the material.

- The conventional BCS theory accounts for superconductivity only at low temperatures. At higher temperatures, thermal degradation breaks the Cooper pairs leading to ordinary electrical conductivity. Thus it is unable to explain the formation of high temperature superconductivity having high  $T_c$  ( $= 80$  K).

## 28.5 LONDON EQUATION

Electrical conduction in the normal state of a metal is described by Ohm's law  $\mathbf{j} = \sigma \mathbf{E}$ . We modify this to describe conduction and the Meissner effect in the superconducting state. We assume that in the superconducting state the current density is directly proportional to the vector potential  $\mathbf{A}$  of the local magnetic field. Here,  $\mathbf{B} = \text{curl } \mathbf{A}$ . The constant of proportionality is  $-1/\mu_0 \lambda_L^2$ . Thus

$$\mathbf{j} = -\frac{1}{\mu_0 \lambda_L^2} \mathbf{A} \quad \dots(1)$$

Here,  $\lambda_L$  is a constant with the dimensions of length.

This is the London equation.

Taking the curl of both sides, we get

$$\text{curl } \mathbf{J} = -\frac{1}{\mu_0 \lambda_L^2} \mathbf{B} \quad \dots(2)$$

Now, consider the Maxwell equation

$$\text{curl } \mathbf{B} = \mu_0 \mathbf{j} \quad \dots(3)$$

Taking curl, we obtain

$$\text{curl curl } \mathbf{B} = \mu_0 \text{curl } \mathbf{j} \quad \dots(4)$$

Now,  $\text{curl curl } \mathbf{B} = \text{grad div } \mathbf{B} - \nabla^2 \mathbf{B}$

$$\text{curl curl } \mathbf{B} = -\nabla^2 \mathbf{B} \quad (\because \text{div } \mathbf{B} = 0) \quad \dots(5)$$

On combining Eqs. (4) and (5),

$$-\nabla^2 \mathbf{B} = \mu_0 \text{curl } \mathbf{j}$$

Using London Eq. (2), we get

$$\begin{aligned} -\nabla^2 \mathbf{B} &= \mu_0 \left( -\frac{1}{\mu_0 \lambda_L^2} \mathbf{B} \right) \\ \therefore \quad \nabla^2 \mathbf{B} &= \mathbf{B}/\lambda_L^2 \end{aligned} \quad \dots(6)$$

Here,  $\lambda_L$  is called the *London penetration depth*.

For a superconductor to the right of the plane  $x = 0$ , Eq. (6) has the solution

$$B(x) = B(0) \exp(-x/\lambda_L), \quad \dots(7)$$

Here,  $B(0)$  is the field at the plane boundary.

The penetration depth  $\lambda_L$  is defined as the distance in which the field decreases by the factor  $e^{-1}$ .

Eq. (7) indicates that  $\mathbf{B}$  does not penetrate very deeply into a superconductor. Therefore, it implies the *Meissner effect*.

A magnetic field external to the superconductor decays exponentially as it penetrates the superconductor (Fig. 28.18). The decay constant is called the penetration depth  $\lambda_L$ . The vacuum-superconducting interface is at  $x = 0$ . Eq. (7) agrees well with the experimental observations that the magnetic field does not drop to zero abruptly at the surface but decays inside the superconductor in a characteristic distance  $\lambda_L$ , called *London penetration depth*. Thus the flux vanishes in the interior of the superconductor. The values of  $\lambda_L$  found experimentally range from 30 nm to 500 nm.

### Energy Gap

Experiments have shown that in superconductors, for temperatures in the vicinity of absolute zero, a forbidden energy gap just above the Fermi level is observed. Figure 28.19 (a) shows the conduction band in the normal state. Fig. 28.19 (b) depicts an energy gap equal to  $2E_g$  at the Fermi level in the superconducting state. Thus, the Fermi level in a superconductor is midway between the ground state and the first excited state so that each lies an energy distance  $= E_g$  away from the Fermi level. Electrons in excited states above the gap behave as normal electrons. At absolute zero, there are no electrons above the gap.  $E_g$  is typically of the order of  $10^{-4}$  eV.

$E_g$  is found to be a function of temperature  $T$ . Figure 28.20 shows reduced values of observed energy gap  $E_g(T)/E_g(0)$  as a function of the reduced temperature  $T/T_c$ . We observe that the energy gap decreases continuously to zero as the temperature is increased to  $T_c$ . The transition from the superconducting state to the normal state is observed to be a *second-order phase transition*. In such a transition, there is no latent heat, but there is a discontinuity in the heat capacity.

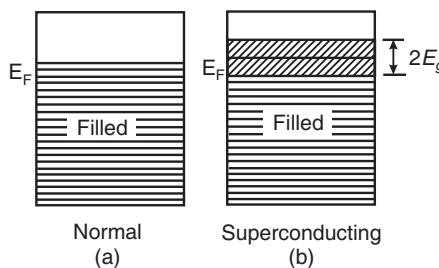


Fig. 28.19

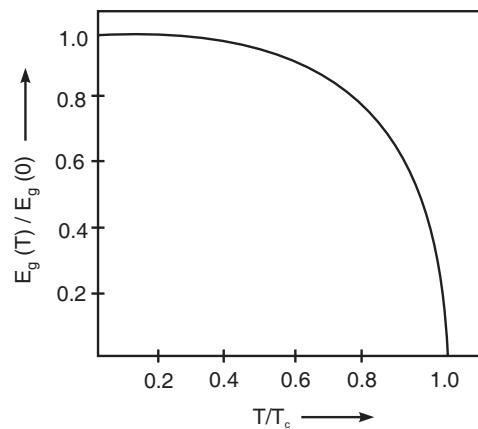


Fig. 28.20

### Microwave and Infrared Absorptions

The response of a metal to electromagnetic radiation is determined by the frequency dependent conductivity. This in turn depends on the available mechanisms for energy absorption by the conduction electrons at the given frequency. The electronic excitation spectrum in the superconducting state is characterised by an energy gap  $2E_g$ . So we expect the AC conductivity to differ substantially from its normal state form at frequencies small compared with  $2E_g/\hbar$  and to be essentially the same in the superconducting and normal states at frequencies large compared with  $2E_g/\hbar$ . The value of  $2E_g/\hbar$  is typically in the range between microwave and infrared frequencies. In the superconducting state, an AC behaviour is observed which is indistinguishable from that in

the normal state at optical frequencies. Deviations from normal state behaviour first appear in the infrared. At microwave frequencies *AC* behaviour fully displaying the lack of electronic absorption characteristic of an energy gap becomes completely developed.

## 28.6 HIGH- $T_c$ SUPERCONDUCTORS (HIGH-TEMPERATURE SUPERCONDUCTORS)

- Any superconductor with a transition temperature above 30 K is in general called high  $T_c$  superconductor.
- Till the year 1986, the highest known transition temperature  $T_c$  was 23 K in the  $\text{Nb}_3\text{Ge}$  alloy.
- Till the year 1986, the highest known transition temperature  $T_c$  was 23 K in the  $\text{Nb}_3\text{Ge}$  alloy.
- In 1986, Bednorz and Muller discovered the first superconductor of the high- $T_c$  family, viz.,  $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$  with  $T_c$  of  $\sim 35$  K.

### Examples of high temperature superconductors.

Table 28.1 gives some important high- $T_c$  superconductors and their  $T_c$  values.

**Table 28.1. High- $T_c$  superconductors**

S. No.	Material	Transition temperature $T_c$ in K
1.	$\text{La}_{1.85}\text{Ba}_{0.15}\text{CuO}_4$	36
2.	$\text{YBa}_2\text{Cu}_3\text{O}_7$	90
3.	$\text{Tl}_{12}\text{Ba}_2\text{Ca}_2\text{Cu}_3\text{O}_{10}$	125

### 1. Characteristics of high-temperature superconductors

- (1) They have high transition temperatures ( $T_c$ ), accompanied by high critical currents ( $i_c$ ), and high critical magnetic field ( $H_c$ ).
- (2) They have a modified perovskite crystal structure. The structure is highly *anisotropic*.
- (3) They are oxides of copper in combination with other elements. They have a layered structure consisting of one or more  $\text{CuO}_2$  layers.
- (4) The most important characteristic of high  $T_c$  superconducting oxides is their *metallic properties*. The conductivity is metallic mainly in the  $\text{CuO}_2$  planes; the conductivity is much smaller in a direction perpendicular to  $\text{CuO}_2$  planes.

### 2. Preparation of high $T_c$ ceramic superconductor- $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$

Let us see the preparation of yttrium barium copper oxide superconductor by the ‘shake and bake’ method. This method involves a four step process. 1. Mixing the chemicals. 2. Calcination (initial firing) 3. The intermediate firing (oxygen annealing) and 4. The final oxygen annealing.

- (1) **Mixing the chemicals.** The powders of yttrium oxide ( $\text{Y}_2\text{O}_3$ ), barium carbonate ( $\text{BaCO}_3$ ), and cupric oxide ( $\text{CuO}$ ) are mixed in the correct ratio.
- (2) **Calcination.** After the powders are thoroughly mixed, they are kept in an alumina crucible and heated in a furnace of about 950°C for about 18 hours. The following chemical reaction takes place.



This initial heat treatment is called calcination and results in a porous dark gray or black clump and the basic crystal structure is developed.

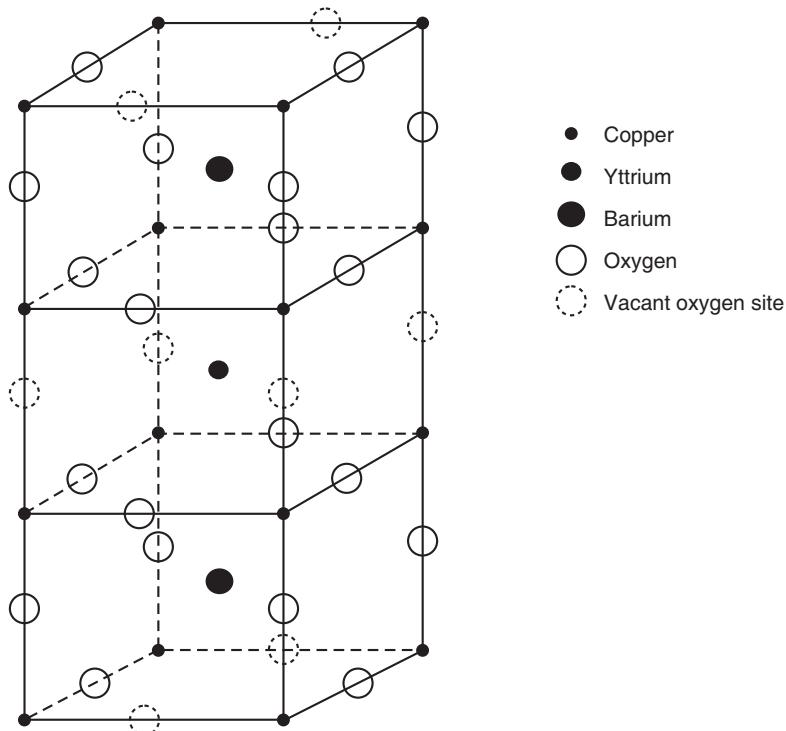
(3) **Intermediate firing.** Oxygen annealing treatment under oxygen flow is done (at temperatures between 950°C and 400°C) to produce  $\text{YBa}_2\text{Cu}_3\text{O}_7$ .

(4) **The final oxygen annealing.** This is done with adequate oxygen flow.

- The crystal structure of the powder product obtained is related to the cubic perovskite structure illustrated in Fig. 28.21.

#### Crystal structure of yttrium barium copper oxide ( $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ )

Figure 28.21 shows the crystal structure of yttrium-barium-copper oxide.



**Fig. 28.21**

Three body-centred cubic unit cells are stacked one above another.

The atom distributions in the unit cells are as follows:

$$\text{Copper atom at the body corners: } 8 \times \frac{1}{8} \times 3 = 3$$

$$\text{Barium atom at two body centres : } 1 \times 2 = 2$$

$$\text{Yttrium atom at one body centre : } 1 \times 1 = 1$$

$$\text{Oxygen atom at mid-points of edges : } 12 \times \frac{1}{4} \times 3 = 9$$

Some oxygen positions are vacant such that their number is between 6.5 and 7. The superconducting properties appear to be a sensitive function of the oxygen content and, therefore, of the partial pressure of oxygen during heat treatment.

## 28.7 APPLICATIONS OF SUPERCONDUCTORS-SQUID, CRYOTRON, MAGNETIC LEVITATION

### 1. SQUID (Superconducting Quantum Interference Device)

**Principle.** A DC magnetic field applied through a superconducting ring containing two Josephson junctions causes the maximum supercurrent to show interference effects as a function of magnetic field intensity. This effect is used in SQUID.

- SQUID is a sensitive magnetometer used to detect and measure weak magnetic fields. This forms an important application of Josephson junctions.

**Description and working.** SQUID is a double junction quantum interferometer. Two Josephson junctions mounted on a superconducting ring form this interferometer (Fig. 28.22).

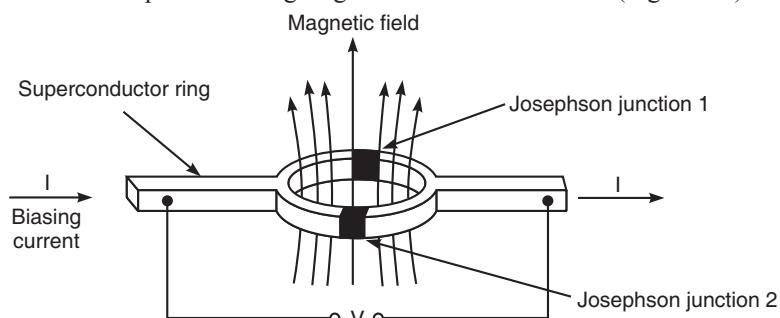


Fig. 28.22

The SQUID is based on the flux quantization in a superconducting ring. The total magnetic flux passing through the ring is quantized.

- *A small change in magnetic field produces variation in the quantum flux.* When the magnetic field is applied perpendicular to the plane of the ring, current is induced at the two Josephson junctions. The induced current produces the interference pattern and it flows around the ring. The total magnetic flux passing through the ring is quantized.

#### Applications of SQUID

- (1) Very minute magnetic signals are detected by the SQUID sensor.
- (2) SQUID is used as storage device for magnetic field.
- (3) The SQUID is used to study tiny magnetic signals from the brain and heart.
- (4) SQUID magnetometer can detect the paramagnetic response in the liver and give the amount of iron held in the liver of the body accurately.

### 2. Cryotron

- *Cryotron is a magnetically operated current switch.*

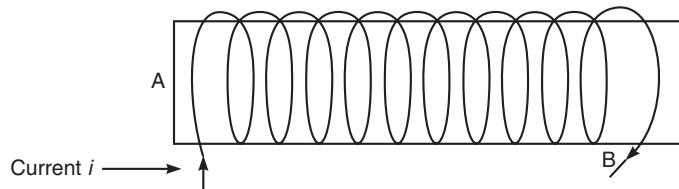
**Principle.** The superconducting property of a material disappears when the applied magnetic field is greater than the critical magnetic field ( $H_c$ ).

A superconductor possesses two states, the superconducting and normal. The application of a magnetic field greater than  $H_c$  can initiate a change from superconducting to normal and removal of the field reverses the process. This principle is applied in development of switching element *cryotron*.

**Description.** A superconducting material *A* is surrounded by another superconducting material *B* (Fig. 28.23). The critical magnetic field ( $H_{c_A}$ ) of material *A* is less than the critical magnetic field ( $H_{c_B}$ ) of material *B*.

$$H_{C_A} < H_{C_B}.$$

Let the temperature of the system be below the transition temperatures of the two materials.



**Fig. 28.23**

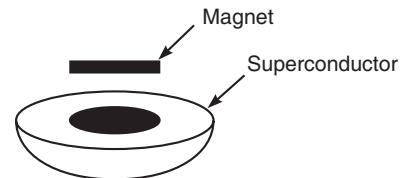
**Working.** A current  $i$  is passed through the material  $B$ .

When the magnetic field produced by material  $B$  exceeds the critical magnetic field of  $A$ , the material  $A$  becomes a normal conductor.

But,  $B$  will not become a normal conductor at the critical magnetic field of  $A$  because  $H_{C_B} > H_{C_A}$ . Thus the current in material  $A$  can be controlled by the current in material  $B$ . Hence, this system can act as a relay or switching element.

### 3. Magnetic levitation.

**Principle.** The diamagnetic property of a superconductor is the basis of magnetic levitation. When a material is superconducting, the magnetic flux lines will be expelled from the material (*Meissner effect*). So superconducting materials strongly repel external magnets.



**Fig. 28.24**

If a small light and powerful magnet is kept over a superconducting material, the magnet will be levitated (lifted up) and it will float in air (Fig. 28.24). This is called *magnetic levitation*.

**Demonstration.** Magnetic levitation can be demonstrated using a high  $T_c$  superconductor like  $\text{Y}_1\text{Ba}_2\text{Cu}_3\text{O}_7$  compound taken in bulk and a light weight and powerful magnet (samarium-cobalt magnet). The Y-123 compound is cooled by pouring liquid nitrogen over it. The magnet kept over it just floats in air.

#### Use of magnetic levitation

*Magnetic levitation* effect can be used for high speed transportation such as super fast trains, without frictional loss.

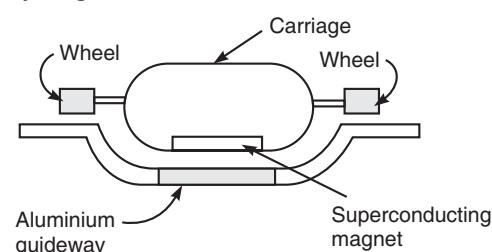
### 4. Magnetic levitated train (MAGLEV train).

- Maglev is the acronym of magnetic levitation. This train has superconducting magnet built into its base (Fig. 28.25).

The guideway for the maglev is similar to the rail tracks of the conventional railways. The train will be set afloat above the aluminium guideway by magnetic levitation.

Magnetic levitation is brought about by enormous repulsion between two highly powerful magnetic fields, one produced by the superconducting magnet inside the train and the other by the electric current in the aluminium guideway.

The current in the guideway produces the necessary magnetic field to levitate the train and helps in propelling the train forward.



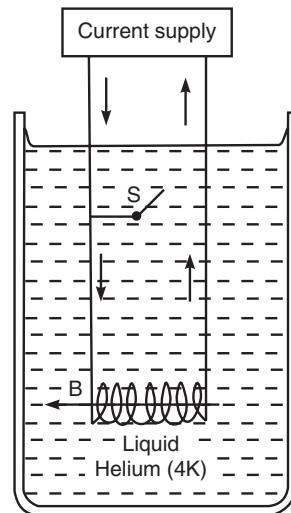
**Fig. 28.25**

### 5. Superconducting electromagnets

Superconductors are used in the construction of superconducting electromagnets that carry large resistanceless currents and, therefore, produce large magnetic fields ( $\approx 5 - 10$  Tesla). If we use superconducting wire for magnet windings, currents of the order of 100 amperes can be carried by very fine wires. Thus small-size magnets can be constructed. The principle is illustrated in Fig. 28.26, in which a superconducting coil is immersed in liquid helium (4 K). Starting from zero, the current is increased until the magnetic field  $B$  reaches the desired value. At that point the switch  $S$  is closed. The current now flows through the switch. The power supply is then turned off. The current will continue to flow through the coil indefinitely without resistive losses.

### 6. Superconducting Cables

Superconducting cables can be used to transmit electric power over long distances without resistive losses. This is possible only if the cost of keeping the cable below its critical temperature is less than the value of the power lost using ordinary cables.



**Fig. 28.26**

### EXERCISE

1. Phenomenon of superconductivity was discovered by  
(a) K. Onnes      (b) Meissner      (c) Silsbee      (d) Josephson  
*(B.U. 2011)*
2. The transition temperature of mercury is  
(a) 1 K      (b) 1.14 K      (c) 4.12 K      (d) 9.22 K  
*(B.U. 2011)*
3. Transition temperature  $T_c$  and critical field  $H_c$  for a superconductor are related as  
(a)  $H_c = H_0(T_c - 1)$       (b)  $H_c = H_0(T_c + 1)$   
(c)  $H_c = H_0 \left[ 1 - \left( \frac{T}{T_c} \right)^2 \right]$       (d)  $T_c = T_0 \left[ 1 - \left( \frac{H_0}{H_c} \right)^2 \right]$   
*(B.U. 2010)*
4. Which type magnetic material is suited for superconductivity?  
(a) paramagnetic      (b) diamagnetic      (c) ferromagnetic      (d) ferrimagnetic  
*(B.U. 2015)*
5. A relation connecting isotopic mass 'M' of a superconductor with critical temperature ' $T_c$ ' is given by  
(a)  $M = kT_c$       (b)  $M^{1/2}T_c = \text{a constant}$   
(c)  $MT_c^{1/2} = \text{a constant}$       (d)  $MT_c^2 = \text{a constant}$   
*(B.U. M.Sc. 2006)*  
[k-Boltzman's constant]
6. Which of the following is Type I Superconductor?  
(a) Vanadium      (b) Gold      (c) Niobium      (d) Lead  
*(B.U. 2012)*  
[Ans. 1. (a), 2.(c), 3.(c), 4.(b), 5.(b), 6.(d)]

7. Explain the various properties of superconducting materials. (B.U. 2011)
8. Explain the effect of magnetic field in superconductor. (B.U. 2011)
9. Derive the thermodynamical relation connecting the critical magnetic field and the transition temperature of a superconductor. (B.U. 2010)
10. The critical field for vanadium is  $10^5$  A/m at 8 K and  $2 \times 10^5$  A/m at 0 K. Calculate the transition temperature of the element. (B.U. 2011)
11. What is Meissner effect? Obtain an expression for the London penetration depth of magnetic field for a superconductor. (Raj. 1976)
12. Explain the Meissner effect in superconductivity. (B.U. 2015)
13. State and explain isotope effect in superconductors. (B.U. 2013)
14. Discuss the thermodynamic effect of superconductors. (B.U. 2015)
15. Write short notes on Type I and Type II superconductors. (B.U. 2012)
16. Discuss the formation of Cooper pairs. (Madras 2006)
17. What are Cooper pairs? Give qualitative ideas of BCS theory of superconductivity.
18. Derive London equation. (Madras 2006)
19. Explain the practical uses of superconductivity. (Madras 2006)

## DIELECTRICS AND FERROELECTRICS

**After reading this chapter, you should be able to**

- ◆ Discuss different types of polarizability
- ◆ Discuss and explain Clausius-Mossotti relation
- ◆ List properties and applications of different types of insulating materials
- ◆ Explain the phenomenon of ferroelectricity
- ◆ Name any two ferroelectric materials and their applications

## 29.1 INTRODUCTION

**Dielectric.** A dielectric (or an insulator) is a material in which all the electrons are tightly bound to the nuclei of the atoms. Thus there are no free electrons to carry current. Hence the electrical conductivity of a dielectric is very low. (The conductivity of an ideal dielectric is zero). Glass, plastic, mica, oil, etc., are examples of dielectrics.

**Nonpolar and Polar Molecules**—The molecules of dielectrics may be classified as ‘nonpolar’ and ‘polar’.

A **non-polar molecule** is one in which the centre of gravity of the positive charges (protons) coincides with the centre of gravity of the negative charges (electrons). The non-polar molecules have symmetrical structure and zero electric dipole moment. Examples are:  $\text{H}_2$ ,  $\text{N}_2$ ,  $\text{O}_2$ ,  $\text{CO}_2$ , benzene, etc. (Fig. 29.1).

A polar molecule is one in which the centre of gravity of the positive charges is separated from the centre of gravity of the negative charges by a finite distance. The polar molecule is thus an electric dipole and has an intrinsic permanent dipole moment. Examples are:  $\text{H}_2\text{O}$ ,  $\text{HCl}$ ,  $\text{CO}$ ,  $\text{N}_2\text{O}$ ,  $\text{NH}_3$ , etc., (Fig. 29.2). The permanent dipole moment of a molecule of water has the magnitude  $p = 6.3 \times 10^{-30} \text{ Cm}$  and is directed from  $\text{O}^-$  ion toward the midpoint of the line connecting the  $\text{H}^+$  ions.

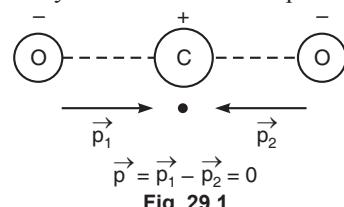


Fig. 29.1

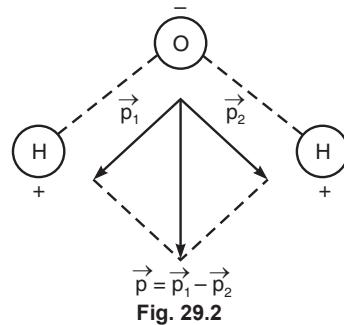


Fig. 29.2

### Fundamental definitions in dielectrics

#### 1. Dielectric constant ( $\epsilon_r$ )

Dielectric constant is the ratio between the absolute permittivity of the medium ( $\epsilon$ ) and the permittivity of free space ( $\epsilon_0$ )

$$\text{Dielectric constant } (\epsilon_r) = \frac{\text{Absolute permittivity of the medium } (\epsilon)}{\text{Permittivity of free space } (\epsilon_0)}$$

#### 2. Electric polarisation

The process of producing electric dipoles by an electric field is called polarisation in dielectrics.

When an electric field is applied to a crystal or a glass containing positive and negative charges, the positive charges are displaced in the direction of the field, while the negative charges are displaced in the opposite direction. This displacement produces local dipoles throughout the solid.

**Electric dipole.** A system consisting of two equal and opposite charges  $+q$  and  $-q$  separated by a distance  $d$  is called an *electric dipole* (Fig. 29.3).

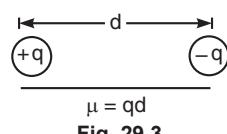


Fig. 29.3

**Electric dipole moment ( $\mu$ ).** The product of the magnitude of the charge ( $q$ ) and distance between two charges ( $d$ ) is called *electric dipole moment* ( $\mu$ ).

$$\mu = qd.$$

Unit of dipole moment is **coulomb-metre (Cm)**.

The dipole moment is a vector. It points from the negative charge to the positive charge.

**3. Polarizability ( $\alpha$ ).** The strength of the induced electric dipole moment ( $\mu$ ) in dielectrics is proportional to the applied electric field ( $E$ ).

$$\mu \propto E$$

or

$$\mu = \alpha E.$$

Here,  $\alpha$  is the constant of proportionality, called the **polarisability**.

$$\alpha = \frac{\mu}{E}$$

*Atomic polarisability ( $\alpha$ ) is defined as the electric dipole moment induced in the atom by an electric field of unit strength. Molecular polarisability ( $\alpha$ ) is defined as the electric dipole moment induced in the molecule by an electric field of unit strength.*

**Definition.** Polarisability is defined as the ratio of induced dipole moment to the electric field applied.

- Polarisability is defined as induced dipole moment per unit electric field.

For a non-spherical atom  $\alpha$  is a tensor.

### 3. Polarisation vector $P$

The polarisation vector  $P$  is the dipole moment per unit volume of the dielectric material.

$$P = N\bar{\mu}$$

Here,

$N$  = number of molecules per unit volume and

$\bar{\mu}$  = average dipole moment per molecule.

### 4. Electric displacement vector $D$

The electric displacement vector  $D$  is given by

$$D = \epsilon_0 \epsilon_r E \quad \dots(1)$$

Here,

$E$  = electric field strength,

$\epsilon_0$  = permittivity of free space (i.e., vacuum),

$\epsilon_r$  = relative permittivity of the material.

Relation between  $D$  and  $P$  is

$$D = \epsilon_0 E + P \quad \dots(2)$$

$$\epsilon_0 \epsilon_r E = \epsilon_0 E + P$$

∴

$$P = \epsilon_0 (\epsilon_r - 1) E$$

### 5. Electric susceptibility ( $\chi_e$ )

When a dielectric material is placed in an electric field, it becomes electrically polarised. For isotropic dielectrics, the polarisation vector  $P$  is proportional to the electric field  $E$ . Thus,

$$P = \epsilon_0 \chi_e E$$

The constant  $\chi_e$  is called the ‘electric susceptibility’ of the dielectric material.

$$\chi_e = \frac{P}{\epsilon_0 E} = \frac{\epsilon_0 (\epsilon_r - 1) E}{\epsilon_0 E} = \epsilon_r - 1$$

∴

$$\chi_e = (\epsilon_r - 1)$$

## 29.2 DIFFERENT TYPES OF ELECTRIC POLARISATION

### 1. Electronic polarisation

The displacement of the positively charged nucleus and the (negative) electrons of an atom in opposite directions on the application of an electric field results in *electronic polarisation*.

Fig. 29.4 shows the electronic polarisation occurring in an atom due to the applied electric field. On applying a field, the electron cloud around the nucleus shifts towards the positive end of the field. As the nucleus and the centre

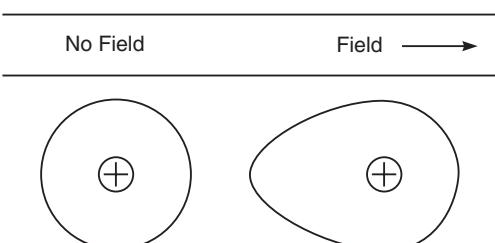


Fig. 29.4

of the electron cloud are separated by a certain distance, dipole moment is created within each atom.

The induced dipole moment ( $\mu$ ) is proportional to the field strength ( $E$ ).

$$\mu \propto E$$

or

$$\mu = \alpha_e E$$

The constant of proportionality  $\alpha_e$  is the electronic polarisability of the atom. The electronic polarisability  $\alpha_e$  is proportional to the volume of the atom. It is independent of temperature.

#### Calculation of electronic polarizability ( $\alpha_e$ ).

##### (i) No electric field

Consider an atom of a dielectric material of atomic number  $Z$  [Fig. 29.5 (a)].  $Ze$  is the nuclear charge. The electrons of charge ( $-Ze$ ) are distributed uniformly throughout the sphere of radius  $R$ . When there is no external electric field, the centres of the electron cloud and nucleus of the atom are same and one.

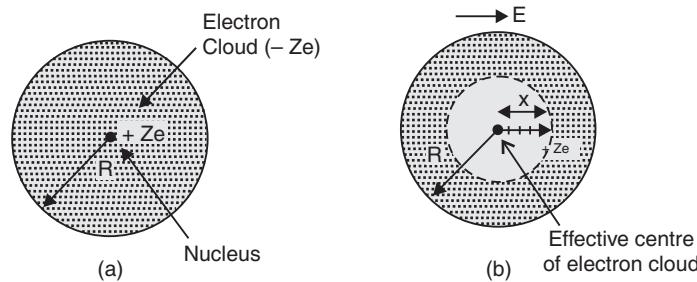


Fig. 29.5

##### (ii) With electric field

When a d.c. electric field  $E$  is applied, the nucleus moves towards the field direction and the electron cloud moves in the opposite direction of the field [Fig. 29.5 (b)]. The electron cloud centre gets a displacement  $x$  with respect to the centre of the nucleus. The displacement  $x \ll R$ . We assume that the spherical shape of the electron cloud is not altered on the application of the electric field.

When the nucleus and electron cloud are shifted from their equilibrium position, an attractive coulomb force is created and it will tend to maintain the original equilibrium position. But the Lorentz force will tend to separate the nucleus and electron cloud of that atom from their equilibrium position. Finally a new equilibrium will be reached when these two forces are equal and opposite.

At equilibrium, *Lorentz force* = *Coulomb force*.

$$\text{The Lorentz force between nucleus and the electron cloud} = -Ze \times E = -ZeE \quad \dots(1)$$

$$\text{The charge density of the charged sphere} = \frac{-Ze}{\frac{4}{3}\pi R^3}$$

$$\left. \begin{array}{l} \text{Total number of negative charges} \\ \text{in the sphere of radius } x \end{array} \right\} = \frac{-Ze}{\frac{4}{3}\pi R^3} \times \frac{4}{3}\pi x^3 = -Ze \frac{x^3}{R^3} \quad \dots(2)$$

$$\left. \begin{array}{l} \text{The coulomb force between the} \\ \text{nucleus and the electron cloud} \end{array} \right\} = \frac{\frac{-Ze x^3}{R^3} (Ze)}{4\pi \epsilon_0 x^2} = \frac{-Z^2 e^2 x}{4\pi \epsilon_0 R^3}$$

Under equilibrium conditions, the coulomb attractive force and the Lorentz repulsive force are equal and opposite. Thus

$$ZeE = \frac{Z^2 e^2 x}{4\pi \epsilon_0 R^3}$$

or  $E = \frac{Zex}{4\pi \epsilon_0 R^3}$  ... (3)

Thus, the displacement of the electron cloud ( $x$ ) is proportional to the applied electric field  $E$ .

$$\text{Induced dipole moment} = \mu = Zex \quad \dots (4)$$

$$\text{By definition} \quad \mu = \alpha_e E \quad \dots (5)$$

$$\text{From Eqs. (4) and (5), } E = \frac{Zex}{\alpha_e} \quad \dots (6)$$

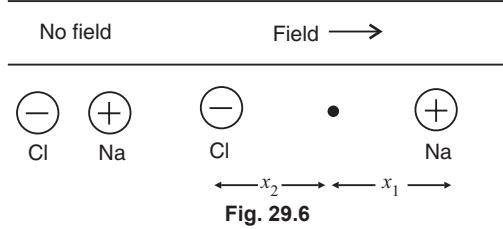
From Eqs. (3) and (6), we get

$$\begin{aligned} \frac{Zex}{\alpha_e} &= \frac{Zex}{4\pi \epsilon_0 R^3} \\ \therefore \alpha_e &= 4\pi \epsilon_0 R^3 \end{aligned} \quad \dots (7)$$

## 2. Ionic polarisation

The ionic polarisation is due to the displacement of cations and anions in opposite directions and occurs in an ionic solid (Fig. 29.6). The displacement of ions is independent of temperature.

Suppose an electric field is applied in the +ve  $x$  direction.



Then the + ve ions move to the right by  $x_1$  and – ve ions move to the left by  $x_2$ .

$m$  is the mass of the + ve ion.

$M$  is the mass of the – ve ion.

We assume that there are one cation and one anion in each unit cell of the ionic crystal.

Due to ionic displacement the resultant dipole moment per unit cell,

$$\mu = e(x_1 + x_2) \quad \dots (1)$$

Here,  $x_1$  = shift of + ve ion and  $x_2$  = shift of – ve ion with respect to their equilibrium positions.

When the electric field is applied, the restoring force ( $F$ ) produced is proportional to the displacement ( $x$ ).

$$\text{Restoring force on + ve ion, } F = \beta_1 x_1 \quad \dots (2)$$

$$\text{Restoring force on – ve ion, } F = \beta_2 x_2 \quad \dots (3)$$

Here,  $\beta_1$  and  $\beta_2$  are restoring force constants. They depend upon the masses of ions and angular frequency of the molecule in which ions are present. Therefore, under equilibrium,

$$\begin{aligned} F &= \beta_1 x_1 = \beta_2 x_2 \\ x_1 &= \frac{F}{\beta_1} = \frac{eE}{m\omega_0^2} \quad \left[ \because F = eE \right] \\ &\quad \left[ \beta_1 = m\omega_0^2 \right] \end{aligned}$$

Here,  $\omega_0$  = natural frequency of the ionic molecule.

Similarly for – ve ion,

$$x_2 = \frac{F}{\beta_2} = \frac{eE}{M\omega_0^2}$$

$$\therefore x_1 + x_2 = \frac{eE}{\omega_0^2} \left( \frac{1}{m} + \frac{1}{M} \right) \quad \dots(4)$$

$$\mu = e(x_1 + x_2) = \frac{e^2 E}{\omega_0^2} \left( \frac{1}{m} + \frac{1}{M} \right)$$

$$\therefore \alpha_i = \frac{\mu}{E} = \frac{e^2}{\omega_0^2} \left( \frac{1}{m} + \frac{1}{M} \right) \quad \dots(5)$$

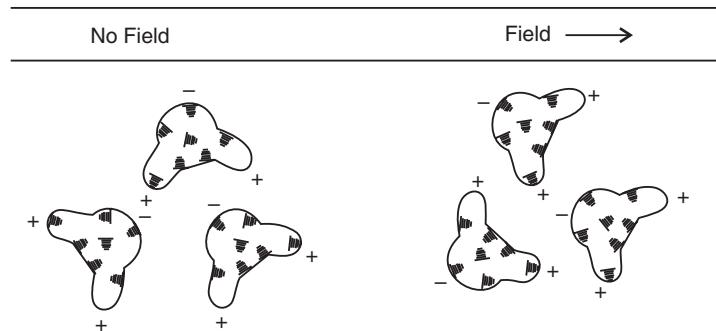
Thus the ionic polarisability ' $\alpha_i$ ' is inversely proportional to the square of the natural frequency of the ionic molecule and to its reduced mass. It is independent of temperature.

### 3. Orientational Polarisation

In methane molecule ( $\text{CH}_4$ ), the centre of the negative and positive charges coincide. It has no permanent dipole moment.

However in  $\text{CH}_3\text{Cl}$ , the positive and negative charges do not coincide. Here chlorine attracts the bonding electrons to itself more strongly than hydrogen. Therefore, even in the absence of an electric field, this molecule has dipole moment. When an electric field is applied, these molecules tend to align themselves in the applied field as shown in Fig. 29.7.

Hence the electric dipole moment increases. The polarisation due to such alignment is called *orientational polarisation*.



**Fig. 29.7**

Orientational polarisation can be shown to be

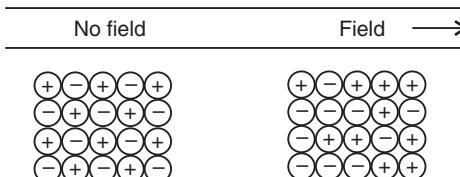
$$P_0 = N \bar{\mu} = N \mu^2 E / 3kT = N\alpha_0 E.$$

$$\therefore \text{orientational polarisability, } \alpha_0 = \frac{P_0}{NE} = \frac{\mu^2}{3kT}.$$

Thus the orientational polarisability  $\alpha_0$  is inversely proportional to absolute temperature of the material.

### 4. Space -charge polarisation

Space-charge polarisation occurs due to the accumulation of charges at the electrodes or at the interfaces in multiphase dielectrics (Fig. 29.8). It is found in ferrites and semiconductors. The ions diffuse over appreciable distances in response to the applied field, giving rise to a redistribution of charges in the dielectric medium.



**Fig. 29.8**

### Total Polarisation

Since the value of space-charge polarisability is very small when compared with other types of polarisabilities, it can be neglected.

The total polarisability of a gas ' $\alpha$ ' can be written as

$$\begin{aligned}\alpha &= \alpha_e + \alpha_i + \alpha_0 \\ &= 4\pi \epsilon_0 R^3 + \frac{e^2}{\omega_0^2} \left( \frac{1}{M} + \frac{1}{m} \right) + \frac{\mu^2}{3kT}\end{aligned}$$

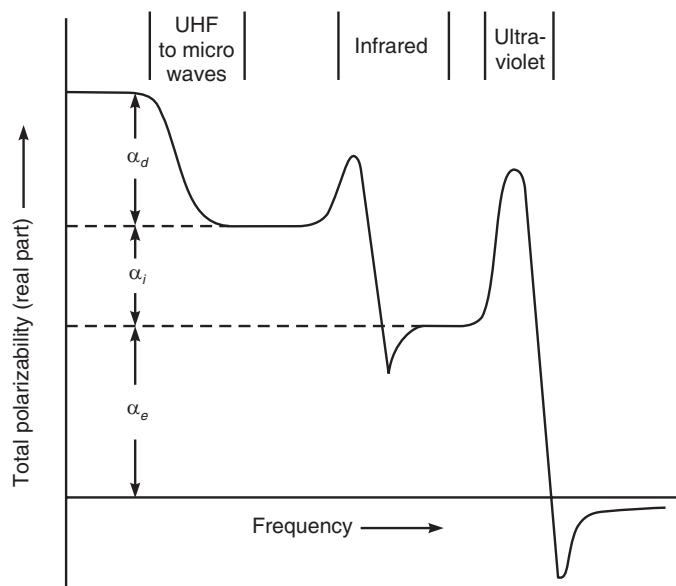
The total polarisation  $P$  can be written as

$$P = N \alpha E = NE \left\{ 4\pi \epsilon_0 R^3 + \frac{e^2}{\omega_0^2} \left( \frac{1}{M} + \frac{1}{m} \right) + \frac{\mu^2}{3kT} \right\}$$

This equation is known as *Langevin - Debye equation*.

### Frequency Dependence of Total Polarizability

The net polarizability of a dielectric material results from three main contributions—the electronic polarizability, ionic polarizability, and dipolar or orientational polarizability. The electronic contribution arises from the displacement of the electron shell relative to a nucleus. The ionic contribution comes from the displacement of a charged ion with respect to other ions. The dipolar polarizability arises from molecules with a permanent electric dipole moment that can change orientation in an applied electric field.



**Fig. 29.9**

The total polarizability of a dielectric is given by the expression:

$$\alpha = \alpha_e + \alpha_i + \alpha_d \quad \dots(1)$$

Figure 29.9 depicts the frequency dependence of the several contributions to the polarizability.

### Effect of Temperature on Polarisation

1. The electronic and the ionic polarisabilities are independent of temperature.

2. The effect of increase of temperature brings about a higher degree of randomness in the molecular orientation in the material. This affects the tendency of permanent dipoles to align along the field direction. Hence, *orientation polarisation decreases with the increase in temperature*.

3. In space charge polarisation, increase in temperature facilitates the diffusion of ions. Thermal energy may also aid in overcoming the activation barrier for the orientation of polar molecules in the direction of the field. Hence, *space charge polarisation increases with increase of temperature*.

### 29.3 DERIVATION OF CLAUSIUS-MOSSOTTI EQUATION

**Lorentz method for finding the internal field for a cubic structure.** Consider a slab of a dielectric placed in the uniform electric field between two parallel plates of a capacitor (Fig. 29.10)

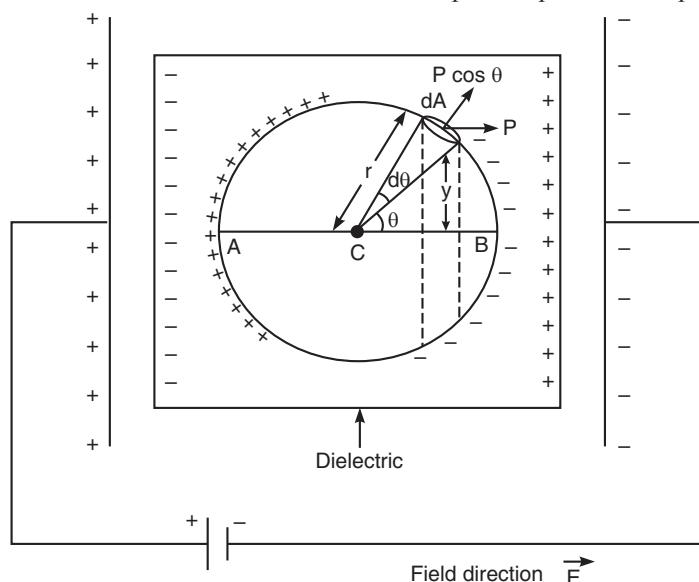


Fig. 29.10

The dielectric material is uniformly polarised.

Let us find the internal field acting on an atom at C.

Consider an imaginary small spherical cavity of radius  $r$  around the atom.

The internal field ( $E_i$ ) at the atom site C is made up of the following four components, namely,  $E_1, E_2, E_3$  and  $E_4$ .

$$E_i = E_1 + E_2 + E_3 + E_4 \quad \dots (1)$$

(i)  $E_1$  = Electric field at point C due to charges on the plates of the capacitor in the absence of the dielectric.

$E_1$  is the field intensity at C due to the charge density on the plates. From the field theory.

$$E_1 = D/\epsilon_0$$

$$D = P + \epsilon_0 E$$

$$E_1 = \frac{P + \epsilon_0 E}{\epsilon_0} = E + \frac{P}{\epsilon_0}$$

$$\therefore E_1 = E + \frac{P}{\epsilon_0} \quad \dots (2)$$

(ii)  $E_2$  = Electric field intensity at C due to the charge density induced on the two sides of the dielectric.

$$E_2 = -\frac{P}{\epsilon_0} \quad \dots (3)$$

(iii)  $E_3$  = Electric field intensity at C due other atoms contained in the cavity.

We are assuming a cubic structure. The lattice has cubic symmetry.

$$\therefore E_3 = 0 \quad \dots (4)$$

(iv)  $E_4$  = Electric field produced by the polarisation charges on the cavity surface.

#### Calculation of $E_4$

Consider an elementary ring of area  $dA$  lying between  $\theta$  and  $\theta + d\theta$  from the polarization direction ( $\vec{P}$ ). The polarization ( $\vec{P}$ ) is along the applied external field ( $\vec{E}$ ).

The radius of the elementary ring =  $y = r \sin \theta$  (Fig. 29.11).

Surface area of the elementary ring,

$$dA = 2\pi y \times \text{thickness of the ring} = 2\pi y \times rd\theta = 2\pi(r \sin \theta)(r d\theta)$$

$$\therefore dA = 2\pi r^2 \sin \theta d\theta$$

Components of ( $\vec{P}$ ) normal to  $dA = P \cos \theta$

The charge  $dq$  on the elementary area  $dA$  is given by the product of the normal component of the polarization and the area.

$$dq = (P \cos \theta)dA = P(2\pi r^2 \sin \theta \cos \theta d\theta)$$

Since only the  $\cos \theta^\circ$  component of  $dE_4$  is effective, we can write,

$$dE_4 = \frac{dq \times \cos \theta}{4\pi\epsilon_0 r^2} = \frac{P(2\pi r^2 \sin \theta \cos \theta d\theta) \times \cos \theta}{4\pi\epsilon_0 r^2}$$

$$dE_4 = \frac{P}{2\epsilon_0} \cos^2 \sin \theta d\theta$$

Thus the total field  $E_4$  due to the charges on the surface of the entire cavity is

$$\begin{aligned} E_4 &= \frac{P}{2\epsilon_0} \int_0^\pi \cos^2 \theta \sin \theta d\theta \\ &= \frac{P}{2\epsilon_0} \int_0^\pi \cos^2 \theta d(-\cos \theta) \\ &= \frac{-P}{2\epsilon_0} \left[ \frac{\cos^3 \theta}{3} \right]_0^\pi \\ &= -\frac{P}{6\epsilon_0} [-1 - 1] = \frac{P}{3\epsilon_0} \\ \therefore E_4 &= \frac{P}{3\epsilon_0} \quad \dots (5) \end{aligned}$$

Now

$$E_i = E_1 + E_2 + E_3 + E_4$$

Substituting the Values of  $E_1$ ,  $E_2$ ,  $E_3$  and  $E_4$  from Eqs. (2), (3), (4) and (5),

$$E_i = \left( E + \frac{P}{\epsilon_0} \right) - \frac{P}{\epsilon_0} + 0 + \frac{P}{3\epsilon_0}$$

$$E_i = E + \frac{P}{3\epsilon_0} \quad \dots (6)$$

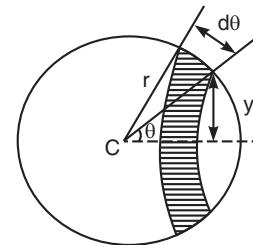


Fig. 29.11

- Eq. (6) is the expression for the local field in a dielectric for a cubic structure.
- Eq. (6) is the expression for the internal field in a dielectric solid material.
- Eq. (6) is the expression for the Lorentz field for elemental dielectrics.

### Derivation of the Clausius-Mossotti Relation

Let  $N$  be the number of molecules per unit volume and  $\alpha$  be the molecular polarisability.

$$\text{Total polarisation, } P = N\alpha E_i \quad \dots (7)$$

Substituting the value of  $E_i$  from Eq. (6) in Eq. (7), we get

$$\begin{aligned} P &= N\alpha \left( E + \frac{P}{3\epsilon_0} \right) \\ P \left\{ 1 - \frac{N\alpha}{3\epsilon_0} \right\} &= N\alpha E \\ P &= \frac{N\alpha E}{\left( 1 - \frac{N\alpha}{3\epsilon_0} \right)} \end{aligned} \quad \dots (8)$$

We know that

$$\begin{aligned} D &= P + \epsilon_0 E \\ P &= D - \epsilon_0 E \\ \frac{P}{E} &= \frac{D}{E} - \epsilon_0 = \epsilon - \epsilon_0 = \epsilon_0 \epsilon_r - \epsilon_0 = \epsilon_0 (\epsilon_r - 1) \end{aligned}$$

$$\therefore P = E \epsilon_0 (\epsilon_r - 1) \quad \dots (9)$$

From equations (8) and (9), we get

$$\begin{aligned} E \epsilon_0 (\epsilon_r - 1) &= \frac{N\alpha E}{1 - \frac{N\alpha}{3\epsilon_0}} \\ 1 - \frac{N\alpha}{3\epsilon_0} &= \frac{N\alpha}{\epsilon_0 (\epsilon_r - 1)} \\ 1 &= \frac{N\alpha}{3\epsilon_0} \left( 1 + \frac{3}{\epsilon_r - 1} \right) \\ \frac{N\alpha}{3\epsilon_0} &= \frac{1}{1 + \left( \frac{3}{\epsilon_r - 1} \right)} = \frac{\epsilon_r - 1}{\epsilon_r + 2} \\ \therefore \left( \frac{\epsilon_r - 1}{\epsilon_r + 2} \right) &= \frac{N\alpha}{3\epsilon_0} \end{aligned} \quad \dots (10)$$

Equation (10) is the *Clausius Mossotti relation*.

- It relates the dielectric constant to the molecular polarisability.
- It relates the dielectric constant to the atomic polarizability provided the condition of cubic symmetry holds.

### 29.4 DIELECTRIC LOSS

*When a dielectric is subjected to the a.c. voltage, the electrical energy is absorbed by the material and is dissipated in the form of heat. This dissipation of electrical energy in the dielectric is called dielectric loss.*

Since this involves heat generation and heat dissipation, this assumes a dominating role in high voltage applications.

#### Expression for dielectric loss (or loss tangent)

(i) When a.c. voltage is applied to perfect insulator like vacuum or purified gas, there is no consumption of energy and the charging current leads the applied voltage by  $90^\circ$  [Fig. 29.12 (b)].

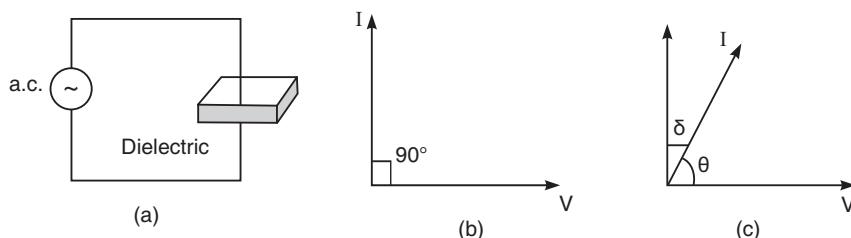


Fig. 29.12

$$\text{Power loss, } P = VI \cos\theta = VI \cos 90^\circ = 0.$$

Thus the ideal dielectrics have no dielectric loss.

(ii) The phase angle between charging current and applied voltage is less than  $90^\circ$  in commercial dielectrics. Complementary angle  $\delta = 90^\circ - \theta$  is called the *dielectric loss angle* [Fig. 29.12(c)]. The loss angle is a measure of the power dissipated in each cycle.

For a dielectric having a capacitance of  $C$  and having a voltage  $V$  applied to it at a frequency  $f$  Hz, the dielectric power loss is given by

$$P = VI \cos\theta$$

$$\text{Since } I = \frac{V}{X_c} \text{ where } X_c = \frac{1}{2\pi f C} = \text{capacitive reactance to A.C.}$$

$$\text{and } \cos\theta = \cos(90^\circ - \delta) = \sin\delta,$$

$$P = \frac{V^2}{X_c} \sin\delta = 2\pi f C V^2 \sin\delta$$

Since  $\delta$  is very small,  $\sin\delta = \tan\delta$ .

$$\therefore \text{Dielectric power loss, } P = V^2 2\pi f C \tan\delta$$

Here  $\tan\delta$  is called *power factor* of the dielectric. The power loss is dependent on  $\tan\delta$  as long as the other factors like voltage, frequency and capacitance are constants. Thus the power loss is found to depend on the frequency and value of applied voltage.

#### 29.5 DETERMINATION OF DIELECTRIC CONSTANT OF A DIELECTRIC MATERIAL

The Schering bridge is used for measuring the dielectric constant of dielectric materials.

**Principle.** It is based on measurement of the capacitance of the capacitor without the dielectric material ( $C_1$ ) and the capacitance of the capacitor with the dielectric material ( $C'_1$ ).

The connections are shown in Fig. 29.13.

$C_1$  = Capacitor whose capacitance is to be measured

$r_1$  = A series resistance representing the loss in the capacitor  $C_1$

$C_2$  = Standard capacitor

$R_3$  = Non-inductive variable resistance

$R_4$  = Non-inductive resistance;

$C_4$  = Variable capacitor

$D$  = A.C. null detector;  
 $S$  = High frequency oscillator.

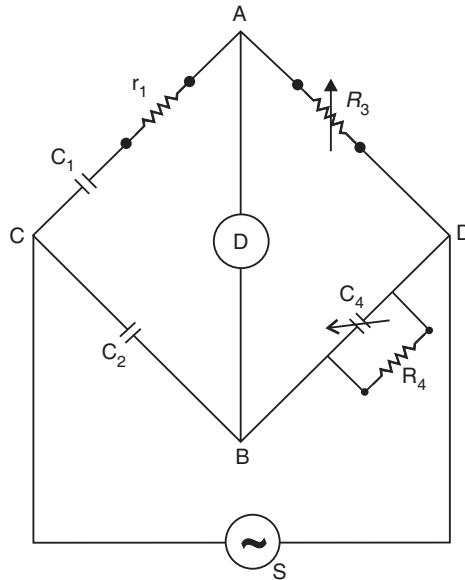


Fig. 29.13

(i) First without inserting dielectric inside  $C_1$ , the bridge is balanced by varying  $C_4$  and  $R_3$ . When current flowing through the detector becomes zero,

$$\frac{\left(r_1 + \frac{1}{j\omega C_1}\right)}{R_3} = \frac{\frac{1}{j\omega C_2}}{\left(\frac{R_4}{(1 + j\omega R_4 C_4)}\right)}$$

$$\begin{aligned} \left(r_1 + \frac{1}{j\omega C_1}\right)R_4 &= \frac{R_3(1 + j\omega R_4 C_4)}{j\omega C_2} \\ r_1 R_4 - \frac{jR_4}{\omega C_1} &= \frac{-j R_3}{\omega C_2} + \frac{R_3 R_4 C_4}{C_2} \end{aligned}$$

Equating real and imaginary parts, we get

$$r_1 = \frac{R_3 C_4}{C_2}$$

$$C_1 = \frac{C_2 R_4}{R_3}$$

Since  $R_4$  and  $C_2$  are fixed, the dial of  $R_3$  is calibrated to read the capacitance  $C_1$  directly.

(ii) The dielectric specimen should be in the size of  $C_1$  (in its area) and is inserted between the plates of  $C_1$ . Now once again the bridge is balanced. Now the dial reading in  $R_3$  will give the value of new capacitance  $C'_1$ .

$$\text{Then } \frac{C'_1}{C_1} = \varepsilon_r = \text{dielectric constant of the specimen.}$$

We can determine the dielectric constant at different frequencies by changing the frequency of the oscillator.

## 29.6 DIELECTRIC BREAKDOWN

*When a dielectric loses its insulation resistance and permits large currents to pass through it, it is said to breakdown.*

Whenever the electric field strength applied to a dielectric exceeds a critical value, very large current flows through it. The dielectric loses its insulating property and becomes conducting. This phenomenon is called dielectric breakdown.

There are five important types of dielectric breakdown.

### 1. Intrinsic breakdown

In a dielectric, the charge displacement increases with increasing field strength. Beyond a critical value of the field strength, there is an electric breakdown due to the physical deterioration of the dielectric material (Fig. 29.14).

The *dielectric strength* is defined as the breakdown voltage per unit thickness of the material.

$$\text{Dielectric strength} = \frac{\text{Dielectric breakdown voltage}}{\text{Thickness of dielectric}}$$

When the applied electric field is large, some of the electrons in the valence band cross over to the conduction band across the large forbidden energy gap. They become conduction electrons and give rise to large conduction currents. The liberation or movement of electrons from valence band is called internal field emission of electrons. This breakdown is called the *Intrinsic breakdown* or *Zener breakdown*.

### Avalanche breakdown

Since the electric field is over the order of  $10^6$  V/cm, the electrons are accelerated to very high velocity, colliding with the other atoms and molecules thereby releasing more electrons and holes breaking covalent bonds between them. The number of bonds broken and hence number of charge carriers released increases enormously with time and finally dielectric breakdown occurs. This type of breakdown is called *Avalanche breakdown*.

This requires relatively large electric fields. Hence this breakdown occurs in thin samples.

### Characteristics

1. This kind of breakdown requires relatively large electric fields.
2. Mostly this kind of breakdown occurs in thin samples.

### 2. Thermal breakdown

*Thermal breakdown* occurs in a dielectric when the rate of heat generation is greater than the rate of dissipation. Energy due to the dielectric loss appears as heat. If the rate of generation of heat is larger than the heat dissipated to the surrounding, the temperature of the dielectric increases which eventually results in local melting. Once melting starts, that particular region becomes highly conductive, enormous current flows through the material and dielectric breakdown occurs.

### Characteristics

1. It occurs only at high temperatures.
2. The strength of the electrical field to create dielectric breakdown depends upon the size and shape of the insulation sample.
3. The breakdown time is of the order of few milliseconds.
4. Since the dielectric loss is directly proportional to frequency, the electric field strength to create dielectric breakdown will be smaller for alternating fields and higher for D.C. fields.

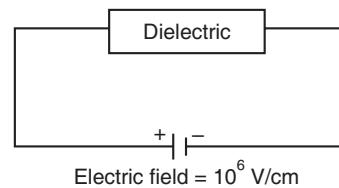


Fig. 29.14

### 3. Electrochemical breakdown

Electrochemical breakdown has a close relationship with thermal breakdown. When temperature rises, mobility of ions increases. Hence electrochemical reaction takes place. When ionic mobility increases, leakage current also increases. This results in dielectric breakdown. The chemical reaction reduces the insulation resistance and finally creates the dielectric breakdown. This type of breakdown is called *electrochemical breakdown*.

#### Characteristics

1. Electrochemical breakdown is determined by the leakage current, density of ions, temperature and permanent dipoles in the material.
2. Electrochemical reactions are accelerated by high temperatures. So to avoid breakdown, insulating materials should not be operated at high temperatures.

### 4. Discharge breakdown

Discharge breakdown is classified as internal or external.

(i) *Internal breakdown* occurs when the insulator contains occluded gas bubbles. Since gases require smaller ionisation potential than solids, the gaseous atoms ionise first and the gaseous ions bombard the solid dielectric causing electrical deterioration. This can occur at low voltages if there are large number of occluded gas bubbles in the insulating material.

(ii) *External breakdown* is caused by a glow or corona discharge. Such discharges are normally observed at sharp edges of electrodes. Carbon is formed on the electrodes due to these discharges. So the damaged areas become conducting leading to power arc and complete failure of the dielectric. Dust and/or moisture on the surface of the dielectric may also cause external discharge breakdown.

When discharge takes place at a point, the surrounding places are burnt and hence their electrical properties are affected. Thus the life of the insulation of an insulating material depends upon the number of discharges which are taking place inside the material. That is, it depends upon the frequency of the applied voltage.

### 5. Defect Breakdown

The surface of the dielectric materials may have defects such as cracks, porosity and blow holes (Fig. 29.15).

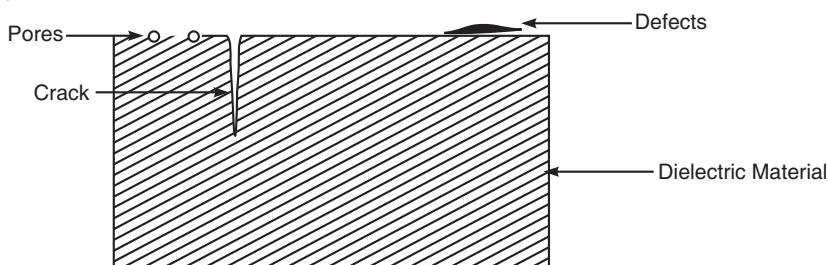


Fig. 29.15

Then impurities like dust or moisture may collect at these discontinuities (defects). This leads to breakdown called *defect breakdown*.

- Glazing is done on ceramic insulators to make the surface nonabsorbent.

## 29.7

### PROPERTIES AND APPLICATIONS OF DIFFERENT TYPES OF INSULATING MATERIALS

Insulating materials are classified into three categories.

1. **Solid insulating materials:** Mica, Porcelain, asbestos, bakelite, rubber, PVC, polythene, glass, cotton, silk, wood and ceramics.

**2. Liquid insulating materials:** Varnish, transformer oil, cable oil, and silicon liquids.

**3. Gaseous insulating materials:** Air, nitrogen, inert gases and sulphur hexafluoride.

#### **1. Solid insulating materials**

##### **(i) Mica**

It is a mineral compound of silicate of aluminium with silicates of soda potash and magnesia.

##### **Properties:**

1. It is crystalline in nature.
2. It can be easily split into very thin flat sheets.
3. It is rigid, tough and strong. It is not affected by moisture .
4. Its dielectric constant varies between 5 and 7.5.
5. Its dielectric strength varies between 700 and 1000 kV / mm .
6. It has high dielectric strength and low power loss.

##### **Uses:**

1. Mica sheets are used as insulator between commutator segments.
2. Thick mica tape is used for taping the stator coils of high voltage alternators.
3. It is used in electrical irons, hot plates and toasters for insulation purposes.
4. It is used as a dielectric material for high frequency applications.

##### **(ii) Ceramics**

They are generally non-metallic inorganic compounds such as silicates, aluminates, oxides, carbides, borides, nitrides and hydroxides.

##### **Properties:**

1. Ceramics are hard, strong, dense and brittle.
2. They are completely stable at high temperatures. They are not affected by moisture and by chemical action except with strong acids and alkalies.
3. They may be either crystalline or amorphous.
4. The ceramics have excellent dielectric and mechanical properties. The dielectric constant of the commonly used ceramics varies between 4 and 10.

##### **Uses:**

1. The ceramics are widely used as insulators for switches, plug holders, cathode heaters, vacuum type ceramic metal seals etc.
2. They are very much used as dielectric materials in capacitors. These capacitors withstand very high temperature and very high voltage.
3. **Refractory materials** are the materials which do not deteriorate or disintegrate, when they are subjected to high temperatures. Alumina is an important refractory material. It is primarily made of aluminium oxide. It has high compressive strength, absorbs less moisture and is excellent in heat conduction. It is used in high temperature appliances like furnaces.

##### **(iii) PVC materials**

When acetylene and hydrogen chloride are combined in presence of catalyst at temperature of about 50°C, polyvinyl chloride (PVC) resin is produced.

##### **Properties:**

1. They have good mechanical and electrical properties.
2. They are non - corrosive.

**Uses:**

1. They are widely used in insulation for wires and cables.
2. PVC films, tapes, and sheets are commonly used as insulation for dry batteries, conductors and cables.

**(iv) Rubber**

They are organic polymers and may be natural or synthetic. Natural rubber has limited applications because of its poor stability at wide temperature range. The synthetic rubbers are produced artificially by copolymerisation of isobutylene and isoprene.

**Properties:**

1. The rubbers have good electrical and thermal properties.
  2. Rubber possesses good dielectric properties and high tensile strength.
- The dielectric constant of rubber varies between 2.5 and 5.

**Uses:**

1. Rubber is widely used as an insulating material for electric wires, cables, tapes, coatings, transformers, motor winding etc.
2. It is used in the construction of storage battery housings and panel boards.

**(v) Glass**

Glass is an inorganic material made by the fusion of different oxides like silicon oxide, zinc oxide and magnesium oxide. Silica glass, Pyrex and fibre glass are the important glass insulating materials.

**Properties:**

1. It is brittle and hard. It is insoluble in water.
2. It has low dielectric loss and has good mechanical strength.
3. It is highly chemical resistant to most corrosive agents. It is a good insulator with good appearance.

**Uses:**

1. Glass is used as dielectric in capacitors.
2. It is used as an insulator in radio and television tubes, electrical lamps and laminated boards.
3. Toughened glass is used for insulation in extra high voltage lines having voltages above 100 kilovolts.

**(vi) Asbestos**

It is a naturally occurring mineral material of fibrous structure. Asbestos generally consists of magnesium silicate composition.

**Properties:**

1. It has high dielectric loss and low dielectric strength.
2. It can withstand very high temperature (400°C).

**Uses:**

1. Asbestos finds extensive use in electrical machines because of its ability to withstand very high temperature.
2. It is used for making of cloth tapes and paper boards for insulation purposes.
3. It is used for covering/insulating the wire in electric heating devices, ovens, electric irons, etc.

**2. Liquid insulating materials**

Liquid insulating materials can be divided into three groups.

- (i) **Mineral insulating oils** : Transformer oil, cable oil, capacitor oil, etc.
- (ii) **Synthetic insulating oils** : Askarels, aroclors, sovol and sovtol.
- (iii) **Miscellaneous insulating oils** : Vegetable oils, vaseline and silicon liquids.

**(i) Mineral insulating oils**

These oils are obtained from crude petroleum by distillation and have high oxidation resistance and good thermal stability.

**Properties and uses**

Transformer oil is the important mineral insulating oil with high dielectric strength, viscosity and cooling properties. This oil is used for insulation and cooling of transformer. It transfers heat from windings and core to the cooling surfaces by convection . It maintains the insulation of the windings.

**(ii) Synthetic insulating oils**

Now-a-days synthetic oil is used as an insulator in transformers in the place of transformer oil (mineral oil) because synthetic oils are very much resistant to oxidation and to fire hazards. Due to longer life and safety in operating conditions, synthetic oil is used as coolant and insulator in H. V. transformers. Particularly askarels have thermal stability upto 110°C and are manufactured from chlorinated hydrocarbons.

**(iii) Miscellaneous insulating oils**

Silicon liquids have stability upto 200°C and are costly. The dielectric strength of these liquids is same as that of mineral oils and power factor is very low. They are used in H.V. Transformers. Further these are used to increase the surface resistivity of ceramic insulators. Vaseline has high viscosity and high dielectric constant. It is used for impregnation of papers used in capacitors.

**3. Gaseous insulating materials**

**(i) Air**

It is the most important insulating material available in nature. The dielectric constant of air increases linearly with the increase of pressure. It is used in air capacitors. The power loss is practically zero. It is used for insulation between the two conductors of high voltage transmission lines. Air can be used as insulation only in the low voltage applications since at higher field strengths air may get ionised.

**(ii) Nitrogen**

Nitrogen is chemically inert. It prevents oxidation and reduces the rate of deterioration. In oil filled transformers, the nitrogen is used to replace oxidizing atmosphere. This is also used in capacitors and in cables under pressure.

**(iii) Inert gases**

They are used in electronic tubes and discharge tubes as insulators.

**(iv) Sulphur hexafluoride**

It is produced when sulphur is burnt in fluorine atmosphere. It has high dielectric strength. It has superior cooling properties than those of air and nitrogen. It has high chemical stability upto 100°C and is used in transformers and electric switches. Further it is also used in Van de Graff generator, Voltage stabilizer and X-ray apparatus for insulation purposes.

**FERROELECTRIC CRYSTALS**

**29.8 INTRODUCTION**

- A ferroelectric crystal exhibits an electric dipole moment even in the absence of an external electric field. In the ferroelectric state, the center of positive charge of the crystal does not coincide with the centre of negative charge.

- The plot of polarizations versus electric field for the ferroelectric state shows a hysteresis loop. A crystal in a normal dielectric state usually does not show significant hysteresis when the electric field is increased and then reversed, both slowly.

### Ferroelectric materials

Materials which exhibit electric polarization even in the absence of the applied electric field are called Ferroelectric materials. They have permanent dipole moment in each atom or molecule. They have a high dielectric constant.

*A ferroelectric material exhibits spontaneous polarization even in the absence of an external electric field.*

### Examples of Ferroelectric Materials

- (1) Barium titanate ( $\text{BaTiO}_3$ ),
- (2) Potassium dihydrogen phosphate ( $\text{KH}_2\text{PO}_4$ ),
- (3) Ammonium dihydrogen phosphate ( $\text{NH}_4\text{H}_2\text{PO}_4$ ),
- (4) Lithium Niobate ( $\text{LiNbO}_3$ ),
- (5) Rochelle salt ( $\text{NaKC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$ )

### Ferroelectricity

*Ferroelectricity refers to the creation of enormous value of induced dipole moment in a weak electric field as well as existence of electric polarisation even in the absence of applied electric field.*

- Ferroelectricity is a result of dielectric hysteresis. Since these materials exhibit hysteresis effects, similar to those observed in ferromagnetic materials, they are called ferroelectric materials.
- Ferroelectricity usually disappears above a certain temperature called the transition temperature. Above the transition, the crystal is said to be in a *paraelectric* state (analogous to paramagnetism); similarly, there is a rapid drop in the dielectric constant as the temperature increases.

## 29.9 PROPERTIES OF FERROELECTRIC MATERIALS

- (1) Ferroelectric materials can be easily polarised even by very weak electric fields.
- (2) They exhibit dielectric hysteresis. Lagging of polarization behind the applied electric field is called dielectric hysteresis. Ferroelectricity is a result of dielectric hysteresis (Fig. 29.16).
- (3) Ferroelectric materials possess spontaneous polarization.
- (4) They possess permanent electric dipoles and internal electric field which develop spontaneous polarization at or below the ferroelectric Curie temperature. *The electric dipoles in a ferroelectric solid are all aligned in the same direction, even in the absence of an electric field.*
- (5) They have a domain structure which is visible in polarized light.
- (6) Ferroelectric materials exhibit ferroelectricity when the temperature  $T \leq T_c$  where  $T_c$  = ferroelectric Curie temperature. When  $T > T_c$ , they are converted into paraelectric materials.

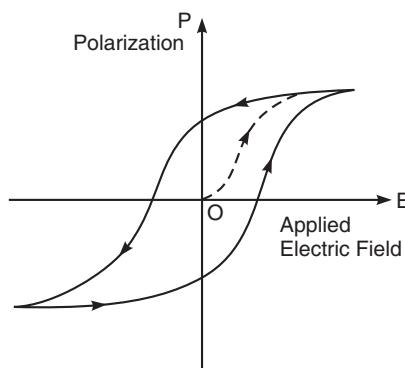


Fig. 29.16

(7) In ferroelectric materials, the dielectric constants are some two orders of magnitude larger than those in ordinary dielectrics. *Barium titanate* is a ferroelectric with a relative dielectric constant of over 2000, compared to less than 10 for ordinary dielectrics.

Fig. 29.17 shows relative permittivity versus temperature curve.

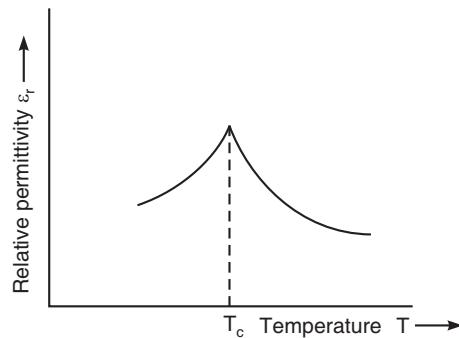


Fig. 29.17

(8) Ferroelectric materials exhibit *piezoelectric property*. Piezoelectricity means the creation of electric polarisation by mechanical stress.

If a compressive stress is applied to the crystal, a potential difference  $V$  develops between the two ends of the crystal.  $V$  will be of the opposite sign, if a tensile stress is applied.

(9) Ferroelectric materials exhibit *pyroelectric property*. Pyroelectricity means the creation of electric polarisation by thermal stress.

#### Structure of $\text{BaTiO}_3$

Fig. 29.18(a) shows the crystal structure of barium titanate.

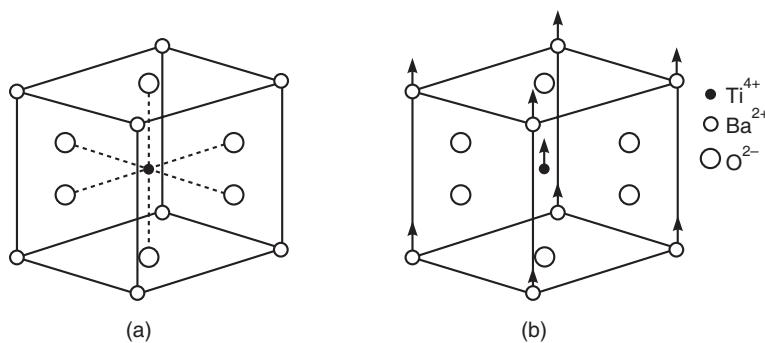


Fig. 29.18

- The prototype crystal is calcium titanate (perovskite).
- The structure is cubic, with  $\text{Ba}^{2+}$  ions at the cube corners,  $\text{O}^{2-}$  ions at the face centers, and a  $\text{Ti}^{4+}$  ion at the body center. (b) Below the Curie temperature, the structure is slightly deformed, with  $\text{Ba}^{2+}$  and  $\text{Ti}^{4+}$  ions displaced relative to the  $\text{O}^{2-}$  ions, thereby developing a dipole moment. The upper and lower oxygen ions may move downward slightly.
- $\text{BaTiO}_3$  is interesting from the viewpoint of practical applications because first, it is chemically and mechanically very stable and secondly, it exhibits ferroelectric properties at above room temperature.

**Applications of Ferroelectricity**

- (i) The dielectric constant of ferroelectrics is very high (thousands or many thousands in a number of cases) and so they are used in manufacture of small size high capacity capacitors.
- (ii) The property of hysteresis makes it possible to use them as memory devices for computers.
- (iii) Piezoelectric acoustic transducers and pyroelectric infrared detectors are the devices based also on ferroelectrics.
- (iv) The piezoelectric constants of quartz are quite small, but the great mechanical and thermal strengths of this solid have made it pre-eminent for highly stable piezoelectric oscillators. These devices control the frequencies of radio transmitters and quartz clocks, particularly quartz wrist watches.

**EXERCISE**

1. The displacement vector is \_\_\_\_\_.  
 (a)  $D = \epsilon_0 E$       (b)  $D = \epsilon_0 E + P$       (c)  $D = \sqrt{\epsilon_0} E$       (d)  $D = \epsilon_0 E - P$
2. \_\_\_\_\_ polarizability is produced by opposite displacement of positive nucleus and negative electrons within the same atom.  
 (a) ionic      (b) electronic      (c) dipolar      (d) orientation  
 (B.U. 2013)
3. Ionic polarizability is \_\_\_\_\_ to the square of the natural frequency of the ionic molecule.  
 (a) inversely proportional      (b) directly proportional  
 (c) equal to      (d) twice  
 (B.U. 2012)
4. The Lorentz field ( $E_L$ ) is given by  
 (a)  $E_L = E + \frac{P}{\epsilon_0}$       (b)  $E_L = E + \frac{P}{3\epsilon_0}$       (c)  $E_L = 2E + \frac{P}{\epsilon_0}$       (d)  $E_L = 3E + \frac{P}{\epsilon_0}$   
 (B.U. 2013)
5. The Clausius-Mossotti equation relates  
 (a) conductivity and dielectric constant      (b) conductivity and polarizability  
 (c) polarizability and dielectric constant      (d) polarizability and dipole moment  
 (B.U. M.Sc. 2006)
6.  $\frac{n\alpha}{3\epsilon_0} = \frac{(\epsilon_r - 1)}{(\epsilon_r + 2)}$  is known as \_\_\_\_\_ relation.  
 (a) Debye      (b) Clausius-Mossotti      (c) Einstein-Debye      (d) Bose-Einstein  
 (B.U. 2013)
7. Dielectric materials are those which are used to store \_\_\_\_\_ energy.  
 (a) potential      (b) kinetic      (c) total      (d) electrical  
 (B.U. 2011)  
 [Ans. 1(b), 2(b), 3(a), 4(b), 5(c), 6(b), 7(d)]
8. What are polar and nonpolar molecules in dielectrics? Give examples.      (B.U. 2010)
9. Define polarization.      (Madras 2006)
10. (i) Explain molecular polarizability.      (B.U. 2013)  
 (ii) Explain atomic polarizability.      (B.U. 2013)
11. Discuss the electronic polarization in Dielectrics.      (B.U. 2012)
12. Discuss in detail the different types of polarization mechanisms in dielectrics.      (B.U. 2011)

13. Give a schematic sketch of the variation of the total polarizability of an atom as a function of the frequency, explaining the physical origin of the various contributions and the relevant frequency ranges.
14. Derive the Clausius-Mossotti relation. *(B.U. 2015)*
15. Derive the Clausius-Mossotti relation for electronic polarizability and describe an experiment to determine the dielectric constant. *(B.U. 2012)*
16. Discuss local field in solid dielectrics. Deduce the Clausius-Mossotti equation. *(B.U. 2010)*
17. Write a note on dielectric breakdown. *(Madras 2006)*
18. Discuss the properties of any five insulating materials. *(Madras 2006)*
19. Give the general properties of Dielectric materials. *(B.U. 2011)*
20. Explain the ferroelectric crystals. *(B.U. 2013)*
21. What is ferroelectricity? Explain. *(B.U. 2011)*
22. Explain dielectric relaxation and dielectric loss in solids. Give an idea of dielectric constant and loss angle for some insulating materials.

**After reading this chapter, you should be able to**

- ◆ Identify the different types of magnetism found in nature
- ◆ Classify the different types of magnetic materials based on their capabilities to take up magnetization
- ◆ Discuss magnetic behaviour of solids, viz., dia-, para-, ferro-, antiferro- and ferri-magnetism, and their theories
- ◆ Explain why ferrites are preferred over ferromagnetic materials at high frequency as core material
- ◆ Utilize your understanding of the magnetic phenomenon to appreciate the operating principles of the various magnetic devices.

### 30.1 INTRODUCTION

- Any material that can be magnetised by the application of external magnetic field is called a **magnetic material**.
- Diamagnetic, paramagnetic, ferro-magnetic, antiferromagnetic and ferrimagnetic materials are the most important magnetic materials. These magnetic materials are widely used in industry.

First let us define the important terms involved in magnetism.

(i) **Magnetic induction ( $B$ ) or magnetic flux density.**

The magnetic induction ( $B$ ) in any material is the number of lines of magnetic force passing perpendicular through unit area (Fig. 30.1).

- Its unit is  $\text{Wbm}^{-2}$  or Tesla.
- The larger the number of field lines crossing per unit normal area, the larger is the magnitude of the magnetic field  $B$ .

(ii) **Magnetic field intensity ( $H$ ).** The magnetic field intensity ( $H$ ) at any point in the magnetic field is the force experienced by a unit north pole placed at that point.

- Its unit is  $\text{Am}^{-1}$ .
- The magnetic induction  $B$  due to a magnetic field of intensity  $H$  applied in vacuum is

$$B = \mu_0 H \quad \dots (1)$$

Here,  $\mu_0$  is the *permeability of free space* (vacuum).

The permeability of free space has a value of  $4\pi \times 10^{-7} \text{ Hm}^{-1}$ .

- If a magnetic field of intensity  $H$  is applied in a solid medium, the magnetic induction ( $B$ ) in the solid is given by

$$B = \mu H \quad \dots (2)$$

Here,  $\mu$  is the *permeability* of the solid material through which the magnetic lines of force pass.

$$\mu = \frac{B}{H} \quad \dots (3)$$

Hence the magnetic permeability ( $\mu$ ) of any material is the ratio of the magnetic induction ( $B$ ) in the sample to the applied magnetic field intensity ( $H$ ).

The ratio of  $\mu/\mu_0$  is called the relative permeability ( $\mu_r$ ) of the solid.

$$\mu_r = \frac{\mu}{\mu_0} \quad \dots (4)$$

(iii) **Intensity of Magnetization ( $M$ ).** Intensity of magnetisation ( $M$ ) of a material is defined as the magnetic moment per unit volume.

- Its unit is  $\text{Am}^{-1}$ .

$$M = \frac{\mu_m}{V}$$

$\mu_m \rightarrow$  magnetic moment of the substance

$V \rightarrow$  volume of the specimen.

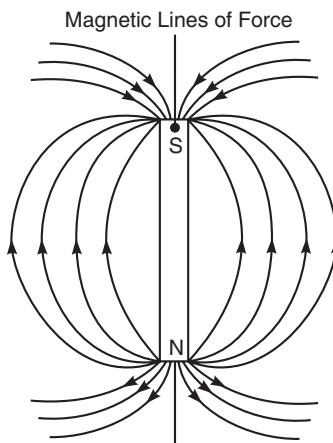


Fig. 30.1

(iv) Magnetic susceptibility ( $\chi$ )

**Definition.** Magnetic susceptibility ( $\chi$ ) of a material is the ratio of the intensity of magnetisation ( $M$ ) produced in the sample to the magnetic field intensity ( $H$ ) which produces the magnetisation.

$$\chi = \frac{M}{H}$$

- It has no units.
- The sign and magnitude of the magnetic susceptibility are used to determine the nature of the magnetic materials.
- Relation between  $\mu_r$  and  $\chi$  is

$$\mu_r = 1 + \chi$$

## DIAMAGNETISM

- The individual atoms of a diamagnetic material do not possess a permanent magnetic moment (Fig. 30.2).

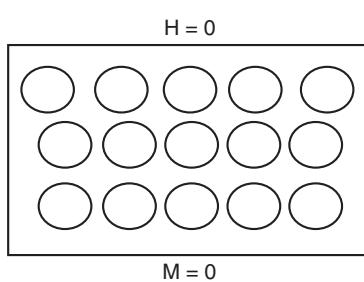


Fig. 30.2

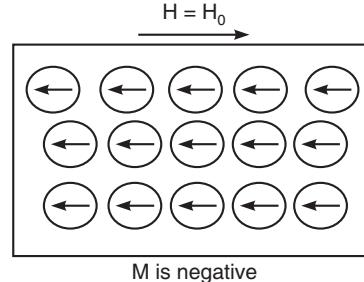


Fig. 30.3

- When an external magnetic field  $H_0$  is applied, the atoms acquire a small induced magnetic moment in a direction opposite to the direction of applied field (Fig. 30.3). The strength of the induced magnetic moment is directly proportional to the applied field  $H_0$ . The induced dipoles and magnetization vanish as soon as the applied magnetic field is removed.
- Diamagnetism is a property of all atoms because of the influence of an applied magnetic field on the motion of electrons in their orbits.

**Properties of diamagnetic materials**

- (1) Permanent dipoles are absent. There is no permanent dipole moment.
- Antimony, bismuth, mercury, gold and copper are some examples of diamagnetic substances.

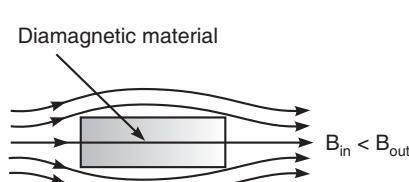


Fig. 30.4

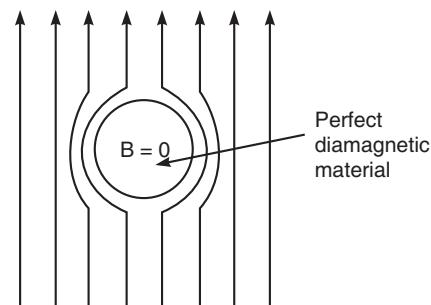


Fig. 30.5

(2) Figure 30.4 shows a bar of diamagnetic material placed in an external magnetic field. The magnetic lines of force are repelled or expelled. The field inside the material is reduced.

Figure 30.5 shows the behaviour of a *perfect diamagnetic material* in the presence of the magnetic field.

- Diamagnetic materials repel the magnetic lines of force.

- The magnetic flux density  $B$  is less inside than outside.

(3) The magnetic susceptibility is negative ( $\chi < 0$ ), i.e., magnetisation opposes the applied field. Magnetic susceptibility is independent of temperature and applied magnetic field strength.

(4) Relative permeability is slightly less than unity ( $\mu_r < 1$ ).

## 30.2 LANGEVIN'S THEORY OF DIAMAGNETISM

Consider an electron (mass =  $m$ , charge =  $e$ ) rotating about the nucleus (charge =  $Ze$ ) in a circular orbit of radius  $r$ . Let  $\omega_0$  be the angular velocity of the electron. Then

$$F_o = m\omega_0^2 r = Ze^2/(4\pi \epsilon_0 r^2)$$

or 
$$\omega_0 = \sqrt{\frac{Ze^2}{4\pi \epsilon_0 mr^3}} \quad \dots(1)$$

The magnetic moment of the electron is

$$\vec{m} = \text{current} \times \text{area} = \frac{e\omega_0}{2\pi} \times \pi r^2 = \frac{e}{2} \omega_0 r^2 \quad \dots(2)$$

Let a magnetic field of induction  $\mathbf{B}$  be now applied.  $\mathbf{B}$  is normal to and into the page (Fig. 30.6).

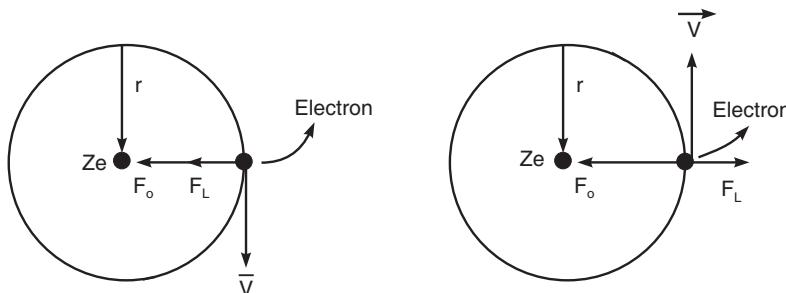


Fig. 30.6

An additional force  $F_L$  called Lorentz force, acts on the electron.

$$\mathbf{F}_L = -e(\mathbf{v} \times \mathbf{B}) = -eBr\omega$$

The condition of stable motion is now given by

$$mr\omega^2 = \frac{Ze^2}{4\pi \epsilon_0 r^2} - eBr\omega \quad \dots(3)$$

or 
$$\omega^2 + \frac{eB}{m}\omega - \frac{Ze^2}{4\pi \epsilon_0 mr^3} = 0$$

Solving the quadratic equation in  $\omega$ ,

$$\omega = \frac{-eB \pm \sqrt{\left(\frac{eB}{m}\right)^2 + 4\left(\frac{Ze^2}{4\pi \epsilon_0 mr^3}\right)}}{2} = \pm \sqrt{\omega_0^2 + \left(\frac{eB}{2m}\right)^2} - \frac{eB}{2m}.$$

$$\text{or } \omega = \pm \omega_0 - \frac{eB}{2m} \quad \left( \because \frac{eB}{2m} \ll \omega_0 \right) \quad \dots(4)$$

Thus the angular frequency is now different from  $\omega_0$ . The result of establishing a field of flux density  $B$  is to set up a precessional motion of the electronic orbits with angular velocity  $-(e/2m)B$ . This is called Larmor theorem. Then

$$\left. \begin{array}{l} \text{change in frequency of} \\ \text{revolution of the electron} \end{array} \right\} = \delta n = -\frac{eB}{4\pi m}$$

The corresponding change in the magnetic moment of the electron is

$$\Delta m = \text{current} \times \text{area} = \left\{ e \times \left( \frac{-eB}{4\pi m} \right) \right\} \times \pi r^2 = -\frac{Be^2 r^2}{4m} \quad \dots(5)$$

On summing over all electrons in the atom, the induced moment per atom becomes

$$\Delta m_{\text{atom}} = -\frac{Be^2 \Sigma r^2}{4m}$$

Let  $N$  be the number of atoms per unit volume. Then the magnetisation  $M$  is given by

$$M = -\frac{NBe^2 \Sigma r^2}{4m} \quad \dots(6)$$

All the electron orbits are not oriented normal to the magnetic field. Hence  $r^2$  in Eq. (6) should be replaced by the average of the square of the projection of orbit radii for various electrons in a plane perpendicular to  $B$ . Hence we should replace  $r^2$  in Eq. (6) by  $\frac{2}{3}r^2$ .

$$\therefore M = -\frac{NBe^2 \Sigma r^2}{6m}$$

Volume susceptibility of the material

$$\chi = \frac{M}{H} = -\frac{NBe^2 \Sigma r^2}{6mH} = -\frac{\mu_0 Ne^2 \Sigma r^2}{6m} \quad (\because B = \mu_0 H)$$

$$\therefore \chi = -\frac{\mu_0 Ne^2 \Sigma r^2}{6m} = -\frac{\mu_0 e^2}{6m} NZ <r^2> \quad \dots(7)$$

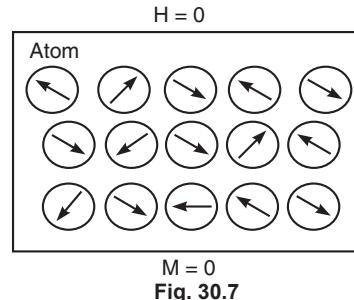
Eq. (7) shows that  $\chi$  is independent of the field strength and temperature. This is in accord with Curie's experimental results.

### PARAMAGNETISM

(i) **In the absence of external magnetic field.** The individual atoms of paramagnetic material possess a permanent magnetic dipole moment of their own (Fig. 30.7).

- Each atom possesses a permanent magnetic moment.
- When  $H = 0$ , all the magnetic moments are randomly oriented because of the ceaseless random thermal motion of the atom. So the net magnetization  $M = 0$ .

(ii) **When an external magnetic field is applied.** When an external magnetic field  $H_0$  is applied, the magnetic dipoles tend to align themselves in the direction of the magnetic field (Fig. 30.8). The individual atomic dipole moments point in the same direction. The material becomes magnetized. This effect is called *paramagnetism*.



- Magnetisation  $M$  and magnetic field  $H$  are in the same direction. Since  $\chi = M/H$ , the susceptibility  $\chi$  is positive.
- Examples of paramagnetic materials:** Platinum, aluminium, ferric oxide, ferrous sulphate, nickel sulphate, etc.

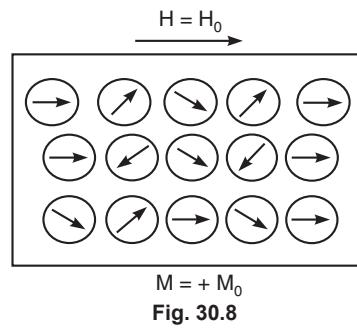


Fig. 30.8

### 30.3 PROPERTIES OF PARAMAGNETIC MATERIALS

- Paramagnetic materials possess permanent magnetic dipoles.
- In the absence of an external applied field, the dipoles are randomly oriented. Hence the net magnetization in any given direction is zero.
- When placed inside a magnetic field, it attracts the magnetic lines of force (Fig. 30.9). The field lines get concentrated inside the material, and the field inside is enhanced. This enhancement is slight, being one part in  $10^5$ .
- Paramagnetic susceptibility is positive and depends on temperature.

- $\chi = \frac{C}{T}$  is Curie's law.

- $\chi = \frac{C}{T - \theta}$  is called Curie-Weiss law.

Here,  $C$  is Curie constant and  $\theta$  is a constant called paramagnetic Curie temperature.

- The value of the paramagnetic susceptibility is independent of the applied magnetic field strength.
- Spin alignment is random (Fig. 30.10).

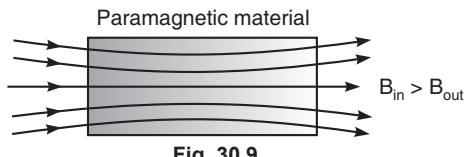


Fig. 30.9



Fig. 30.10

### 30.4 LANGEVIN'S THEORY OF PARAMAGNETISM

He assumes that each atom has a permanent magnetic moment  $m$ . The only force acting on the atom is that due to the external field  $B$ . Let  $\theta$  be the angle of inclination of the axis of the atomic dipole with the direction of the applied field  $B$ . Then magnetic potential energy of the atomic dipole is

$$U = -mB \cos\theta$$

Now, on classical statistics, the number of atoms making an angle between  $\theta$  and  $\theta + d\theta$  is

$$dn = Ce^{mB \cos\theta/kT} \sin\theta d\theta$$

where  $k$  is Boltzmann's constant and  $T$  is the absolute temperature.

Put  $mB/kT = \alpha$ . Then

$$dn = Ce^{\alpha \cos\theta} \sin\theta d\theta \quad \dots(1)$$

Hence the total number of atomic magnets in unit volume of the paramagnetic material

$$n = \int_0^\pi dn = \int_0^\pi Ce^{\alpha \cos\theta} \sin\theta d\theta \quad \dots(2)$$

Put  $\cos \theta = x$ . Then  $-\sin \theta d\theta = dx$ .

$$\begin{aligned} n &= \int_{+1}^{-1} -Ce^{\alpha x} dx = C \int_{-1}^{+1} e^{\alpha x} dx \\ \therefore C &= \frac{n\alpha}{e^\alpha - e^{-\alpha}} \end{aligned} \quad \dots(3)$$

The component of each dipole moment parallel to  $B$  in  $m \cos \theta$ . The total magnetic moment of all the  $n$  atoms contained in unit volume of the gas is the magnetisation  $M$ . It is given by

$$M = \int_0^\pi m \cos \theta dn = \int_0^\pi m \cos \theta Ce^{\alpha \cos \theta} \sin \theta d\theta \quad \dots(4)$$

Put  $\cos \theta = x$ . Then,  $-\sin \theta d\theta = dx$ . Therefore, we get

$$M = \int_{+1}^{-1} -mxCe^{\alpha x} dx = Cm \int_{-1}^{+1} xe^{\alpha x} dx$$

Evaluating this integral and substituting the value of  $C$  from (3), we get

$$\begin{aligned} M &= mn \left[ \frac{e^\alpha + e^{-\alpha}}{e^\alpha - e^{-\alpha}} - \frac{1}{\alpha} \right] \\ &= mn \left[ \coth \alpha - \frac{1}{\alpha} \right] \\ &= mn L(\alpha) \end{aligned} \quad \dots(5)$$

where  $L(\alpha) = \left[ \coth \alpha - \frac{1}{\alpha} \right]$  is called the *Langevin function*.

The variation of  $M$  with  $\alpha$  is shown in Fig. 30.11.

**Case (i):** At low temperatures or large applied field,  $L(\alpha) \rightarrow 1$ .

Hence, magnetisation  $M$  in this case will be

$$M = mn \quad \dots(6)$$

So saturation is reached when all the atomic dipoles are parallel to  $B$ .

**Case (ii) :** Under normal conditions  $\alpha$  is very small. Then,

$$L(\alpha) = \coth \alpha - \frac{1}{\alpha} \approx \frac{\alpha}{3} \quad \dots(7)$$

$$\therefore M = mn \frac{\alpha}{3} = \frac{nm^2 B}{3kT} = \frac{nm^2 \mu_0 H}{3kT} \quad \dots(8)$$

$$\chi = \frac{M}{H} = \frac{\mu_0 nm^2}{3kT} = \frac{C}{T} \quad \dots(9)$$

where  $C = \mu_0 nm^2 / 3k$  is called the Curie constant.

**Failure of Langevin Theory.** (i) Langevin's theory was unable to explain a more complicated dependence of susceptibility upon temperature exhibited by several paramagnetics such as highly compressed and cooled gases, very concentrated solutions of salts, etc.

(ii) Langevin's theory could not account for the intimate relation between para-and ferro-magnetism.

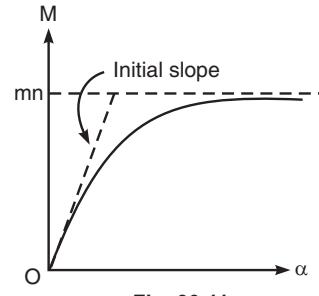


Fig. 30.11

**Weiss Modification:** Langevin's theory applies strictly only to gases, where the molecules are sufficiently far apart for their mutual interactions to be negligible. In liquids and solids such interactions may be large, and many substances obey the modified *Curie-Weiss law*

$$\chi = \frac{C}{T-\theta} \quad \dots(10)$$

$\theta$  is called the *Curie temperature* and is characteristic of the substance. Eq. (10) holds only at temperatures where  $T > |\theta|$ . Eq. (10) is of the same form as Eq. (9), except that the origin of temperature is shifted from 0 to  $\theta$ .

### 30.5 WEISS THEORY OF PARAMAGNETISM

Weiss introduced the concept of *internal molecular field* in order to explain the complicated type of dependence of susceptibility. In a real gas, the molecules are mutually influenced by their magnetic moments and consequently, there should exist within the gas a molecular field. This field produced at any point by all the neighbouring molecules, is proportional to and acting in the same sense as the intensity of magnetization ( $M$ ). Let this internal molecular field be  $H_i$ . Now

$$H_i = \lambda M \quad \dots(1)$$

Here,  $\lambda$  is molecular field coefficient.

Therefore, the net effective field should be

$$H_e = (H + H_i) \quad \dots(2)$$

Here,  $H$  is external applied field.

Following the Langevin theory along with this effective field,

$$M = \frac{Nm^2\mu_0 H_e}{3kT} = \frac{Nm^2 \mu_0 (H + \lambda M)}{3kT} \quad [\text{Refer Eq. (8) in Langevin theory}]$$

Here,

$N$  = number of dipoles/unit volume,

$m$  = magnetic moment of each atomic dipole,

$k$  is Boltzmann's constant and  $T$  is the absolute temperature.

Rearranging the terms, we get

$$M \left( 1 - \frac{Nm^2 \mu_0 \lambda}{3kT} \right) = \frac{Nm^2 \mu_0 H}{3kT}$$

Let  $C = \mu_0 Nm^2 / 3k$  and  $\theta = C\lambda$ . Then we get

$$\begin{aligned} M \left( 1 - \frac{\theta}{T} \right) &= \frac{CH}{T} \\ \therefore \chi &= \frac{M}{H} = \frac{C}{T \left( 1 - \frac{\theta}{T} \right)} = \frac{C}{T - \theta} \end{aligned} \quad \dots(3)$$

Here  $C$  is called the Curie constant.

$\theta$  is called *paramagnetic Curie point* or the *Curie temperature*.

Eq. (3) is called *Curie-Weiss law*.

Below Curie temperature ( $T < \theta$ ), susceptibility becomes negative; i.e., paramagnetics would become diamagnetics. However, for most of the paramagnetic substances, Curie temperature is quite low so that a situation for which  $T < \theta$  is rare.

### 30.6 QUANTUM THEORY OF PARAMAGNETISM

The magnetic moment of an atom or ion in free space is given by

$$\vec{\mu} = \gamma \cdot \vec{\hbar} \mathbf{J} = -g \mu_B \mathbf{J} \quad \dots(1)$$

Here, total angular momentum  $\vec{\hbar} \mathbf{J}$  = orbital angular momentum  $\vec{\hbar} \mathbf{L}$  + spin angular momentum  $\vec{\hbar} \mathbf{S}$ .

The ratio of the magnitude of  $\vec{\mu}$  to that of  $\mathbf{J}$  is called the *gyromagnetic ratio* ( $\gamma$ ) or *magnetogyric ratio*.  $\mu_B$  is the Bohr magneton. It is defined as  $e \hbar / 2m$ .

For a free atom the  $g$  factor is given by the Lande equation

$$g = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \quad \dots(2)$$

For an electron spin  $g = 2.00$ .

The energy levels of the system in a magnetic field are

$$U = -\vec{\mu} \cdot \vec{\mathbf{B}} = m_J g \mu_B \mathbf{B} \quad \dots(3)$$

Here,  $m_J$  is the azimuthal quantum number.

$$m_J = J, (J-1), \dots, -(J-1), -J$$

For a single spin with no orbital moment, we have  $m_J = \pm \frac{1}{2}$  and  $g = 2$ .

$$\therefore U = \pm \mu_B \mathbf{B}. \quad \dots(4)$$

Thus an energy difference exists between spin magnetic dipole moments parallel and anti-parallel to the field. The magnitude of energy difference is  $2\mu_B \mathbf{B}$ . This splitting is shown in Fig. 30.12.

Let us suppose that in unit volume of the paramagnetic material there is a total of  $N$  atoms. The spin moments of  $N_1$  atoms are parallel to the field and  $N_2$  spin moments are antiparallel.

If a system has only two levels the equilibrium populations are,

$$\text{with } \tau = k_B T,$$

$$\frac{N_1}{N} = \frac{\exp(\mu B / \tau)}{\exp(\mu B / \tau) + \exp(-\mu B / \tau)}, \quad \dots(5)$$

$$\frac{N_2}{N} = \frac{\exp(-\mu B / \tau)}{\exp(\mu B / \tau) + \exp(-\mu B / \tau)}. \quad \dots(6)$$

Here  $N_1$  and  $N_2$  are the populations of the lower and upper levels, and  $N = N_1 + N_2$  is the total number of atoms. The fractional populations are plotted in Fig. 30.13.

The projection of the magnetic moment of the upper state along the field directions is  $-\mu$  and of the lower state is  $\mu$ . The resultant magnetization for  $N$  atoms per unit volume is

$$M = (N_1 - N_2)\mu = N\mu \cdot \frac{e^x - e^{-x}}{e^x + e^{-x}} = N\mu \tanh x \quad \dots(7)$$

$$\text{Here, } x \equiv \mu B / k_B T \quad \dots(8)$$

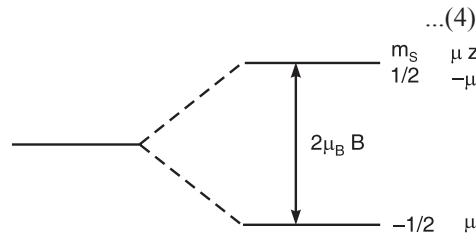


Fig. 30.12

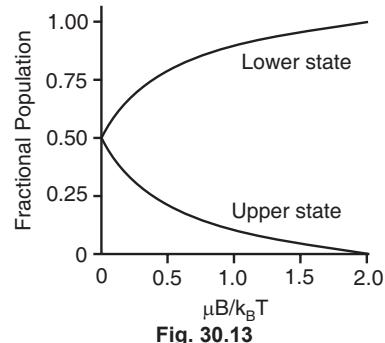


Fig. 30.13

For  $x \ll 1$ ,  $\tan h x \approx x$ , and we have

$$M \approx N\mu_B (\mu_B/k_B T) \quad \dots(9)$$

In a magnetic field an atom with angular momentum quantum number  $J$  has  $(2J+1)$  equally spaced energy levels. The Magnetization is given by

$$M = Ng\mu_B B_J(x) \quad \dots(10)$$

Here,

$$x \equiv gJ\mu_B B/k_B T \quad \dots(11)$$

Eq. (10) is called the *Curie-Brillouin law*.

$B_J(x)$  is called the *Brillouin function* defined as

$$B_J(x) = \frac{2J+1}{2J} \operatorname{ctnh} \left( \frac{(2J+1)x}{2J} \right) - \frac{1}{2J} \operatorname{ctnh} \left( \frac{x}{2J} \right) \quad \dots(12)$$

Equation (7) is a special case of Eq. (12) for  $J = \frac{1}{2}$ .

Fig. 30.14 shows the plot of magnetic moment versus  $B/T$  for spherical samples of (I) potassium chromium alum, (II) ferric ammonium alum, and (III) gadolinium sulphate octahydrate.

For  $x \ll 1$ , we have

$$\operatorname{ctnh} = \frac{1}{x} + \frac{x}{3} - \frac{x^3}{45} + \dots, \quad \dots(13)$$

The susceptibility per unit volume is

$$\chi = \frac{M}{B} \approx \frac{N\mu_B (2J+1)}{3k_B T} = \frac{Np^2 \mu_B^2}{3k_B T} = \frac{C}{T} \quad \dots(14)$$

Here  $p$  is called the *effective number of Bohr Magnetons*. It is defined as

$$p \equiv g [J(J+1)]^{1/2} \quad \dots(15)$$

Eq. (14) is called the *Curie law*.  $C$  is called the *Curie constant*.

### Rare Earth Ions

Equation (14) is successfully employed to predict the values of susceptibility for various paramagnetic crystals particularly rare earth ions. The value of  $J$  is determined by applying the Hund's rules.

Experiments on rare-earth ions in crystals show that they obey the Curie law, with an effective number of magnetons in agreement with the theory of spin-orbit interaction. In these ions, therefore, the angular momenta  $\mathbf{L}$  and  $\mathbf{S}$  are strongly coupled, and the moment of the ion can respond freely to the external field.

In these ions—from La to Lu in the periodic table—the  $4f$  shell is incompletely filled. The outer  $5p$  shell is completely filled, while the  $5d$  and  $6s$  shells which are still further out are stripped of their electrons to form the ionic crystal (Fig. 30.5). Thus the only incomplete shell is the  $4f$  shell. This is the one in which the magnetic behaviour occurs. Electrons in this shell lie deep within the ion, screened by the outer  $5p$  and  $5d$  shells. So they are not appreciably affected by other ions in the crystal. Magnetically their behaviour is much like that of a free ion.

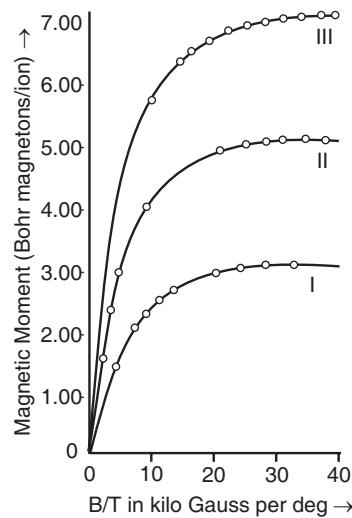


Fig. 30.14

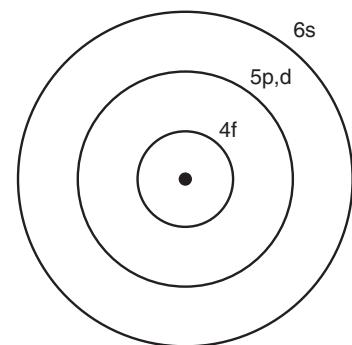


Fig. 30.15

### Iron-group ions

Equation (14) is unable to account for the experimental observations of susceptibility for the ions of the iron group. The reason for this is that for the iron group, the ground state of the ions in a crystal is very different from that of the free ion. This is because, the partly filled electron shell, which is responsible for producing the magnetic moment, is the outermost ( $3d$ ) shell of these ions. It is, therefore, very strongly influenced by the crystal field. The interaction is so strong that the orbital angular momentum of the electrons in the  $3d$  shell is very often reduced to zero. This does not mean that the electrons no longer encircle the nucleus, but rather that they will be travelling around it as often in one sense as in the other sense. This effect when  $L$  averages to zero is termed *quenching* of the orbital angular momentum and it means that when the ion is in the crystal, there is no orbital magnetic moment. In quenching, the crystal field hardly interacts with the spin and hence, the spin magnetic moment still exists. Therefore, while calculating susceptibility on the basis of Eq. (14), we should replace  $J$  by  $S$ . When this is done, quite good agreement is obtained between the theoretical and experimental Curie constant.

## 30.7 COOLING BY ADIABATIC DEMAGNETIZATION OF A PARAMAGNETIC SALT

**Experimental arrangement.** Fig. 30.16 shows the experimental arrangement.

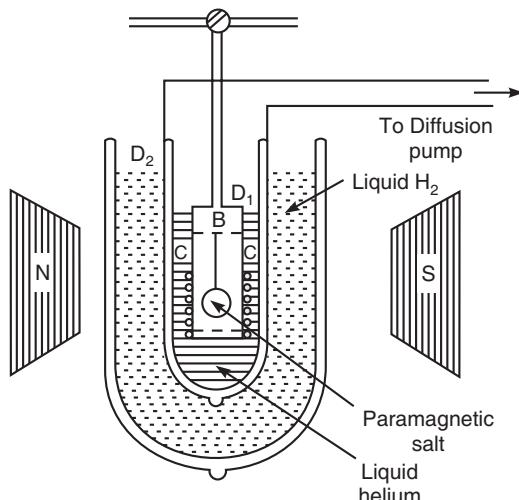


Fig. 30.16

The paramagnetic salt (gadolium sulphate) is suspended inside a glass bulb  $B$ . Bulb  $B$  is surrounded by Dewar flasks  $D_1$  and  $D_2$  containing liquid helium and liquid hydrogen respectively. The whole arrangement is placed between the poles of a strong electromagnet.

### Working

- The magnetic field is switched on, so that the specimen is magnetised.
  - The heat due to magnetisation is removed by first introducing hydrogen gas into  $B$  and then pumping it off with a high vacuum pump. Now the cold magnetised specimen is thermally isolated from  $D_1$  and  $D_2$ .
  - The magnetic field is now switched off.
- Adiabatic demagnetisation of the specimen takes place and its temperature falls.
- The final temperature of the specimen ( $T_2$ ) is determined by measuring the susceptibility of the substance at the beginning and at the end of the experiment by using a *solenoid coil CC*.

**Calculation.** Let  $\chi_1$  and  $\chi_2$  be the susceptibilities of the specimen salt at the initial and final temperature  $T_1$  and  $T_2$ . Then, according to Curie's law,

$$\frac{\chi_1}{\chi_2} = \frac{T_2}{T_1}$$

or  $T_2 = \frac{\chi_1}{\chi_2} T_1$ .

- Using this method with the salt gadolinium sulphate, Giauque and Macdougall reached a temperature of 0.25 K.

**Principle of the method.** At a fixed temperature, the entropy of a system of magnetic moments is lowered by the application of a magnetic field.

The entropy is a measure of the disorder of a system. Therefore, the greater the disorder, the higher is the entropy. In the magnetic field the magnetic moments will be partly lined up *i.e.*, partly ordered, so that the entropy is lowered by the magnetic field. The entropy is also lowered if the temperature is lowered, as more of the magnetic moments line up.

Suppose the magnetic field is removed without changing the entropy of the spin system. Then, the order of the spin system will look like a lower temperature than the same degree of order in the presence of magnetic field.

When the specimen is demagnetized adiabatically, entropy can flow into the spin system only from the lattice vibrations (Fig. 30.17).

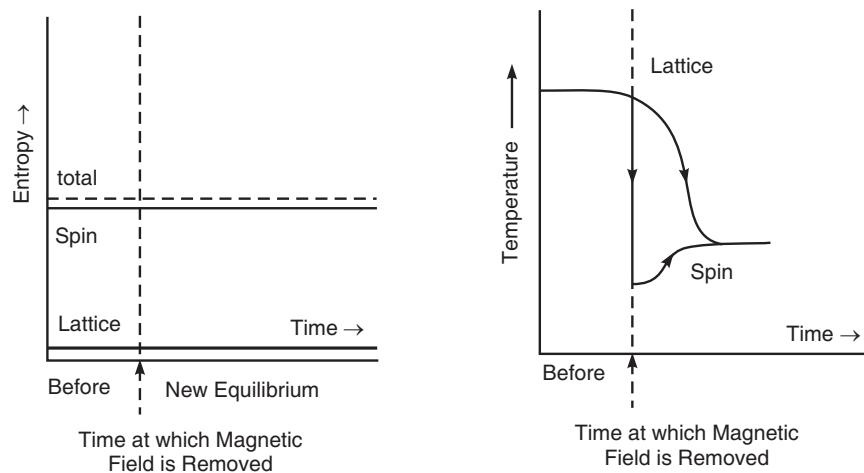


Fig. 30.17

- At very low temperatures, the entropy of lattice vibrations is usually negligible. Thus, the entropy of the spin system will be essentially constant during adiabatic demagnetization of the specimen. It is important to mention that magnetic cooling is a one-shot operation, and it is not cyclic.

#### Expression for the spin entropy

Let us find an expression for the spin entropy of a system of  $N$  ions, each of spin  $S$ , at a temperature sufficiently high that the spin system is entirely disordered. That is,  $T$  is supposed to be much higher than some temperature  $\theta$  which characterizes the energy of the interactions ( $E_{int} = k_B \theta$ ) tending to orient the spins preferentially. The entropy  $\sigma$  of a system of  $G$  accessible states is  $\sigma = k_B \ln G$ . At a temperature so high that all of the  $(2S + 1)$  states of each ion are nearly equally populated,  $G$  is the number of ways of arranging  $N$  spins in  $(2S + 1)$  states. Thus, we have

$$G = (2S + 1)^N.$$

The spin entropy

$$\sigma_S = k_B \ln (2S+1)^N = Nk_B \ln (2S+1) \quad \dots(1)$$

This spin entropy is reduced by a magnetic field if the lower levels gain in population when the field separates the  $2S+1$  states in energy.

**Steps in the cooling process.** Fig 30.18 shows the steps carried out in the cooling process.

(i) Magnetic field is applied at temperature  $T_1$  with the specimen in good thermal contact with the surroundings, giving the isothermal path  $ab$ . At temperature  $T_1$ , the thermal contact is provided by helium gas and by removing the gas with a pump the thermal contact is broken.

(ii) The specimen is then insulated ( $\Delta\sigma = 0$ ) and the magnetic field is removed. Thus the specimen follows the constant entropy path  $bc$ , ending up at temperature  $T_2$ .

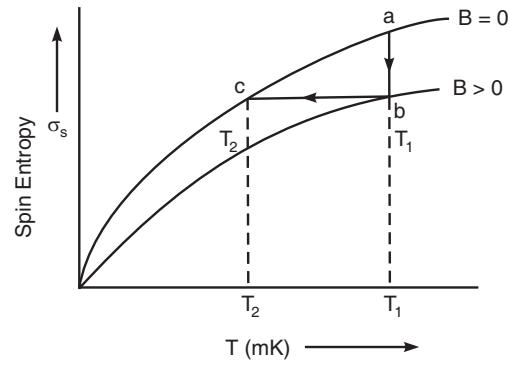


Fig. 30.18

### FERROMAGNETISM

- Ferromagnetism is the existence of a *spontaneous magnetization*, even in zero applied field. Ferromagnetic materials have a small amount of magnetisation even in the absence of an external magnetic field. This indicates that there is a strong internal field within the material which makes the atomic magnetic moments align with each other.
- When placed in a magnetic field, ferromagnetic materials become strongly magnetized in the direction of the applied field. The direction of magnetization is the same as that of the external field.

**Origin of ferromagnetism.** Ferromagnetism arises due to permanent magnetic moments in the atoms or molecules of the material. When an external field is applied, the magnetic moments line up in the same direction as that of the applied field.

**Examples of ferromagnetic materials.** Iron (Fe), Cobalt (Co), Nickel (Ni), and Gadolinium (Gd).

### 30.8 PROPERTIES OF FERROMAGNETIC MATERIALS

- (1) All the dipoles are aligned parallel to each other due to the magnetic interaction between any two dipoles. Figure 30.19 shows the dipole alignment.
- (2) Ferromagnetic materials have permanent dipole moment.
- (3) When placed inside a magnetic field, a ferromagnetic material attracts the magnetic line of forces very strongly (Fig. 30.20).

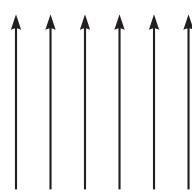


Fig. 30.19

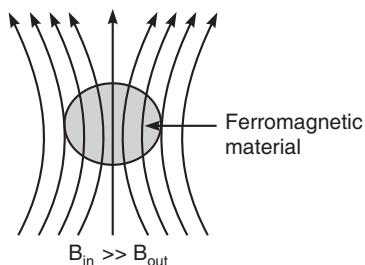


Fig. 30.20

(4) They exhibit magnetisation even in the absence of a magnetic field. This property of ferromagnetic materials is called **spontaneous magnetisation**.

(5) Ferromagnetic materials exhibit the phenomenon of hysteresis.

(6) On heating, they lose their magnetisation slowly.

(7) The magnetic susceptibility ( $\chi$ ) of ferromagnetic materials is very high and depends on **temperature ( $T$ )**.

- As temperature increases, the value of susceptibility decreases.
- The ferromagnetic property depends on temperature. At high enough temperatures, a ferromagnet becomes a paramagnet.

The temperature of transition from ferromagnetism to paramagnetism is called the *paramagnetic Curie temperature  $\theta$* .

The susceptibility above the Curie temperature, i.e., in the paramagnetic phase is described by,

$$\chi = \frac{C}{T-\theta} \quad (T > \theta).$$

Here,  $C$  is the Curie constant.

For  $T > \theta$ , paramagnetic behaviour.

For  $T < \theta$ , ferromagnetic behaviour.

Figure 30.21 shows the variation of susceptibility with temperature for ferromagnetic materials.

(8) The relative permeability  $\mu_r$  of ferromagnetic materials is very high. The relative magnetic permeability is  $> 1000$ !

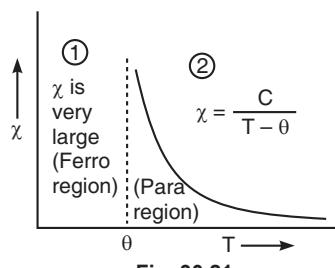


Fig. 30.21

**Table 30.1. Comparison of dia, para and ferromagnetic materials**

		Diamagnetic material	Paramagnetic material	Ferromagnetic material
1.	Magnetic moment	There is no permanent dipole moment (or) magnetic moment in each atom.	There is permanent dipole moment (or) magnetic moment in each atom.	There is enormous permanent dipole moment (or) magnetic moment in each atom.
2.	Spin alignment	No spin.	All spins are randomly oriented	Spin alignment is parallel in the same direction.
3.	Susceptibility and its temperature dependence	Susceptibility is always negative. It is independent of the temperature and strength of applied magnetic field.	It is always positive and small. It is inversely proportional to absolute temperature of the material. $\chi \propto \frac{1}{T} \text{ or } \chi = \frac{C}{T}$	It is always positive and very large. $\chi = \frac{C}{T-\theta}$ (Curie-Weiss law) <ul style="list-style-type: none"> <li>(i) For <math>T &gt; \theta</math>, paramagnetic behaviour.</li> <li>(ii) For <math>T &lt; \theta</math>, ferromagnetic behaviour.</li> </ul>

4.	Behaviour of material in the presence of magnetic field	When the material is placed in the magnetic field, the magnetic lines of force are repelled away from the material. $B_{\text{out}} > B_{\text{in}}$ .	The magnetic lines of force are attracted towards the centre of the material. $B_{\text{in}} > B_{\text{out}}$ .	The magnetic lines of force are highly attracted towards the centre of the material. $B_{\text{in}} >> B_{\text{out}}$ .
5.	Relative magnetic permeability ( $\mu_r$ )	$\mu_r$ is slightly less than 1.	$\mu_r$ is slightly greater than 1.	$\mu_r$ is very much greater than 1. $\mu_r \gg 1$
6.	Examples	Hydrogen, bismuth, antimony, gold and superconducting materials like Niobium.	Aluminium, platinum, sodium, titanium, zirconium and chromium.	Iron, nickel, cobalt, gadolinium.

### 30.9 DOMAIN THEORY OF FERROMAGNETISM

What is domain theory of ferromagnetism?

- Weiss proposed the concept of *domains* in order to explain the properties of ferromagnetic materials.

**Concept of magnetic domains.** The group of atomic dipoles (atoms with permanent magnetic moment) organised into tiny bounded regions in the ferromagnetic materials are called *magnetic domains*. Ferromagnetic material contains a large number of domains.

- Magnetic domain.** In demagnetized state, a ferromagnetic material is divided into a number of small regions called *domains*. Each domain is spontaneously magnetized.
- The *boundaries between different domains are called domain walls*. The domain walls are also called *Bloch walls*.
- In each individual domain, the magnetic moments of the atoms are aligned in the same direction.

Hence the domain is a region of the ferromagnetic material in which all the magnetic moments are aligned to produce a net magnetic moment in one direction only. Thus it behaves like a magnet with its own magnetic moment and axis.

(i) **Magnetic domains in a demagnetized ferromagnetic material.** In a demagnetized ferromagnetic material, the domains are *randomly oriented* (Fig. 30.22). So the magnetization of the material as a whole is zero.

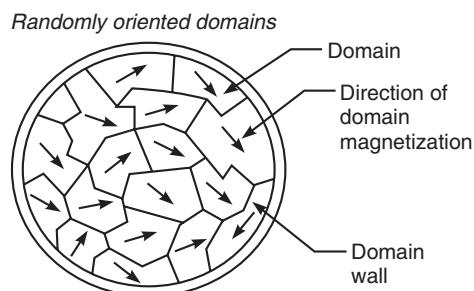


Fig. 30.22

(ii) **Domain configuration in a magnetized body.** When an external magnetic field is applied to a ferromagnetic material, then the domains align themselves in the direction of the magnetic field (Fig. 30.23). This results in a *large net magnetization of the material*.

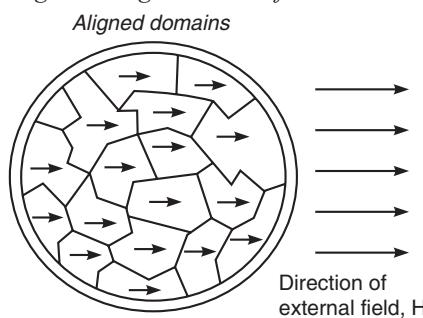


Fig. 30.23

### Ferromagnetic domains

A ferromagnet has a spontaneous magnetic moment — a magnetic moment even in zero applied magnetic field. The existence of a spontaneous moment suggests that electron spins and magnetic moments are arranged in a regular manner (Fig. 30.24).

*The magnetization is defined as the magnetic moment per unit volume.*

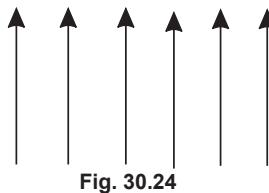


Fig. 30.24

If *domains* (regions magnetized in different directions) are present, the magnetization refers to the value within a domain.

Weiss introduced the concept of *ferromagnetic domains*. He postulates that any specimen consists of a large number of small regions or domains which are spontaneously magnetized. In the demagnetized state, the magnetization vectors of these domains are oriented in such a way that the specimen does not possess a resultant magnetic moment. Fig. 30.25 shows a single crystal domain arrangement for zero resultant magnetic moment. Magnetization occurs only when some of these vectors are caused to align themselves more or less parallel to an applied field.

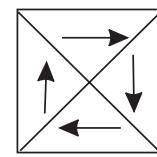


Fig. 30.25

The increase in the gross magnetic moment of a ferromagnetic specimen in an applied magnetic field takes place by two independent processes :

#### (i) By the Motion of Domain Walls

In weak fields, the volume of domains [Fig. 30.26(b)] favourably oriented with respect to the field increases at the expense of unfavourably oriented domains. The magnetization changes by means of domain boundary displacements and so the domains change in size.

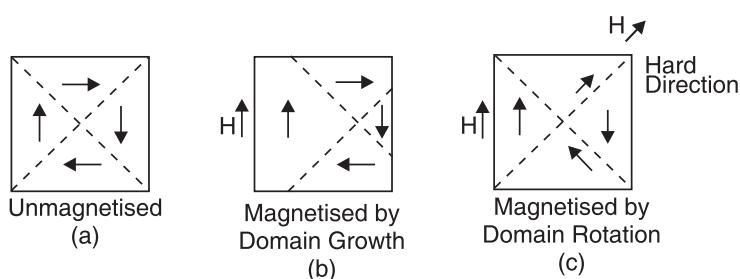
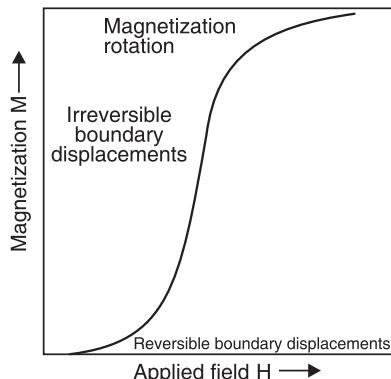


Fig. 30.26

### (ii) By Rotation of Domains

In strong applied fields, the domain magnetization rotates toward the direction of the field [Fig. 30.26 (c)]. The magnetization changes by means of rotation of direction of magnetization.

Fig. 30.27 depicts the magnetization curve, showing the dominant magnetization processes in the different regions of the curve.



**Fig. 30.27**

### Four Types of Energy Involved in the Process of Domain Growth

Domain structure occurs in order to minimize the total energy of the ferromagnetic substance. The total energy comprises the sum of the *exchange* energy, the *anisotropy* energy, and the *magnetic* energy.

#### (i) Exchange Energy

It is the energy associated with the quantum mechanical coupling that aligns the individual atomic dipoles within a single domain. It depends upon the interatomic distance. It arises from interaction of electron spins. The exchange energy of two electrons may be written in the form

$$U = -2J \vec{s}_1 \cdot \vec{s}_2 \quad \dots (1)$$

Here,  $J$  is the exchange integral and is related to the overlap of the charge distributions of the atoms 1, 2.  $\vec{s}_1, \vec{s}_2$  are the electron spins of atoms 1, 2.

#### (ii) Anisotropy Energy

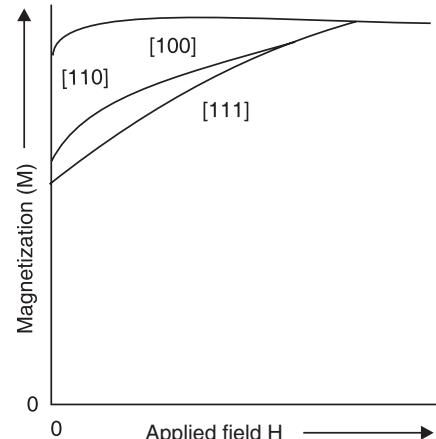
There is an energy in a ferromagnetic crystal which directs the magnetization along certain crystallographic axes called directions of easy magnetization. This energy is called the *magnetocrystalline or anisotropy energy*.

**Definition.** The excess of energy required to magnetize a specimen in a particular direction over that required to magnetize it along the easy direction is called *crystalline anisotropy energy*.

Iron is a cubic crystal. The cube edges are the directions of easy magnetization. Magnetization curves for single crystal iron are shown in Fig. 30.28. Along the  $<100>$  direction, a very small magnitude of the magnetic field is required to produce a large magnetization. It is ‘easy’ to magnetize an iron crystal along  $<100>$  direction, while it is ‘hard’ to do so along, say,  $<111>$  direction.

To represent the anisotropy energy of iron magnetized in an arbitrary direction with direction cosines  $\alpha_1, \alpha_2, \alpha_3$  referred to the cube edges, we are guided by cubic symmetry. The expression for the anisotropy energy must be an even power of each  $\alpha_i$ , provided opposite ends of a crystal axis are equivalent magnetically, and it must be invariant under interchanges of the  $\alpha_i$  among themselves.

The lowest order combination satisfying the symmetry requirements is  $\alpha_1^2 + \alpha_2^2 + \alpha_3^2$ , but this is identically equal to unity and does not describe anisotropy effects. The next combination is of the



**Fig. 30.28**

fourth degree:  $\alpha_1^2 \alpha_2^2 + \alpha_1^2 \alpha_3^2 + \alpha_3^2 \alpha_2^2$ , and then of the sixth degree :  $\alpha_1^2 \alpha_2^2 \alpha_3^2$ . Thus

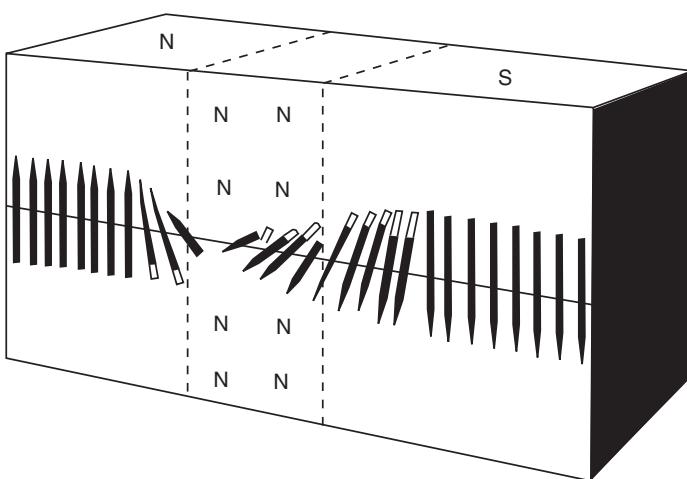
$$U_K = K_1(\alpha_1^2 \alpha_2^2 + \alpha_2^2 \alpha_3^2 + \alpha_3^2 \alpha_1^2) + K_2 \alpha_1^2 \alpha_2^2 \alpha_3^2.$$

Here,  $K_1, K_2$  are constants.

### (iii) Domain Wall Energy (Bloch Wall Energy)

*The transition region between the two oppositely magnetized domains is called the Bloch wall.*

Fig. 30.29 shows the structure of the Bloch wall separating domains. The arrows represent individual atomic magnetic moments. In iron the thickness of the transition region is about 300 lattice constants.



**Fig. 30.29**

The entire change in spin direction between domains does not occur in one discontinuous jump across a single atomic plane, but takes place in a gradual way over many atomic planes.

*Domain wall energy is the sum of contributions from the exchange and crystalline anisotropy energies in the domain wall region.*

### (iv) Magnetostrictive Energy

Magnetization of a crystal in a certain direction always produces changes in atomic spacing in that direction. This is called *magnetostriction*.

*The magnetostrictive energy is the energy due to the mechanical stresses generated by magnetostriction in the domains.*

#### Reversible and Irreversible Domains

If we apply a small external field on the ferromagnetic specimen, the domain wall is displaced slightly away from the minimum energy. But it returns to the original position when the field is removed. This gives a reversible domain wall movement and the domain is called reversible domain.

In larger external fields, the domain wall may be shifted to a more distant position where the energy curve has passed through a maximum and then diminished. On removing the field, the domain wall cannot cross the energy maximum. So it is unable to return to its initial position. This gives an irreversible domain wall movement.

#### Explanation of Hysteresis Curve on the Basis of Domain Theory

The Hysteresis of ferromagnetic materials refers to the lag of magnetisation ( $M$ ) behind the magnetising field ( $H$ ).

Technical terms defined by the  $B$ - $H$  hysteresis loop are shown in Fig. 30.30.

The coercivity  $H_c$  is the reverse field that reduces  $B$  to zero.

The remanence  $B_r$  is the value of  $B$  at  $H = 0$ .

(i) When a weak magnetic field is applied, domains where the magnetisation is parallel or at a small angle with the field grow at the expense of those where the magnetisation is antiparallel so that the boundary between domains is displaced. This boundary displacement is reversible and is indicated by the path  $OA$ .

(ii) When the magnetic field becomes stronger, the Bloch wall movement is sharp and is irreversible. The steeper part  $AC$  of the magnetisation curve is due to larger, irreversible displacements.

(iii) Above the knee of the curve ( $CS$ ), magnetization proceeds by rotation of the direction of magnetization of whole domains. Such a process is rather difficult and the increase in magnetisation is relatively slow. At  $S$ , all the domains are in the field direction and the specimen is said to be saturated.

(iv) When the applied field is reduced, there is a little change in the domain structure so that the magnetisation remains quite high, until high reverse fields are applied. Further even when the external field is zero, there is a residual magnetisation in the specimen and that can be destroyed by applying a high reverse field.

Thus the reversible and irreversible domain wall movements give rise to hysteresis in the ferromagnetic materials.

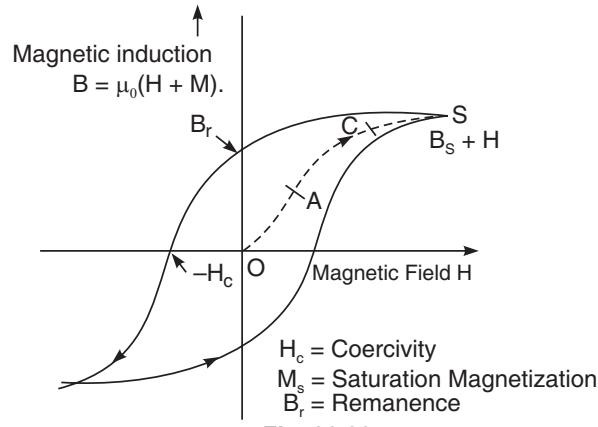


Fig. 30.30

### 30.10 SOFT AND HARD MAGNETIC MATERIALS

The process of magnetisation of a ferromagnetic material consists of moving the domain walls so that favourably oriented domains grow and unfavourably oriented domains shrink.

- If the domain walls are easy to move, the coercive field is low. It is easy to magnetise the material. Such a material is called a *soft magnetic material*.
- If it is difficult to move the domain walls, the coercive field is large and the material is magnetically *hard*.

Fig. 30.31 shows magnetisation curves for soft and hard magnetic materials.

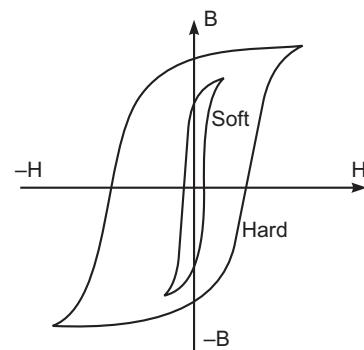


Fig. 30.31

#### Soft Magnetic Materials

Fig. 30.32 shows the nature of hysteresis loop of soft magnetic material (soft iron).

- Soft magnetic materials have low hysteresis loss due to small hysteresis loop area.
- In these materials, the domain wall movement is relatively easier. Even for small changes in the magnetizing field, magnetization changes by large amounts.
- The coercivity and retentivity are small. Hence these materials can be easily magnetized and demagnetized.

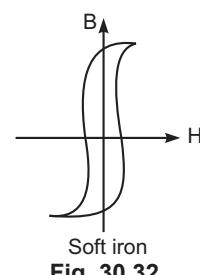


Fig. 30.32

- Soft magnetic materials are used in applications requiring frequent reversals of the directions of magnetisation such as cores of transformers.
- In soft magnetic materials, the hysteresis losses must be kept down to a minimum. When the magnetic induction is large for a small applied field, the loop area is small and the hysteresis loss is reduced. The key factor in the design of a soft magnet is then to have easily moving domain walls. Soft magnetic materials should be free of impurities and inclusions.

Usually there are easy and hard magnetisation directions in a crystal. Fig. 30.33 shows magnetisation curves for single crystals of iron. Iron magnetises more easily along (1 0 0) than along (1 1 1). (1 1 1) is the hard direction for iron. This property can be exploited to reduce the area under the hysteresis loop by manufacturing materials with a preferred orientation of grains.

- The other source of energy loss in soft magnetic materials is the eddy current loss (changing magnetic flux in a medium induces an emf). The induced emf is proportional to the rate of change of flux and hence to the frequency of the alternating current. The induced emf sets up eddy current. The power loss due to these is equal to  $V^2/R$ . Here,  $V$  is the induced emf and  $R$  is the resistance of the medium. Eddy current losses can be minimised by increasing the resistivity of the medium.

Iron, which used to be the material for transformer cores, is now almost entirely replaced by Fe–Si alloys, which has substantially higher resistivity than iron. Fe–Si alloys are suitable for operation at power frequencies of 50–60 Hz. At microwave frequencies, ferrites (48% MnO·Fe<sub>2</sub>O<sub>3</sub>, 52% ZnO·Fe<sub>2</sub>O<sub>3</sub>; 36% NiO·Fe<sub>2</sub>O<sub>3</sub>, 64% ZnO·Fe<sub>2</sub>O<sub>3</sub>) and garnets (3Y<sub>2</sub>O<sub>3</sub>·5Fe<sub>2</sub>O<sub>3</sub>) are preferred.

### Hard Magnetic Materials

Fig. 30.34 shows the nature of hysteresis loop of hard magnetic material (steel).

- Hard magnetic materials have large hysteresis loss due to large hysteresis loop area.
- In these materials, the domain wall movement is difficult because of presence of impurities and crystal imperfections and it is irreversible in nature.
- The coercivity and retentivity are large.

Hence, these materials cannot be easily magnetized and demagnetized.

Hard magnetic materials are used to produce permanent magnets. Hysteresis losses are of no significance here as no repeated reversals of magnetisation is involved in a permanent magnet. The permanent magnets must have high residual induction  $B_r$  and large coercive field  $H_c$ . The area of the hysteresis loop between  $B_r$  and  $H_c$  represents the energy required to demagnetise a permanent magnet. The maximum value of this area ( $= B_r H_r$ ), called the *energy product*, must be as large as possible for permanent magnets. High carbon steels and other low alloy tungsten and chromium steels are used for making permanent magnets.

### Distinction between soft and hard magnetic materials.

Table 30.2 gives the “*Differences between Soft and Hard magnetic Materials*”

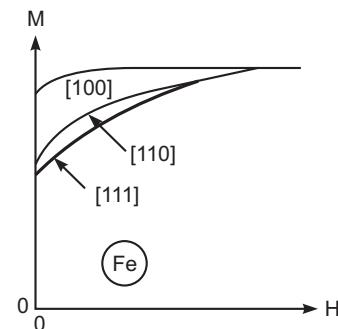


Fig. 30.33

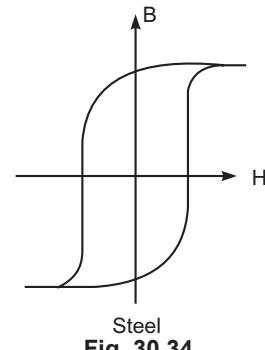
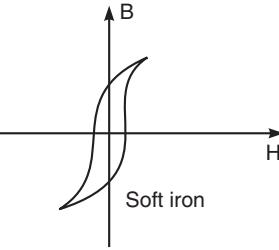
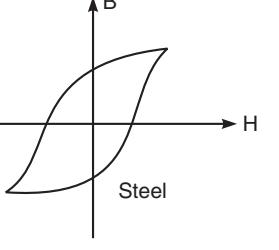


Fig. 30.34

Table 30.2. Distinction between Soft and Hard Materials

S.No.	Soft magnetic materials	Hard magnetic materials
1.	They can be easily magnetised and demagnetised.	They cannot be easily magnetised and demagnetised.
2.	They have low hysteresis loss due to small hysteresis loop area.  	They have large hysteresis loss due to large hysteresis loop area.  
3.	Magnetic energy stored is not high.	Magnetic energy stored is high.
4.	They have large values for permeability and susceptibility.	They have small values for permeability and susceptibility.
5.	Coercivity and retentivity are small.	Coercivity and retentivity are large.
6.	The eddy current loss is small due to its high resistivity.	The eddy current loss is more due to its small resistivity.
7.	The domain walls are easy to move.	The domain walls are hard to move.
8.	They are used to make electromagnets.	They are used to make permanent magnets.
9.	<b>Examples:</b> Iron silicon alloy, Nickel iron alloy, silicon steels and ferrites.	<b>Examples.</b> Tungsten steel, Cobalt steel, Alini, Alnico and Cunife.
10.	They are free from irregularities (in the crystal structure) like strains or impurities. Their magnetostatic energy is very small.	In these materials, the irregularities (in the crystal structure) like mechanical strains will be more. Their magnetostatic energy is large.

### 30.11 WEISS THEORY OF FERROMAGNETISM

According to Weiss, the atomic magnets of a ferromagnetic substance are grouped into certain regions or *domains*. When the substance is in the unmagnetised condition, the domains form closed chains with no free poles. When the substance is magnetized, the chains break up and the domains gradually set themselves with their magnetic axes all pointing in the field direction. Thus ferromagnetism is a *crystal phenomenon*.

Weiss assumed that a molecular magnetic field exists at the position of every atom or molecule. This field arises due to the interaction of all neighbouring molecules. The molecular field is proportional to the magnetisation vector  $I$ .

$$\therefore \text{molecular field } B_i = \beta I$$

Here,  $\beta$  = molecular field coefficient.

The effective field strength  $B_e$  may be regarded as the vector sum of external applied field strength  $B$  and the internal molecular field strength  $B_i$ .

Hence

$$B_e = B + B_i = B + \beta I \quad \dots(1)$$

Consider 1 gram mole of the substance. Let

$\rho$  = density of the substance

$M$  = molecular weight

$\sigma$  = gram molecular magnetic moment

$\sigma_0$  = saturation value of gram molecular magnetic moment.

Then

$$I = \frac{\sigma}{(M/\rho)} \text{ and } I_s = \frac{\sigma_0}{(M/\rho)}$$

$\therefore$

$$\frac{I}{I_s} = \frac{\sigma}{\sigma_0} \quad \dots(2)$$

The domains will obey the general theory of paramagnetism by Langevin.

$$\therefore \frac{I}{I_s} = \frac{\sigma}{\sigma_0} = \left( \coth \alpha - \frac{1}{\alpha} \right) \quad \dots(3)$$

Here,

$$\alpha = \frac{\mu B_e}{kT} = \frac{\mu(B + \beta I)}{kT}$$

$\mu$  = dipole moment of individual atom (molecule)

$k$  = Boltzmann constant

$T$  = absolute temperature of substance.

When external field is zero,

$$B_e = \beta I = \beta \frac{\sigma \rho}{M}$$

Now,

$$\alpha = \frac{\mu \beta \sigma \rho}{kTM} = \frac{\sigma_0}{N} \times \frac{\beta \sigma \rho}{kTM}$$

Here,  $N$  = Avogadro's number

$$\text{or } \alpha = \frac{\sigma_0 \beta \sigma \rho}{RTM}$$

$$\text{or } \frac{\sigma}{\sigma_0} = \frac{RTM}{\sigma_0^2 \beta \rho} \alpha \quad \dots(4)$$

Eqs. (3) and (4) simultaneously determine the condition of spontaneous magnetisation. Fig. 30.35 shows a graph drawn between  $\sigma/\sigma_0$  and  $\alpha$  corresponding to Eqs. (3) and (4). Curve 1 is Langevin curve corresponding to Eq. (3) while straight line 2 is corresponding to Eq. (4). The two curves meet at 0 and A. The solution  $\frac{\sigma}{\sigma_0} = 0$  is not true. Hence  $\frac{\sigma}{\sigma_0} = AD$  is

the true solution. It is obvious from the graph that A represents the stable state of spontaneous magnetisation. If the molecules in a domain assume state C, then the local magnetisation is less than the corresponding equilibrium state A. Now the magnetisation and the

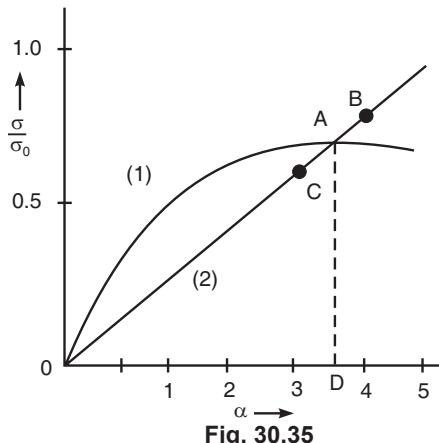


Fig. 30.35

value of  $\alpha$  will in consequence increase until the state  $AD$  is reached. On the other hand, if the molecules in a domain assume state  $B$ , then the local magnetisation is more than the equilibrium value. Now the magnetisation and the value of  $\alpha$  will tend to decrease until the state  $AD$  is reached.

We know that the slope of the tangent at the origin of the Langevin curve is  $\frac{1}{3}$  since  $\frac{\sigma}{\sigma_0} = \frac{\alpha}{3}$ ,

when  $\alpha$  is small. Hence, the condition for *stable spontaneous magnetisation* is given by

$$\frac{RTM}{\sigma_0^2 \beta \rho} < \frac{1}{3} \text{ or } T < \frac{\sigma_0^2 \beta \rho}{3RM}$$

But  $\frac{\sigma_0^2 \beta \rho}{3RM} = \theta$ , the Curie point. Hence  $T < \theta$ .

Hence below the Curie point  $\theta$ , in the absence of the external field, the domains are spontaneously magnetised to a degree depending upon temperature, approaching the saturation value as the temperature approaches absolute zero. Above the Curie point  $\theta$ , spontaneous magnetisation no longer occurs, the ferromagnetic properties disappear and the substance becomes paramagnetic. Therefore, Curie temperature is transition temperature. Above Curie point, the substance obeys Curie-Weiss law.

The chief drawbacks of Weiss theory are that (i) it could not explain why and how internal field between the molecules of ferromagnetic materials possess such large values, and (ii) why the linear relationship expressed by Curie-Weiss law breaks near Curie point.

### 30.12 NUCLEAR MAGNETIC RESONANCE

**Principle.** A nucleus with a nonzero spin acts like a tiny magnet. When placed in a magnetic field, the magnetic moment vector precesses round the field direction. This is called Larmor precession. This precession is quantised. If this precessing nucleus is placed in the path of a beam of electromagnetic radiation there will be coherent interaction between the particle and the radiation provided the frequency of the radiation is equal to that of precession. The phenomenon is thus one of *resonance*. In the case of nuclei, the phenomenon is called *nuclear magnetic resonance* (N. M. R.).

#### Expression for Resonance frequency

If a nucleus of magnetic moment  $\mu$  is placed in a uniform magnetic field  $B_0$  in the  $z$ -direction, then its energy is

$$U = -\mu_z B_0 \quad \dots(1)$$

The magnetic moment  $\mu_z$  is quantized according to the usual rule

$$\mu_z = g m_I \frac{e\hbar}{2m_p} \quad m_I = I, I-1, \dots, 1-I, -I \quad \dots(2)$$

where  $g$  is the  $g$ -factor for the nucleus. Hence the energy levels of the nucleus in the magnetic field are

$$U = -g m_I \frac{e\hbar}{2m_p} B_0 \quad m_I = I, I-1, \dots, 1-I, -I \quad \dots(3)$$

The energy difference between adjacent energy levels is

$$\Delta U = -g e \hbar B_0 / 2(m_p) \quad \dots(4)$$

The corresponding transition frequency is

$$v = \frac{\Delta U}{h} = \frac{g e B_0}{4\pi m_p}.$$

**NMR Spectrometer.** Fig. 30.36 shows a schematic arrangement for observing the nuclear magnetic resonance, and is based on the energy absorption method.

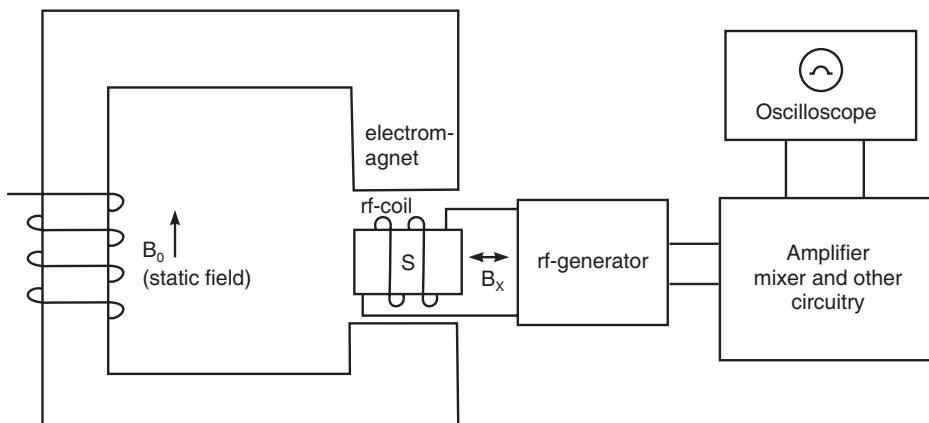


Fig. 30.36

The specimen  $S$ , about 1 c.c. of the material being investigated, is placed between the poles of an electromagnet. This magnet produces the field  $B_0$ . An r-f coil surrounding the specimen is carefully positioned to produce a second field  $B_x$  perpendicular to  $B_0$ . An r-f generator not only serves to drive the coil but also supplies a signal to auxiliary circuitry which measures the r-f power absorbed by the specimen. To trace out the absorption line, an auxiliary low-frequency oscillator supplies power to secondary coils wound on the main magnet core. These coils permit the main field  $B_0$  to be varied by a small amount on both sides of the resonant field. Associated electronic circuitry displays on an oscilloscope, a signal proportional to the energy absorbed as a function of field. Such a signal is sketched in Fig. 30.37.

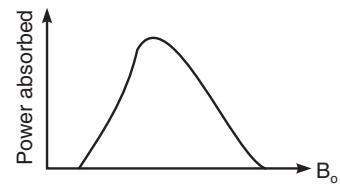


Fig. 30.37

#### Applications

- NMR is used for the accurate determination of nuclear moments.
- In a sensitive form of magnetometer, NMR is used for measuring magnetic fields.
- In medicine, NMR tomography has been developed. In this technique, images of tissues are produced by magnetic resonance. Whole plants, small animals or parts of bigger animals are placed between the pole pieces of a suitably designed magnet. Using surface coils of wire, radio frequency radiation is directed towards the part under study. Spectra obtained in this way are used for medical diagnosis. This method is supposed to be superior to diagnosis by X-rays because the potential damage to living tissue by X-rays is not present in the case of radio frequency radiation.
- The main application of NMR is in its use for chemical analysis and determination of molecular structures.

### 30.13

#### HEISENBERG'S INTERPRETATION OF INTERNAL FIELD OR WEISS MOLECULAR FIELD

The first theoretical explanation of the large Weiss internal field in ferromagnetic materials was proposed by Heisenberg.

1. The magnetization of ferromagnetic materials arises due to electron spin rather than due to orbital motion.

2. High value of internal fields can be explained in terms of quantum *exchange interactions* between the electrons.

3. An exchange force is caused by a static coulomb interaction and its magnitude is much larger than that of the magnetic interaction.

4. This exchange force appears in the form of spin-spin interaction.

5. The strength of interaction depends upon the interatomic separation and it may change its sign as the separation is varied.

6. Assume in each atom unpaired electrons are in parallel orientation. If the interatomic separation decreases more and more, exchange forces decrease until finally they pass through zero and an antiparallel spin alignment is favoured.

### Theory

Consider two ferromagnetic atoms *A* and *B* in the lattice (Fig. 30.38).

$\Psi_A$  and  $\Psi_B$  denote the atomic wavefunctions.

1 and 2 are the two unpaired electrons having negative charges.

The interaction potential ' $V_{AB}$ ' between the ions of *A* and *B* and electrons 1 and 2 is

$$V_{AB} = \frac{e^2}{4\pi\epsilon_0} \left[ \frac{1}{r_{AB}} + \frac{1}{r_{12}} - \frac{1}{r_{B1}} - \frac{1}{r_{B2}} - \frac{1}{r_{A1}} - \frac{1}{r_{A2}} \right]$$

Here,  $r_{AB}$  = distance between atoms *A* and *B*,

$r_{B1}$  = distance between atom *B* and electron 1

$r_{B2}$  = distance between atom *B* and electron 2

$r_{A1}$  = distance between atom *A* and electron 1

$r_{A2}$  = distance between atom *A* and electron 2.

The charge of each ion (atom) is equal to  $e$ .

The total exchange interaction energy between spins of unpaired electrons of the system is given by

$$E_{ex} = -2 \sum_{ij} (J_{ij})_e \vec{S}_i \cdot \vec{S}_j$$

Here

$S_i = S_j = 1/2$  = spin magnetic moment.

For parallel spins,  $\theta = 0$  and for antiparallel spins  $\theta = \pi$ .

Here exchange integral  $J_e = \int \Psi_A^*(1) \Psi_B^*(2) V_{AB} \Psi_A(1) \Psi_B(2) d\mathbf{r}_1 d\mathbf{r}_2$

which depends on overlap of atomic orbitals.

The total energy of the system is

$$E_{ex} = K \pm J_e \quad \dots(1)$$

Here,  $K$  is the coulomb interaction energy and  $J_e$  is the exchange integral.

The positive sign in Eq. (1) stands for the state when the spins of the electrons are antiparallel, i.e., non-magnetic state. The negative sign is for the state when the spins of the two electrons are parallel, i.e., for magnetic state.

Fig. 30.39 shows variation of exchange integral with distance ratio  $r_{AB}/r_0$ .

$r_0$  is the orbital radius of unpaired electrons.

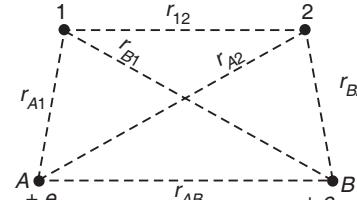


Fig. 30.38

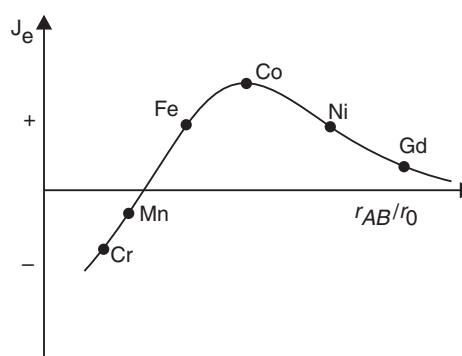


Fig. 30.39

A magnetic state is favoured when  $r_{AB}/r_0$  is slightly larger than 3. For Fe, Co and Ni this ratio has the values 3.26, 3.64 and 3.94 respectively.

So Fe, Co, and Ni are ferromagnetic. For Cr and Mn this ratio has the value 2.60 and 2.24, *i.e.*, these are not ferromagnetic.

### 30.14 QUANTUM THEORY OF FERROMAGNETISM

Ferromagnetic substances are those substances which possess a spontaneous magnetic moment *i.e.*, a magnetic moment even in the absence of an applied magnetic field. The Curie temperature  $\theta$  is the temperature at and above which the spontaneous magnetisation vanishes.

Weiss gave the theory of ferromagnetism on the basis of two hypotheses —

(1) A ferromagnetic specimen contains in general a large number of small regions called domains which are spontaneously magnetised. The spontaneous magnetisation of the specimen is determined by the vector sum of the magnetic moments of the individual domains.

(2) Within each domain the spontaneous magnetisation is due to existence of molecular field which tends to produce a parallel alignment of the atomic dipoles.

Existence of molecular moment leads to co-operation or interaction between the atomic dipoles. Tendency of co-operation is to produce parallel alignment.

To explain this fact Weiss assumed that magnetic field actually acting on the specimen is

$$H_{\text{eff}} = H + \lambda M \quad \dots(1)$$

On these assumptions we shall develop the quantum theory of ferromagnetism.

According to quantum theory, since the magnetic moments are quantized, the magnetic dipole moment  $\mu$  and its component  $\mu_z$  in the direction of the applied field can not have arbitrary values. We have, in general, a direct relationship between the magnetic dipole moment  $\mu$  of an atom or ion in free space and its angular momentum  $J$  as

$$\mu = -g \mu_B J \quad \dots(2)$$

$\mu_B$  is the Bohr magneton. It is defined as  $e\hbar/2m$ .

The  $g$  factor is given by the Lande equation

$$g = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \quad \dots(3)$$

Here,  $S$  and  $L$  represent the spin and orbital quantum numbers of the dipole respectively.

The orientations of the magnetic moment  $\mu$  with respect to the direction of the applied magnetic field are specified by the rule that the possible components of  $\mu$  along the field direction are given by

$$\begin{aligned} \mu_z &= -g \mu_B m_J \\ m_J &= J, (J-1), \dots, -(J-1), -J. \end{aligned} \quad \dots(4)$$

For each value of  $J$ ,  $m_J$  can have  $(2J+1)$  values which means that the magnetic moments of the atom can have  $(2J+1)$  different orientations relative to the field.

The potential energy of the magnetic dipole in the presence of a magnetic field  $H$  is

$$U = -\vec{\mu} \cdot \vec{H} = m_J \mu_B g H_{\text{eff}} \quad \dots(5)$$

The number of atoms with a given value of  $m_J$  is proportional to  $\exp(m_J g \mu_B H_{\text{eff}} / kT)$ .

Consider a ferromagnetic solid containing  $N$  atoms per unit volume each having a total angular momentum number  $J$ . Then the magnetization in the direction of the field is given by

$$M = \frac{N \sum_{m_J=-J}^J m_J g\mu_B \exp(m_J g\mu_B H_{eff} / kT)}{\sum_{m_J=-J}^J \exp(m_J g\mu_B H_{eff} / kT)} \quad \dots(6)$$

Substituting  $\frac{g\mu_B H_{eff}}{kT} = x$  in the above equation, we get

$$M = Ng \mu_B \frac{\sum_{m_J=-J}^J m_J \exp(m_J x)}{\sum_{m_J=-J}^J \exp(m_J x)} \quad \dots(7)$$

After some algebraic manipulations, Eq. (7) reduces to,

$$M = Ng\mu_B \left[ \frac{2J+1}{2} \coth \frac{2J+1}{2}x - \frac{1}{2} \coth \frac{x}{2} \right]$$

Substituting  $x = a/J$  in the above equation, we get

$$\begin{aligned} M &= Ng\mu_B \left[ \frac{2J+1}{2} \coth \frac{2J+1}{2J}a - \frac{1}{2} \coth \frac{a}{2J} \right] \\ &= Ng\mu_B J \left[ \frac{2J+1}{2J} \coth \frac{2J+1}{2J}a - \frac{1}{2J} \coth \frac{a}{2J} \right] \end{aligned}$$

or

$$M = Ng J \mu_B B_J(a) \quad \dots(8)$$

Here  $B_J(a)$  is the *Brillouin function* and it is defined as

$$B_J(a) = \frac{2J+1}{2J} \coth \left( \frac{2J+1}{2J}a \right) - \frac{1}{2J} \coth \left( \frac{a}{2J} \right) \quad \dots(9)$$

From the two substitutions made in the above treatment, we have

$$\begin{aligned} \frac{a}{J} &= \frac{g\mu_B H_{eff}}{kT} \\ \text{or} \quad a &= \frac{gJ\mu_B H_{eff}}{kT} = \frac{g J \mu_B}{kT} (H + \lambda M) \end{aligned} \quad \dots(10)$$

For spontaneous magnetisation  $H = 0$ , Eq. (10) becomes

$$a = \frac{gJ\mu_B \lambda M}{kT} \quad \dots(11)$$

$$\therefore M(T) = \frac{akT}{gJ\mu_B \lambda} \quad \dots(12)$$

As  $T \rightarrow 0$  or  $a \rightarrow \infty$ ,  $B_J(a) \rightarrow 1$ ; the magnetic moments align themselves parallel to the field and the magnetization  $M$  becomes the *saturation magnetization*,  $M_S(0)$ . Thus, from Eq. (8), we get

$$M_S(0) = Ng J \mu_B \quad \dots(13)$$

From Eqs. (12) and (13), we obtain,

$$\frac{M(T)}{M_S(0)} = \frac{akT}{\lambda Ng^2 J^2 \mu_B^2} \quad \dots(14)$$

Eqs. (8) and (13) give

$$\frac{M(T)}{M_s(0)} = B_J(a) \quad \dots(15)$$

$\frac{M(T)}{M_s(0)}$  must simultaneously satisfy both equations (14) and (15). So its magnitude at a given temperature is obtained graphically as the intersection of the two  $\frac{M(T)}{M_s(0)}$  versus  $a$  plots (Fig. 30.40).

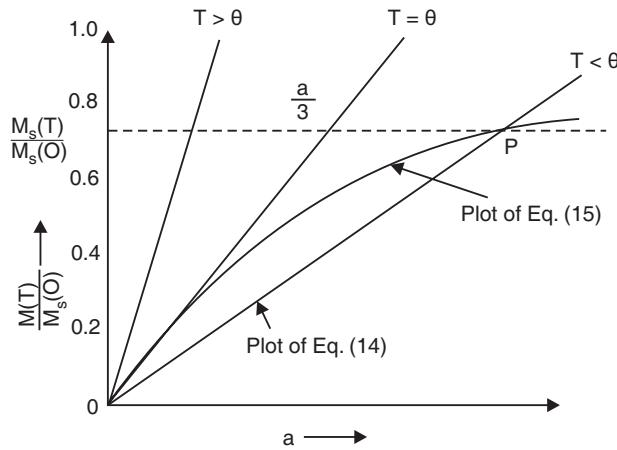


Fig. 30.40

- (i) For  $T < \theta$  (curie temperature), spontaneous magnetization results.
- (ii) For  $T \geq \theta$ , the two curves do not intersect and there is no spontaneous magnetization.

### ANTIFERROMAGNETISM

#### 30.15 ANTIFERROMAGNETIC MATERIALS

- Antiferromagnetic materials are crystalline materials which possess a small positive susceptibility of the order of  $10^{-3}$  to  $10^{-5}$ .
- Consider an antiferromagnetic crystal containing two types of atoms  $A$  and  $B$  distributed over two interpenetrating lattices (Fig. 30.41).

$B$  atoms occupy the corner points of a body centered cubic lattice.

$A$  atoms occupy the body centres of these cubic lattices.  
The atom at  $A$ -site has its spin up.

The atom at  $B$ -site has its spin down (Fig. 30.42).

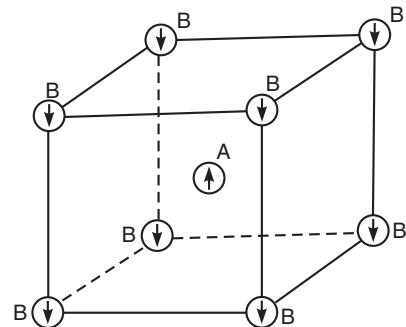


Fig. 30.41

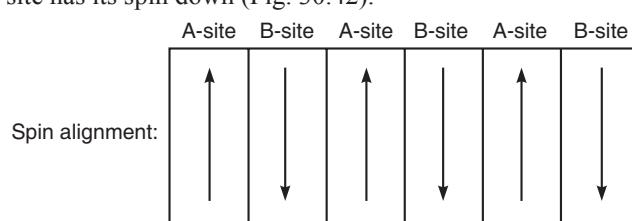


Fig. 30.42

The spin alignment of adjacent sites (*A* sites and *B* sites) is in an antiparallel manner. Hence their intensity of magnetisation is equal and opposite resulting in zero net magnetisation at  $T = 0\text{K}$ . Thus the susceptibility of antiferromagnetic materials at 0K is zero.

Above 0 K, the alignment of spin magnetic moments in *A* sites and *B* sites are varying in a different manner. This will lead to a positive and small value of susceptibility for these materials.

**Examples.** FeO (ferrous oxide), MnO (Manganese oxide),  $\text{Cr}_2\text{O}_3$  (Chromium oxide) and salts of transition elements.

#### Definition of Antiferromagnetism

- If the *A* and *B* sublattice moments are exactly equal but opposite, the *net moment is zero*. This type of magnetic ordering is called *antiferromagnetism*.

In antiferromagnetism, the magnetic moments of sublattices in crystal cell are equal in magnitude but opposite in direction. So they cancel each other giving rise to net zero magnetization.

#### Variation of susceptibility with temperature of an antiferromagnetic material

Figure 30.43 shows the variation of susceptibility with temperature.

- The susceptibility increases with increasing temperature and it reaches a maximum at a certain temperature called *Neel temperature*,  $T_N$ .

$$\chi \propto T \text{ when } T < T_N$$

The material is *antiferromagnetic* below  $T_N$ .

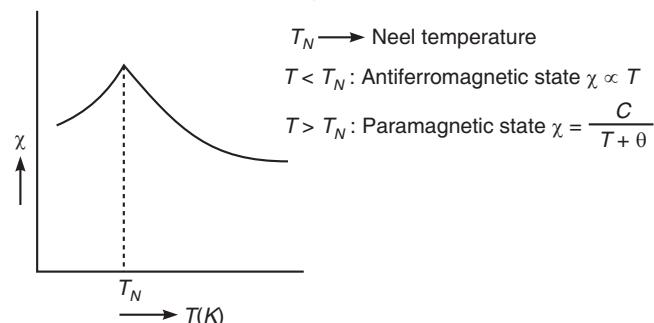


Fig. 30.43

- With further increase in temperature, the material reaches paramagnetic state. Above Neel temperature, the susceptibility decreases with temperature.

$$\chi = \frac{C}{T + \theta}$$

Here,  $C$  is the *Curie constant* and  $\theta$  the *paramagnetic Curie temperature*.

#### Properties of Antiferromagnetic materials

- (1) Electron spin of neighbouring atoms are aligned antiparallel; i.e.,

Spin alignment is antiparallel (Fig. 30.44).

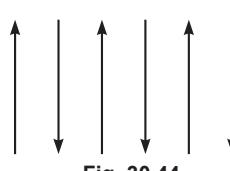


Fig. 30.44

- (2) Antiferromagnetic susceptibility depends greatly on temperature.

Magnitude of susceptibility	Temperature dependence	Examples
Small, positive	$\chi \propto T$ when $T < T_N$ $\chi = \frac{C}{T + \theta}$ when $T > T_N$	FeO (ferrous oxide), MnO (Manganese oxide), $\text{Cr}_2\text{O}_3$ (chromium oxide), salts of transition elements.

## FERRIMAGNETISM

### 30.16 FERRITES AND THEIR APPLICATIONS

- Ferrimagnetism is a special case of antiferromagnetism in which the opposite magnetic moments on each of the two sub-lattices are not exactly equal. Thus, when spontaneous anti-parallel arrangement of spins occurs, the material has a net paramagnetic moment and hence, a net magnetization too. This is called *ferrimagnetism*.

In *ferrimagnetic materials* (also called ferrites) such as  $\text{MnFe}_2\text{O}_4$ , the magnetic moments of adjacent ions are antiparallel and of unequal strength (Fig. 30.45). So there is a finite net magnetisation. Ferrites exhibit spontaneous magnetization below the ferromagnetic Neel temperature  $T_{FN}$ .

The general chemical formula of ferrites may be written as  $XY_2Z_4$ , where  $X$  is a divalent negative ion,  $Y$  is  $\text{Fe}^{3+}$ , and  $Z$  is mostly the divalent oxygen ion,  $\text{O}^{2-}$ . A familiar example is  $\text{Fe}_3\text{O}_4$  (or  $\text{Fe}^{2+}\text{Fe}^{3+}\text{O}^{2-}$ ).



Fig. 30.45

**Structure of Ferrites:** A ferrite crystal has the *spinel* structure.

The mineral *spinel* is  $\text{MgAl}_2\text{O}_4$ .

Figure 30.46 shows the regular spinel structure.

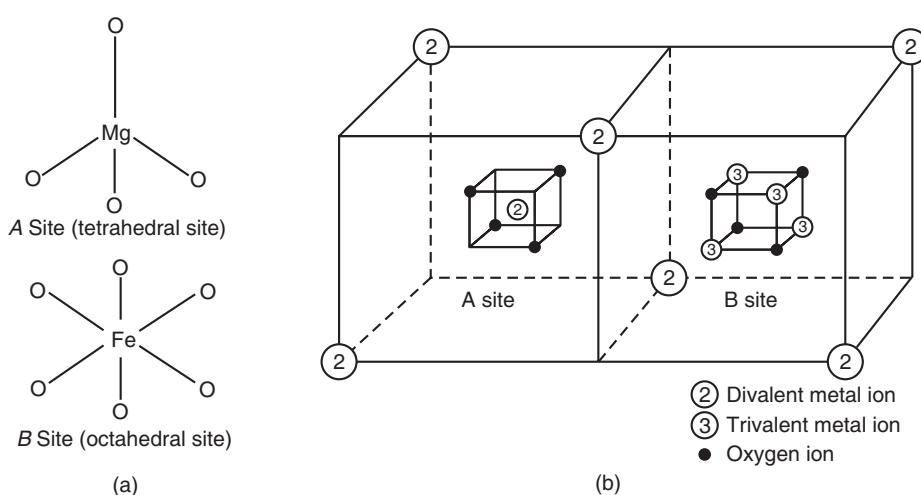


Fig. 30.46

(1) Each divalent metal ion is surrounded by  $\text{O}^{2-}$  ions in a *tetrahedral* fashion. For example, in  $\text{Mg}^{2+} [\text{Fe}_2^{3+}] \text{O}_4^{2-}$ , the structure of  $\text{Mg}^{2+}$  is given in Fig. (a). It is called *A site*. There are 8 tetrahedral sites in the unit cell.

(2) Each trivalent metal ion ( $\text{Fe}^{3+}$ ) is surrounded by 6  $\text{O}^{2-}$  ions in the *octahedral* fashion as shown in Fig. (a). It is called *B site*. There are 16 octahedral sites in the unit cell.

- Thus in a regular spinel, each divalent metal ion ( $\text{Mg}^{2+}$ ) exists in a tetrahedral site (*A site*) and each trivalent metal ion ( $\text{Fe}^{3+}$ ) exists in an octahedral site (*B site*).
- Hence, the sites *A* and *B* combine together to form a regular spinel ferrite structure as shown in Fig. (b).

## Properties:

1. These are metal oxides, but not metals.
  2. These materials exhibit hysteresis property.
  3. They are insulators with very high electrical resistivity ( $\sim 10^{12} \Omega \text{ cm}$ ). As a consequence, there will be no eddy current loss as usually noted with iron, at high frequencies.
  4. High microwave dielectric constant ( $\sim 10$  to  $12$ ) (low dielectric loss).
  5. A high magnetic permeability ( $500 - 1000$ ) for the mixed Ni-Zn ferrites ( $\text{Ni}_{0.36}\text{Zn}_{0.64}\text{Fe}_2\text{O}_4$ ) at low frequencies and  $\sim 10$  at high frequencies ( $\geq 300 \text{ MHz}$ ).

## **Applications of Ferrites**

1. Ferrites are used to produce low frequency ultrasonic waves by magnetostriction principle. Further these are used in the electromechanical transducers.
  2. Ferrite rods are used in radio receivers (particularly in medium wave coil) to increase the sensitivity and selectivity of receiver.
  3. Ferrites like Nickel Zinc ferrites are used as cores in audio and T.V. transformers.
  4. Since for ferrites eddy current loss and hysteresis loss are small at microwave frequencies, these are widely used in non-reciprocal microwave devices like *gyrator*, *circulator* and *isolator*.
  5. Ferrites are also used in digital computers and data processing circuits. Normally here ferrites with rectangular hysteresis loops are used as magnetic storage elements.
  6. Based on nonlinear tensor permeability property, ferrites can be used in devices for power limiting and harmonic generation.

## **EXERCISE**

8. Domain formation is necessary feature of \_\_\_\_\_  
 (a) diamagnetism      (b) paramagnetism      (c) ferrimagnetism      (d) ferromagnetism  
 (B.U. 2014)
9. At Curie temperature, the spontaneous magnetization for ferromagnetic material is \_\_\_\_\_  
 (a) one      (b) infinity      (c) zero      (d) none of these  
 (B.U. 2012)
10. Ferrites are  
 (a) ferromagnetic material      (b) ferrimagnetic material  
 (c) paramagnetic material      (d) antiferromagnetic material  
 (B.U. MSc. 2006)
- [Ans. 1. (a), 2. (b), 3. (c), 4. (c), 5. (c), 6. (a), 7. (d), 8. (d), 9. (c), 10. (b)]
11. Explain in detail Langevin's theory of diamagnetism. Find an expression for diamagnetic susceptibility.  
 (B.U. 2012)
12. Distinguish between para and diamagnetism.  
 (B.U. 2011)
13. Explain the Langevin's theory of paramagnetism.  
 (B.U. 2013)
14. Obtain an expression for paramagnetic susceptibility using Langevin's theory.  
 (B.U. 2010)
15. Discuss the origin and nature of Weiss molecular field.  
 (B.U. M.Sc. 2006)
16. Give an account of the quantum theory of paramagnetism and derive an expression for the susceptibility. How does this theory account for the experimentally observed susceptibility of the rare earth ions?  
 (Raj., 1978)
17. Explain the quantum theory of Paramagnetism.  
 (B.U. 2014)
18. Obtain an expression for  $\chi_m$  from quantum theory of paramagnetism.  
 (B.U. M.Sc. 2006)
19. Describe the process of cooling by adiabatic demagnetization of paramagnetic salts.  
 (B.U. 2015)
20. Differentiate between paramagnetism and ferromagnetism.  
 (Madras 2006)
21. Explain the domain theory of Ferromagnetism.  
 (B.U. 2012)
22. Explain the domain structure in ferromagnetic materials.  
 (Madras, 2006)
23. Discuss the Weiss theory of ferromagnetism.  
 (B.U. 2011)
24. Explain the nuclear magnetic resonance.  
 (B.U. 2013)
25. How was Weiss field in ferromagnets explained by Heisenberg?
26. Define Neel temperature.  
 (Madras 2006)
27. Outline any five applications of ferrites.  
 (B.U. M.Sc. 2006)
28. Give an account of quantum theory of ferromagnetism.
29. Explain Ferrimagnetism. What are ferrites? Discuss their applications.  
 (H.P.U. 1992)

**P A R T – VII**

# **FRONTIERS OF PHYSICS**

**31. Astrophysics (Large scale frontier of physics)**

**32. Nanotechnology (Small-scale frontier of physics)**

# ASTROPHYSICS

**After reading this chapter, you should be able to**

- ◆ Develop a basic knowledge about the solar system, stars, galaxies, the universe, and understanding of relevant physical processes
- ◆ Explain the birth and death of stars
- ◆ Derive an expression for gravitational potential energy of a star
- ◆ Derive an expression for the average temperature of a star
- ◆ Derive an expression for the core pressure of a star

### 31.1 INTRODUCTION

Astrophysics is a branch of astronomy that treats of the physical properties of celestial bodies, such as luminosity, size, mass, density, temperature, and chemical composition, and their origin and evolution. Astrophysics is primarily an observational subject. Among many other accomplishments, it has provided a convincing picture of the birth, life and death of stars—stellar evolution and the internal structure of stars.

Our knowledge of celestial objects such as the sun, the comets, the stars and the galaxies is mainly based on the analysis of the electromagnetic radiation which arrives to us from these objects.

### 31.2 CLASSIFICATION OF STARS — THE HARVARD CLASSIFICATION SYSTEM

Absorption lines were present in the spectra of all the stars (A Continuous spectrum was present too). When the absorption spectra of stars were first studied, it was realized that stars could be classified into several different types, called *spectral classes*. At first, the ‘Harvard Classification System’ was proposed. In this system, the stars were classified according to the strength of the hydrogen lines in their spectra, and were lettered alphabetically. Class *A* corresponds to the stars having the strongest hydrogen lines, class *B* the next strongest, and so on. It was later realized that the types of spectra varied primarily because of differing temperatures of the stellar atmospheres. So the spectral classes correspond in fact to different surface temperatures. When we now list the spectral types of stars in order of decreasing temperature, they are no longer in alphabetical order. In order of decreasing temperature, the spectral types are *O B A F G K M*. The spectral types can be remembered by the mnemonic : *Oh, Be, A Fine Girl, Kiss Me*. Each class is subdivided into 10 subclasses. For example, the hottest *B* stars are *BO*, followed by *B1, B2, B3 ... B9*. Table 31.1 describes the main characteristics of the spectral classes of stars.

**Table 31.1. Characteristics of the spectral classes of stars**

Spectral class	Temperature range (K)	Main characteristics of the spectrum	Colour	Typical example
O	30000–50000	Ionized helium lines dominant	Blue	10 Lacertae a Quasar, Zeta Puppis
B	10000–30000	Neutral helium lines dominant, no ionized helium	Blue or Bluish White	Rigel, Spica
A	7500–10000	Hydrogen lines dominant	White	Sirius, Deneb, Vega
F	6000–7500	Ionized calcium lines, Many metal lines (manganese, iron, titanium, strontium)	White	Canopus, Polaris
G	4500–6000	Very large number of metal lines. Strong ionized calcium lines, ionized and neutral iron	Yellow	Sun, Alpha Centauri
K	3500–4500	Large number of neutral metal lines	Red	Arcturus, Aldebaran
M	2000–3500	Band spectra of molecules (particularly of titanium oxide molecule)	Red	Betelgeuse, Antares

Today we know that the stars are nearly uniform in composition, being composed mainly of hydrogen and helium. the Indian astrophysicist M.N. Saha suggested that the differences in stellar spectra are principally due to different surface temperatures of stars. *M*-type star is the coolest. the gases occur in the form of molecules and molecular band spectra are observed. The spectra of *K* and *M* stars are dominated by molecular bands because they do not have sufficient temperatures to split molecule into atoms. With the increase in temperatures, molecules break into neutral atoms and ions. metals have low ionisation potential. Thus, ionised metal lines are dominant in *F* and *G* type stars. On further increase of temperatures, hydrogen is excited and the spectra of *A* type stars show hydrogen lines. At extremely high temperatures, even helium is ionised. Therefore, lines corresponding to ionised helium are present only in *O* — type stars.

### 31.3 HERTZSPRUNG — RUSSEL DIAGRAM

It is a graphical representation of the absolute magnitude of stars plotted against the spectral class (Fig. 31.1). The y-axis then represents the energy output of the star (luminosity) and the x-axis its surface temperature. The temperature is plotted from right to left. The majority of stars on such a diagram fall on a band running from the top left to the bottom right of the graph. These are called *main-sequence stars*. At the left and below the main sequence we have a group of stars called *white dwarfs*. The *giants* fall in a cluster above the main sequence and the *supergiants* are above the giants. The H-R diagram forms the basis of the theory of stellar evolution. It has been established that the different groups of stars in the diagram represent different stages in the evolution of a star.

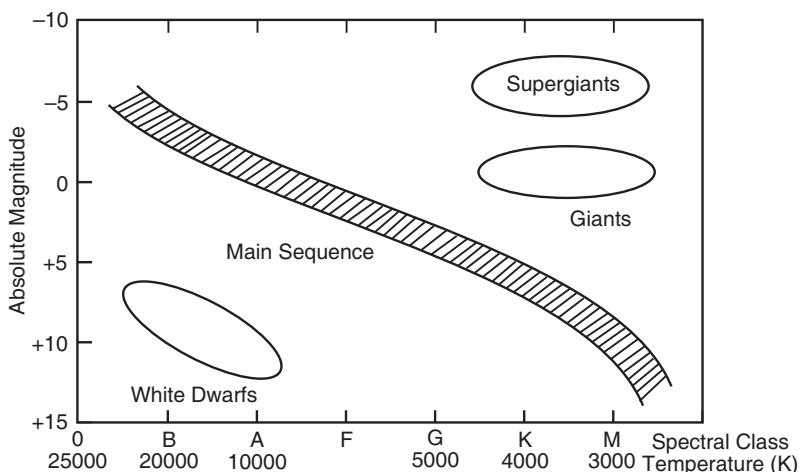


Fig. 31.1

### 31.4 LUMINOSITY OF A STAR

The total amount of light energy given out by a star in all directions per unit time is called luminosity  $L$ . By Stefan's law,

$$L = 4\pi R^2 \sigma T^4 \quad \dots(1)$$

Here  $R$  = radius of the star,  $T$  = surface temperature of the star and  $\sigma$  = Stefan constant.

Let  $d$  be the distance of the star from us.

$$\left. \begin{array}{l} \text{Energy flux on a unit area} \\ \text{at the observer's eye} \end{array} \right\} = f = \frac{L}{4\pi d^2} \quad \dots(2)$$

Each star is assigned a number called its magnitude, on the basis of its apparent brightness. Since it tells how bright a star *appears*, it is called apparent magnitude ( $m$ ).

The relation between  $m$  and  $f$  is

$$m = C - 2.5 \log_{10} f \quad \dots(3)$$

Here,  $C$  is a constant which determines the zero of the magnitude scale.

Substituting for  $f$  from Eq. (2), we get

$$m = (C + 2.5 \log_{10} 4\pi) - 2.5 \log_{10} L + 5 \log_{10} d \quad \dots(4)$$

Thus, the apparent magnitude ( $m$ ) of a star depends on its intrinsic luminosity ( $L$ ) and its distance ( $d$ ).

### 31.5 STELLAR EVOLUTION

The life cycle of a one solar mass star such as the sun is expected to follow the following stages: star formation, proto-star stage, pre-main sequence stage, main sequence stage, red giant stage, white dwarf stage.

#### Birth of a Star

The formation of stars is supposed to be the consequence of condensation of the primordial hydrogen under the effect of gravitational attraction. During their random motion in space a few atoms of hydrogen may at some instant of time be converging towards some point. When they are near each other they may exert a small gravitational attraction on each other. If this attraction is sufficiently strong, the atoms would be prevented from flying apart. They would form a pocket of gas. For this to happen the number of atoms converging towards a point must be really very large. Such a pocket of hydrogen atoms forms a beginning of a star. It is called a *protostar*.

Every atom in a protostar is pulled towards the centre of gravity. The protostar as a result shrinks in size. As the atoms fall towards the centre their speed increases. The temperature of the gas increases from about 100 K to about 50,000 K in about a year. At this temperature the hydrogen atoms collide with each other with sufficient violence to dislodge the electrons from the atoms. Instead of the original hydrogen gas there are now two gases, one of protons and the other of electrons. Because of the heat generated the protostar now becomes luminous. But most of the radiation is in the infra-red region. The shrinkage continues with rising temperature.

The shrinkage of the protostar is at first rapid. It then slows down because the density gradually increases and the atoms moving towards each other meet with more resistance. After 10 million years a protostar will have shrunk from its original size of trillions of kilometre to about 2.5 million kilometre in diameter. The internal temperature of the protostar will then be about  $10^6$  K.

At this temperature, thermonuclear conversion of hydrogen to helium begins. The loss of energy from the outer surface is then just compensated by the total nuclear energy production within the whole interior of the star. When this stage is reached, the star is said to have reached the *main-sequence*. A star is said to be born at this stage. The surface temperature of the new born star is about 3600 K.

The Hertzberg-Russell diagram is the graph obtained by plotting the luminosity of stars (in

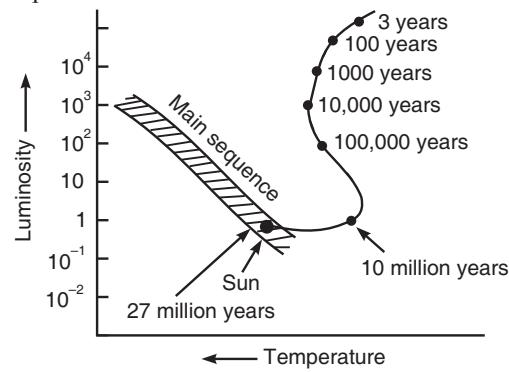


Fig. 31.2

terms of the luminosity of the sun) against the surface temperature of the stars. The track of a new star of solar mass across the H-R diagram is shown in Fig. 31.2. The age of the star is reckoned from the instant it enters the main sequence.

### Maturity

During the period the star remains in the main sequence, the release of nuclear energy at the centre balances the inward gravitational force. In the process of hydrogen burning, helium is continually formed as a result of the consumption of hydrogen in its core. This proceeds either by the proton-proton chain reactions or by the carbon-nitrogen cycle. As nuclear energy is released at the centre, the temperature of the star rises. The star becomes more and more luminous. The release of nuclear energy at the centre balances the inward gravitational force. This process of fusion of hydrogen into helium dominates 99% of the life of a star of a solar mass. During this period, the star remains in the main sequence in the H-R diagram.

### Ageing stars

The helium formed by fusion collects at the centre. Now there is no burning of hydrogen at the centre. The core contracts. The outer layer expands outwards. As the size of the star increases, the amount of energy radiated per unit area of its surface decreases. As a result, the surface temperature of the star decreases to a value between 3000 K to 4000 K. The temperature drop changes the colour from white to red. The star now leaves the main sequence and heads towards the giants stage. Fig. 31.3 shows the track of a star to the red giant region in the H-R diagram.

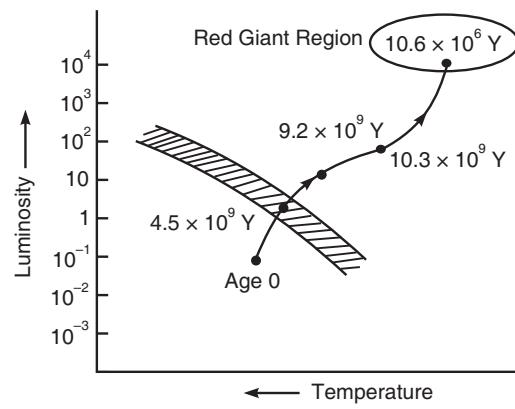


Fig. 31.3

### Death of a Star

When the temperature of the helium core reaches  $10^8$  K, helium nuclei start fusing into carbon through the triple-alpha process. But a star of low mass relatively soon runs out of helium and the core collapses into a *white dwarf*, while the outer regions drift away into space. Larger stars have sufficient helium for the process to continue so that heavier elements, up to iron are formed. But iron is the heaviest element that can be formed with the production of energy. When the helium has all been consumed there is a catastrophic collapse of the whole massive structure, resulting in a supernova explosion. Thereafter the burnt out remnants become neutron stars and black holes.

## 31.6 WHITE DWARFS

*A small star consisting of elements lighter than iron which has reached the stage where no further nuclear burning is possible is called a white dwarf.* Perhaps 10% of the stars in our galaxy are white dwarfs. These are stars in the final stages of their evolution with original masses that were less than about 7 solar masses. After the nuclear reactions that provided it with energy run out of fuel, such a star becomes unstable and eventually throws off its outer layer. The remaining core then cools and contracts gravitationally until its atoms collapse into nuclei and electrons packed closely together. A typical white dwarf has a mass of two-thirds that of the sun but is only about the size of the earth. The density of the white dwarf is quite large. The white dwarf is small and is therefore not easy to detect. The first white dwarf to be detected is the so-called Companion of Sirius.

### Electrons in a White Dwarf Star

Despite the high temperatures found in stellar interiors, the electrons in a white dwarf are quite degenerate because the value of Fermi temperature  $T_F$  is so large. When a star burns all its hydrogen fuel, it is made up mainly of a plasma of helium nuclei and electrons. The gravitational forces between the helium nuclei cause the star to collapse until the Pauli pressure of the electrons given by

$$P = \frac{2}{5} \rho \epsilon_F(0) \quad \dots(1)$$

brings a halt to the collapse.

Here,  $\rho$  = Density of the gas,  $\epsilon_F(0)$  = Fermi-energy at absolute zero.

The negative (inward) gravitational pressure is

$$P_G \sim \frac{E_G}{V} \sim \frac{GM^2}{R} \frac{1}{R^3} \quad \dots(2)$$

Here,  $E_G$  is the gravitational potential energy,  $G$  is the gravitational constant,  $R$  is the radius of the star, and  $M \sim N m_{He}$  is the mass of the star made up of  $N$  helium atoms each of mass  $m_{He}$ .

The outward Pauli pressure is given by

$$P_P \sim \frac{N \hbar^2}{R^3 m_e} \left( \frac{N}{R^3} \right)^{2/3} \quad \dots(3)$$

Here,  $m_e$  is the electron mass and the number of electrons is actually  $2N$ , twice the number of helium nuclei. Equating  $P_P$  and  $P_G$  gives the radius of the star.

$$R \sim \frac{\hbar^2}{G m_e m_{He}^{5/3} M^{1/3}} \quad \dots(4)$$

We have dropped all numerical constants from this expression to get just an order of magnitude estimate. If we put in numbers we get  $R = 7 \times 10^2$  km for  $M = M_0$ , the solar mass. This means a white dwarf is a very compact object.

Eq. (4) is derived on the basis of nonrelativistic theory. But the electron system of the white dwarf star must be treated relativistically. The relativistic treatment yields the result that, for stellar masses greater than about  $1.4 M_0$ , the Pauli pressure can no longer stop the gravitational collapse. The star continues its infall, a supernova occurs, and ultimately a neutron star or a black hole is formed. This critical mass is known as the *Chandrasekhar limit*.

### 31.7 CHANDRASEKHAR LIMIT

The core of a star that has exhausted its nuclear fuel can end up in different ways depending on its mass. In the case of a star like our sun as a result of gravitational compression the core is left with protons, with electrons moving around like the molecules of a gas. This electron gas is able to withstand the gravitational force. As a result a stable equilibrium can be achieved. Such a star is called a white dwarf.

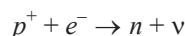
White dwarfs are quite different from all the usual stars. The particles of *classical gases* obey *classical mechanics*. The particles of *quantum gases* obey *quantum mechanics*. Quantum gases help us to understand the behaviour of white dwarfs.

Fowler regarded the core of a white dwarf as a Fermi gas. The pressure in such a gas is of quantum mechanical origin. It is called degeneracy pressure, and has its origins in the Pauli principle. According to Fowler, it was degeneracy pressure which arrested the total collapse of a white dwarf. Fowler applied quantum mechanical ideas to stars.

When the white dwarf gradually cools, at some stage it ceases to be self luminous and ends up as a black dwarf. S. Chandrasekhar discovered theoretically in the 1930s that such a stable condition can occur only for cores which have masses up to  $1.4 M_0$  where  $M_0$  is the mass of the sun. This is known as the *Chandrasekhar limit*. If the mass of the core exceeds the Chandrasekhar limit, the gravitational compression becomes so large that the electrons are forced into the nuclei. This process is known as inverse beta decay. The star is now composed only of neutrons. These neutrons can withstand the gravitational compression. We now have a neutron star.

### 31.8 NEUTRON STARS

Let us now discuss the fate of the core of a star that explodes as a supernova. As iron fills the core of a massive star, the temperatures are so high that the iron nuclei begin to break apart into smaller units like alpha particles (helium nuclei). The pressure is no longer high enough to counteract gravity and the core collapses. As the density increases, the electrons are squeezed into the nuclei and react with the protons there to produce neutrons and neutrinos.



The neutrinos escape. A gas composed mainly of neutrons is left behind in the dense core as the outer layers explode as a supernova. Now the core has a very high density. At this density a condition called *neutron degeneracy*, in which the neutrons cannot be packed any more tightly, appears. The pressure caused by neutron degeneracy balances the gravitational force that tends to collapse the core. As a result, the core reaches equilibrium as a *neutron star*.

Typically neutron stars have a radius of about 10 km, although in mass they are comparable with the sun. The chunk of their matter, the size of a small sugar cube, contains about  $10^8$  tons of neutrons. Neutron stars have masses between 1.4 and  $\sim 3 M_{\text{sun}}$ .

Before it collapses, the core has only a weak magnetic field. But as the core collapses, the magnetic field is concentrated, and grows stronger as a result. By the time the core shrinks to neutron star size, it has an extremely powerful magnetic field.

Neutron stars were postulated in the 1930s and many astronomical observations confirm their existence. Neutron stars are believed to be pulsars if they rotate sufficiently rapidly and have sufficiently strong magnetic fields.

**Pulsars.** First discovered in 1968, a pulsar is believed to be a rotating neutron star. They emit radiation from relatively a small spot on the star. This situation produces an effect rather like the beam of a lighthouse. The distant observer receives a pulse of radiation as the beam sweeps across his position on each rotation of the star.

A pulsar gives off pulses of radio waves. A very few objects (example : Crab and the Vela pulsars) are known to pulse  $\gamma$ -rays, X-rays, or visible light. The periods of known pulsars range between 33 milliseconds and 3.75 seconds. Pulse durations range from 2 to about 150 milliseconds with longer period pulsars generally having a longer pulse duration.

### 31.9 BLACK HOLES

Consider stars with very large mass, say 5 to  $10 M_{\text{sun}}$ . For such a star, the contraction cannot be arrested either at the white dwarf stage or at the stage of the neutron star. A star may continue to collapse beyond the neutron star stage. When the radius of the star is of the order of 15 kilometre, relativity predicts a most extraordinary phenomenon. According to the theory of relativity, a ray of light should possess mass and hence be subject to gravity. A ray of light emitted by a star will therefore be pulled back by the star's gravity. If the star is large in size, gravity will not be strong at its surface and a ray of light will be able to escape from the star. If the star shrinks to a size of about

15 kilometre radius, the force of gravity at its surface will be billions of times stronger than the force of gravity at the surface of the sun. A ray of light trying to leave the star will therefore be pulled back and it cannot escape from the star. And when light cannot escape, nothing else can escape from it. The star then becomes invisible. It becomes a *black hole* in space.

The contraction continues inside a black hole. There is no force in nature which is strong enough to halt this contraction. The volume of the star must become smaller and smaller. The mass of the star however remains constant. Ultimately trillions and trillions of quintals of matter will have to be packed in a size less than a pinhead.

Observational evidence of objects thought to be black holes comes from their effect on surrounding matter. Thus, if a black hole forms a binary system with another star it will attract and capture matter from this star. The material leaving the star first forms a rotating disc around the black hole, in which the matter becomes compressed. The disc is heated to such an extent that it emits X-rays. In the constellation Cygnus there is an X-ray source, Cygnus X-1, which consists of a supergiant revolving around a small invisible companion with a mass ten times that of the sun. The companion is thought to be a black hole.

### The Basic Physics of Black Holes

Consider an escape velocity such that, when an object leaves with just the escape velocity, it will have zero velocity out at "infinity". There, its total energy is

$$TE = (KE + PE) = (1/2) mv^2 - GmM/R = 0.$$

Since total energy must be conserved, at the moment of launch we require

$$\begin{aligned} TE &= 0 = (1/2) mV_{\text{esc}}^2 - GmM/R \\ (1/2) V_{\text{esc}}^2 &= GM/R \\ V_{\text{esc}} &= (2GM/R)^{1/2} \end{aligned}$$

Now no object can travel faster than the speed of light, and so the maximum escape velocity is  $c$ . Then the equation for the black-hole radius is

$$R = 2GM/c^2.$$

This critical radius is called the *Schwarzschild radius*.

## 31.10 SUPERNOVA EXPLOSION

When a massive star starts rapidly collapsing, the interior gets heavily compressed and therefore also very hot. The core becomes converted into neutrons. If the compression is sudden and the heat generated is very high, then a violent explosion called *supernova explosion* takes place. It tears off the outer layers of the star and hurls them into space. What is left after a supernova explosion may be a remnant whose mass is greater than 1.4  $M_{\text{sun}}$ . The Crab Nebula is the nearest *supernova remnant*.

## 31.11 PHOTON DIFFUSION TIME

Photons are produced in the core of the star due to thermonuclear reactions. Photons move from the core towards outer surface of the star. The photons get scattered large number of times by free electrons before reaching the surface. This results in an apparent dance like motion of the photons inside the stars like the sun (Fig. 31.4).

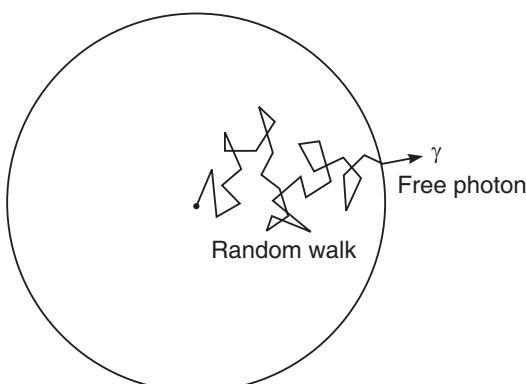


Fig. 31.4

This dance like motion of the photons is called *random walk*. The photon travels along a *zig-zag* path from the core of the sun to the sun's surface. Finally the photon escapes from the sun's surface.

During their random walk, the photons which are X-ray photons at the core become optical photons by the time they reach the surface of the star by *diffusion*.

The amount of energy radiated out per second from the surface of a star is its *luminosity*  $L$ .

The **photon diffusion** time is defined as the ratio of the total radiant energy  $E$  contained in the volume of the star to the luminosity  $L$  of the star. i.e.,

$$\tau = \frac{E}{L} \quad \dots (1)$$

The total amount of light energy given out by a star in all directions per unit time is called luminosity  $L$ . By Stefan's law,

$$L = 4\pi R^2 \sigma T^4 \quad \dots (2)$$

Here  $R$  = radius of the star,  $T$  = surface temperature of the star and  $\sigma$  = Stefan constant.

The energy density of the radiation which is a function of temperature  $T(r)$  is defined as

$$U = \frac{4\sigma}{c} T^4(r) \quad \dots (3)$$

Here,  $c$  = velocity of light.

According to the linear star model, the relation between the core temperature  $T_c$  and the varying temperature  $T(r)$  is given by

$$T(r) = T_c \left[ 1 - \frac{r}{R} \right] \quad \dots (4)$$

$$\therefore U = \frac{4\sigma}{c} T_c^4 \left( 1 - \frac{r}{R} \right)^4 \quad \dots (5)$$

$$\text{The total radiant energy} = E = \int_0^R U dV = \int_0^R \frac{4\sigma}{c} T_c^4 \left( 1 - \frac{r}{R} \right)^4 dV$$

But  $dV = 4\pi r^2 dr$ :

$$\begin{aligned} E &= \int_0^R \frac{4\sigma T_c^4}{c} \left( 1 - \frac{r}{R} \right)^4 4\pi r^2 dr = \frac{16\pi\sigma T_c^4}{c} \int_0^R \left( 1 - \frac{r}{R} \right)^4 r^2 dr \\ \therefore E &= \frac{1}{105} \left( \frac{16\pi\sigma}{c} \right) R^3 T_c^4 \end{aligned} \quad \dots (6)$$

For the sun,  $E = 1.4 \times 10^{39} J$  and  $T_c = 15 \times 10^6 K$ .

Substituting  $E$  and  $L$  from Eqs. (6) and (2) in Eq. (1), we get

$$\begin{aligned} \tau &= \frac{\frac{1}{105} \left( \frac{16\pi\sigma}{c} \right) R^3 T_c^4}{4\pi R^2 \sigma T^4} \\ \therefore \tau &= \frac{4R}{105c} \left( \frac{T_c}{T} \right)^4 \end{aligned} \quad \dots (7)$$

For the sun,  $\tau = 3.46 \times 10^{12} s = 1.2 \times 10^5$  years.

**Mass - Luminosity relation for a star**

$$\text{From Eq. (1), } L = \frac{E}{\tau} \quad \dots (8)$$

$$\text{From Eq. (6), } E \propto R^3 T_c^4 \quad \dots (9)$$

$$\text{It can be shown that } T_c \propto \frac{M}{R} \quad \dots (10)$$

Here,  $M$  is the mass of the star.

$$\therefore E \propto R^3 \left( \frac{M}{R} \right)^4 \propto \frac{M^4}{R} \quad \dots (11)$$

The mean free path ( $s$ ) of the diffusing photon is defined as the distance travelled by the photon between its two successive scatterings.

$$\text{The photon diffusion time } \tau \propto \frac{R^2}{s} \quad \dots (12)$$

The mean free path ( $s$ ) is inversely proportional to the mean photon density. i.e.,

$$\begin{aligned} s &\propto \frac{1}{\rho} \propto \frac{V}{M} \propto \frac{R^3}{M} \\ \therefore \tau &\propto \frac{R^2 M}{R^3} \propto \frac{M}{R} \end{aligned} \quad \dots (13)$$

Substituting for  $E$  and  $\tau$  from Eqs. (11) and (13) in Eq. (8), we get

$$\begin{aligned} L &\propto \frac{M^4}{R} \times \frac{R}{M} \\ \therefore L &\propto M^3 \end{aligned} \quad \dots (14)$$

*The life time of a star is defined as the ratio between the mass and the luminosity of a star.*

$$\text{Life time} \propto \frac{M}{L} \propto \frac{M}{M^3}$$

$$\therefore \text{Life time of a star} \propto \frac{1}{M^2}$$

*The life time of a star is inversely proportional to the square of its mass.*

### 31.12 GRAVITATIONAL POTENTIAL ENERGY OF A STAR

The formation of a star is due to the condensation of the primordial hydrogen under the effect of gravitational attraction. Consider a star of mass  $M$  and radius  $R$  being formed by a slow gravitational contraction of hydrogen gas (Fig. 31.5).

Initially the gas particles are at an infinite distance. Let  $dM(r)$  be the mass of a cloud of spherical shell of thickness  $dr$ , internal radius  $r$  and density  $\rho(r)$  over a sphere of mass  $M(r)$  and radius  $r$ .

$$dM(r) = 4\pi r^2 \rho(r) dr \quad \dots (1)$$

The work done to bring this mass  $dM(r)$  and add it to the sphere of mass  $M(r)$  is given by the gravitational potential energy,

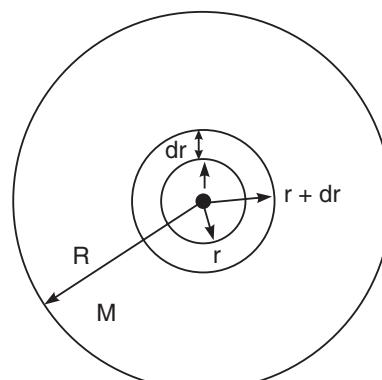


Fig. 31.5

$$dU = -\frac{G M(r) dM(r)}{r} \quad \dots (2)$$

The total gravitational energy required to form the entire star is given by,

$$U = - \int_0^R \frac{G M(r) dM(r)}{r} \quad \dots (3)$$

(i) According to **constant density model of a star**, density is constant throughout the interior of a star.

$$\therefore M(r) = \int_0^r 4\pi r^2 \rho dr = \frac{4}{3} \pi r^3 \rho \quad [\text{from Eq. (1)}]$$

Substituting for  $dM(r)$  and  $M(r)$  in Eq. (3), we have

$$\begin{aligned} U &= -G \int_0^R \frac{\left(\frac{4\pi r^3}{3}\rho\right)(4\pi r^2 \rho dr)}{r} = -G \frac{16\pi^2 \rho^2}{3} \int_0^R r^4 dr = -G \frac{16\pi^2 \rho^2}{15} R^5 = -G \frac{3}{5R} \left(\frac{4}{3} \pi R^3 \rho\right)^2 \\ &M = \frac{4}{3} \pi R^3 \rho, \text{ where } \rho \text{ is the average density.} \\ \therefore U &= -\frac{3}{5} \frac{GM^2}{R}. \end{aligned} \quad \dots (4)$$

Eq. (4) gives the gravitational potential energy of a star as per constant density model of a star.

(ii) According to the *linear star model*, density  $\rho(r)$  decreases from maximum  $\rho_c$  to 0.

$$\rho(r) = \rho_c \left(1 - \frac{r}{R}\right) \quad \dots (5)$$

Substituting this value of  $\rho(r)$  in Eq. (1), we have

$$dM(r) = 4\pi r^2 \rho_c \left(1 - \frac{r}{R}\right) dr \quad \dots (6)$$

$$M(r) = 4\pi \rho_c \int_0^r \left(r^2 - \frac{r^3}{R}\right) dr = 4\pi \rho_c \left[\frac{r^3}{3} - \frac{r^4}{4R}\right] = \frac{4}{3} \pi r^3 \rho_c \left(1 - \frac{3r}{4R}\right)$$

The total mass  $M$  of the star (when  $r = R$ ) is given by

$$\begin{aligned} M &= \frac{\pi R^3}{3} \rho_c \\ \text{or} \quad \rho_c &= \frac{3M}{\pi R^3} \end{aligned} \quad \dots (7)$$

But average density is given by  $\rho = \frac{3M}{4\pi R^3}$

$$\therefore \rho_c = 4\rho \quad \dots (8)$$

Substituting for  $dM(r)$  and  $M(r)$  in Eq. (3),

$$U = -G \int_0^R \frac{\frac{4\pi}{3} \rho_c r^3 \left(1 - \frac{3r}{4R}\right) \times 4\pi r^2 \rho_c \left(1 - \frac{r}{R}\right) dr}{r} = -\frac{G(4\pi\rho_c)^2}{3} \int_0^R r^4 \left(1 - \frac{3r}{4R}\right) \left(1 - \frac{r}{R}\right) dr$$

Substituting for  $\rho_c$ ,

$$U = -\frac{G}{3} \left( \frac{4\pi 3M}{\pi R^3} \right)^2 \int_0^R r^4 \left( 1 - \frac{3r}{4R} \right) \left( 1 - \frac{r}{R} \right) dr = -\frac{144}{3} \frac{GM^2}{R^6} \left[ \frac{312R^5}{20160} \right] = -\frac{26}{21} \left( \frac{3}{5} \frac{GM^2}{R} \right)$$

But  $\frac{26}{21} \approx \frac{5}{4}$ .

$$\therefore U = -\frac{3}{4} \frac{GM^2}{R}. \quad \dots (9)$$

### 31.13 INTERNAL TEMPERATURE OF A STAR

A contracting star radiates energy just like a black body and loses energy obeying Stefan-Boltzmann law ( $E = \sigma T^4$ ). To uphold the law of conservation of energy, half the loss in potential energy  $U$  of the star is gained by the star itself in the form of kinetic energy  $K$  and the remaining half of the energy is lost by radiation.

$$\therefore K = U/2 \quad \dots (1)$$

The interior of a star behaves like an ideal gas. According to kinetic theory of gases, the kinetic energy of the interior gas is given by

$$K = \frac{3}{2} Nk T_{av} \quad \dots (2)$$

Here,  $N$  is the number of gas particles,  $k$  is the Boltzmann constant and  $T_{av}$  is the average temperature of the entire stellar matter.

Let  $m$  = mass of each gas particle and  $M$  = mass of entire star. Then,

$N$  = total number of particles contained in the star =  $M/m$ .

$$\therefore K = \frac{3}{2} \left( \frac{M}{m} \right) k T_{av} \quad \dots (3)$$

According to *Virial theorem*, the kinetic energy  $K$  is equal to the total thermal energy of the star and is given by

$$K = \frac{3}{8} \frac{GM^2}{R} \quad \dots (4)$$

Comparing Eqs. (3) and (4),

$$T_{av} = \frac{GMm}{4kR}$$

If  $\mu$  is the mean molecular weight in terms of atomic mass of hydrogen  $m_H = 1.67 \times 10^{-27}$  kg, then  $m = \mu m_H$ .

$$\therefore T_{av} = \frac{GM\mu m_H}{4kR} \quad \dots (5)$$

This is the first equation of stellar structure.

If all the hydrogen atoms in a star (sun) are considered to be completely ionized, then  $\mu = 0.5$ . Sun is a main sequence star. Its properties like surface temperature, mass, radius, etc., are well known.

The average temperature of any star of similar composition can be determined in terms of solar parameters using the relation

$$T_{av} = \left( \frac{M}{M_{sun}} \right) \left( \frac{R_{sun}}{R} \right) (T_{sun})_{av} \quad \dots (6)$$

Let  $T(r)$  be the internal temperature of the sun at a distance  $r$  from the centre. Let  $T_c$  = core temperature.

$$\text{According to the linear star model } T(r) = T_c \left[ 1 - \frac{r}{R} \right] \quad \dots (7)$$

Here,  $T_c = T_{av}$  at the centre of the star and  $r = R$  at the surface.

### 31.14 INTERNAL PRESSURE OF A STAR

Consider the given star to be made up of large number of concentric shells. Each shell is in *hydrostatic equilibrium*. Consider a shell of radius  $r$  and thickness  $dr$ . Let  $\rho(r)$  be the density of stellar matter in the shell.

$$\text{Mass of the shell, } dM(r) = 4\pi r^2 \rho(r) dr$$

$$\text{or } \frac{dM(r)}{dr} = 4\pi r^2 \rho(r) \quad \dots (1)$$

If the stellar mass distribution is constant, then Eq. (1) represents conservation of stellar mass. Therefore, the outward pressure on the shell from within the star is equal to the gravitational force acting on the shell towards the centre of the star.

The gravitational force between the shell of mass  $dM(r)$  and the stellar sphere of mass  $M(r)$  with radius  $r$  is given by

$$F_{gr} = \frac{-G M(r) dM(r)}{r^2} = \frac{-GM(r) 4\pi r^2 \rho(r) dr}{r^2} \quad \dots (2)$$

Let  $P(r)$  = outward gas pressure on the shell from within the star and

$P(r+dr)$  = inward gas pressure on the shell from above it (Fig. 31.6).

Net outward pressure acting on the shell is given by  
 $dP = P(r+dr) - P(r)$

The total outward force acting on the shell to balance it against the inward gravitational force is given by

$$F_{pr} = 4\pi r^2 dP \quad \dots (3)$$

For *hydrostatic balance*,

$$\begin{aligned} F_{pr} &= F_{gr} \\ \therefore \frac{dP}{dr} &= \frac{-G M(r) \rho(r)}{r^2} \end{aligned} \quad \dots (4)$$

Eq. (4) is known as the equation of *hydrostatic equilibrium* and is the *second equation of stellar structure*. All stable main sequence stars obey this equation.

If the stellar mass is considered as an ideal gas, the ideal gas equation is

$$PV = RT = N_A kT$$

Here,  $N_A$  = Avogadro number.

$$P = \frac{N_A}{V} kT = nkT$$

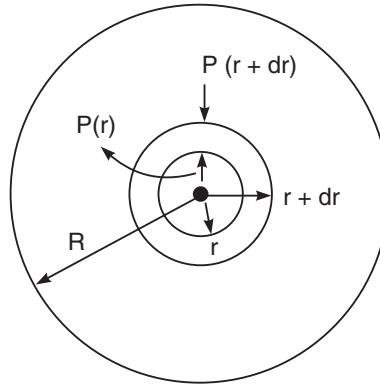


Fig. 31.6

Here,  $n = \frac{N_A}{V}$  is the molecular density.

Let  $\rho_{av}$  = average stellar density and  $\mu$  = mean molecular weight in the star.  
Molecular density,

$$n = \frac{\rho_{av}}{\mu m_H} = \frac{\frac{M}{\left(\frac{4}{3}\pi R^3\right)}}{\mu m_H} \quad \dots (5)$$

The average temperature of a star is given by

$$T_{av} = \frac{GM\mu m_H}{4kR} \quad \dots (6)$$

If the average pressure and temperature of the star are considered, then

$$P_{av} = nkT_{av} = \frac{\rho_{av}}{\mu m_H} kT_{av} \quad \dots (7)$$

Substituting for  $n$  and  $T_{av}$ , we have

$$P_{av} = \frac{3}{16\pi} \frac{GM^2}{R^4} \quad \dots (8)$$

Eq. (8) gives the average internal pressure of a star.

The average pressure for any star can be determined in terms of solar parameters using the relation

$$P_{av} = \left( \frac{M}{M_{sun}} \right)^2 \left( \frac{R_{sun}}{R} \right)^4 (P_{sun})_{av} \quad \dots (9)$$

From Eq. (7), the internal (or core) pressure is given by

$$P_c = \frac{\rho_c}{\mu m_H} kT_c \quad \dots (10)$$

The internal pressure of the Sun at a distance  $r$  from its centre is given by

$$P(r) = P_c \left[ 1 - \frac{r}{R} \right]. \quad \dots (11)$$

## EXERCISE

1. Write a note on chemical composition of a star and hence explain spectral classes and spectral types.
2. What is a HR diagram? Describe how HR diagram is used to account for the evolution of stars.
3. What is photon diffusion and random walk? Obtain an expression for the photon diffusion time.
4. Obtain an expression for the gravitational potential energy of a main sequence star.
5. Obtain an expression for the average temperature and core temperature of a main sequence star.
6. Obtain an expression for the internal pressure of a main sequence star.

## NANOTECHNOLOGY

**After reading this chapter, you should be able to**

- ◆ Explain methods of preparation of clusters
- ◆ Develop an overview of carbon nanotubes and their applications
- ◆ Discuss in detail various fabrication processes of nanomaterials
- ◆ Explain how the observation of materials in the nanoscale is done using STM and AFM

### 32.1 INTRODUCTION

The prefix *nano* in the word *nanotechnology* means a billionth ( $1 \times 10^{-9}$ ). Nanotechnology deals with various structures of matter having dimensions of the order of a billionth of a meter. A nanometer is  $10^{-9}$  m. Nanoparticles are generally considered to be a number of atoms or molecules bonded together with a radius of  $< 100$  nm. Generally, physical properties of materials can be characterized by some critical length, a thermal diffusion length, or a scattering length, for example. The electrical conductivity of a metal is strongly determined by the distance that the electrons travel between collisions with the vibrating atoms or impurities of the solid. This distance is called the *mean free path* or the *scattering length*. If the sizes of particles are less than these characteristic lengths, it is possible that new physics or chemistry may occur. *A working definition of a nanoparticle is an aggregate of atoms between 1 and 100 nm viewed as a subdivision of a bulk material, and of dimension less than the characteristic length of some phenomena.*

Materials with at least one of the dimensions measuring less than 100 nm are known as *nanomaterials*. *We define nanomaterials as those which have a characteristic length scale within about 100 nm.*

A particle diameter, grain size, layer thickness, or width of a conducting line on a device are some examples of length scale.

**EXAMPLE.** *The size of the Quantum dot is 1–5 nm.*

#### Special Features of Nanomaterials

Conventional materials have grains varying in size anywhere from hundreds of microns ( $\mu\text{m}$ ) to millimeters (mm). A nanocrystalline material has grains on the order of 1–100 nm. The properties of bulk materials are mostly retained till the reduction of their dimensions to the micrometer range. But materials in the nanometre scale show remarkably new properties.

**EXAMPLE.** 1. *A nanocrystal of gold serves as an excellent low temperature catalyst whereas the bulk gold does not.*

2. *Bulk semiconductors transform into insulators if their dimensions shrink to a couple of nanometres.*

3. *The melting temperature of crystals in the nanometre scale is very low, lower by 1000°C in some cases. The lattice constants are, too, reduced.*

4. *A nanoscale wire or circuit component does not necessarily obey the Ohm's law of electricity.*

### 32.2 NANOPARTICLES

A group of  $10^6$  or less number of atoms or molecules bonded together in a cluster with the radius of about 100 nm or less forms a nanoparticle. The metal oxide nanoparticles of different sizes in stained-glass windows produce different beautiful colours because a particle scatters only that wavelength which compares with its size. The understanding of structure and properties of nanoparticles is crucial for the interpretation of the characteristic phenomena exhibited by them. A brief description of some well-studied nanoparticles is given below.

### 32.3 METAL NANOCLUSTERS

#### 32.3.1 Magic Numbers

**Experimental set-up to form clusters of metal atoms.** Fig. 32.1 shows the apparatus to make metal nanoparticles by laser induced evaporation of atoms from the surface of a metal.

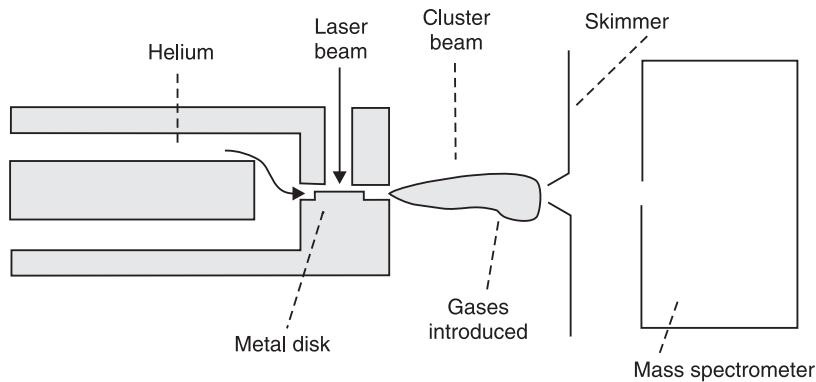


Fig. 32.1

A high-intensity laser beam is incident on a metal rod, causing evaporation of atoms from the surface of the metal. The atoms are then swept away by a burst of helium and passed through an orifice into a vacuum. In the evacuated chamber, the expansion of the gas causes cooling and formation of clusters of the metal atoms. These clusters are then ionized by UV radiation and passed into a mass spectrometer. The mass spectrometer measures their mass : charge ratio.

**Mass spectrum of Pb clusters:** Fig. 32.2 shows the mass spectrum data of lead clusters formed in the experiment. Here, the number of ions (counts) is plotted as a function of the number of atoms in the cluster. (Usually mass spectra data are plotted as counts versus mass over charge.) The data show that clusters of 7 and 10 atoms are more likely than other clusters which means that these clusters are more stable than clusters of other sizes.

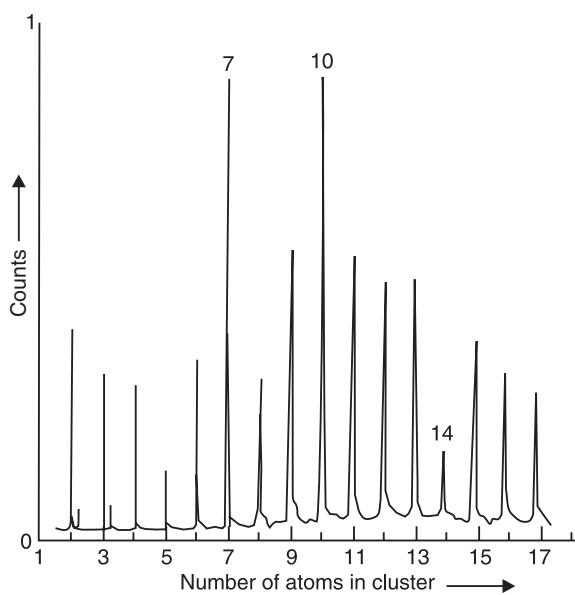


Fig. 32.2

**Ionization Potential:** Fig. 32.3 (a) is a plot of the ionization potential of atoms as a function of their atomic number  $Z$ . The maximum ionization potentials occur for the rare-gas atoms  ${}^2\text{He}$ ,  ${}^{10}\text{Ne}$  and  ${}^{18}\text{Ar}$  because their outermost  $s$  and  $p$  orbitals are filled. More energy is required to remove electrons from filled orbitals than from unfilled orbitals.

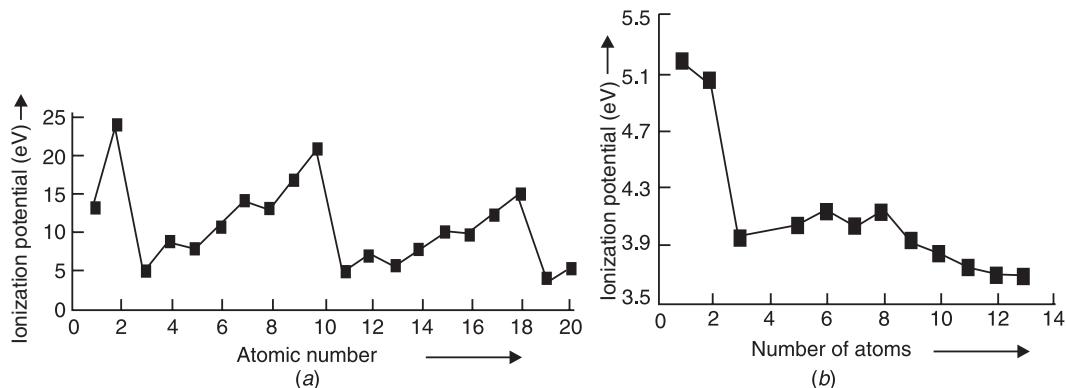


Fig. 32.3

Fig. 32.3 (b) shows the ionization potential of sodium clusters as a function of the number of atoms in a cluster. Peaks are observed at clusters having two and eight atoms. These numbers are referred to as *electronic magic numbers*. Their existence suggests that clusters can be viewed as superatoms. This result motivated the development of the jellium model of clusters. In the case of larger clusters, stability is determined by structure and the magic numbers are referred to as *structural magic numbers*.

### 32.3.2 Theoretical Modeling of Nanoparticles

There exist two well-known theoretical models to calculate the properties of clusters:

(i) the jellium model and (ii) the density functional model (based on the molecular orbital theory).

**(i) The jellium model.** The jellium model envisions a cluster of atoms as a large atom. The positive nuclear charge of each atom of the cluster is assumed to be uniformly distributed over a sphere the size of the cluster. A spherically symmetric potential well is used to represent the potential describing the interaction of the electron with the positive spherical charge distribution. Thus the energy levels can be obtained by solving the Schrödinger equation for this system in a fashion analogous to that for the hydrogen atom.

Fig. 32.4 compares the energy level scheme for the hydrogen atom and the energy-level scheme for a spherical positive-charge distribution. The superscripts refer to the number of electrons that fill a particular energy level. The electronic magic number corresponds to the total number of electrons on the superatom when the top level is filled.

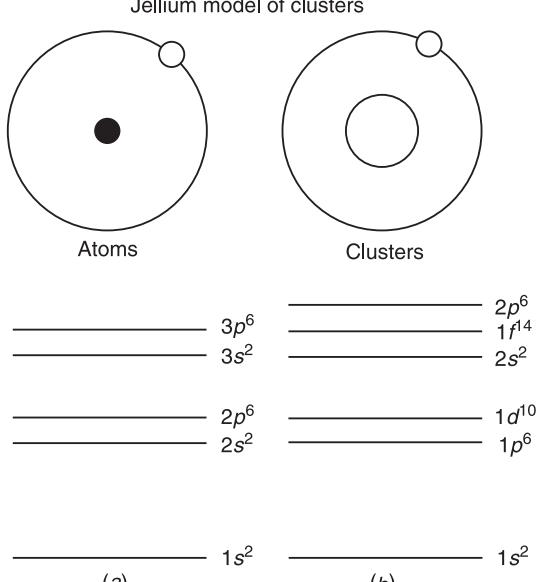


Fig. 32.4. A comparison of the energy levels of the hydrogen atom and those of the jellium model of a cluster

Notice that the order of the levels in the jellium model is different from that of the hydrogen atom. In this model the magic numbers correspond to those clusters having a size in which all the energy levels are filled. The electronic magic numbers of the atoms are 2, 10, 18, and 36 for He, Ne, Ar, and Kr, respectively (the Kr energy levels are not shown on the figure) and 2, 18 and 40 for the clusters.

### (ii) The density functional model

This model treats the clusters as molecules and uses existing molecular orbital theories such as density functional theory to calculate their properties. This approach can be used to calculate the actual geometric and electronic structure of small metal clusters.

The simplest of clusters is a molecule such as  $\text{H}_2^+$  because the Schrödinger equation can be solved exactly for a hydrogen atom. The ground state wavefunction of the hydrogen atom is expressed as

$$\psi(1s) = A \exp\left(-\frac{r}{\rho}\right) \quad \dots(1)$$

Here,  $r$  is the distance of the electron from the nucleus and  $\rho$  is the radius of the first Bohr orbit.

In the case of a molecule such as the  $\text{H}_2^+$  ion, molecular orbital theory assumes that the wavefunction of the electron around the two H nuclei can be described as a linear combination of the wavefunction of the isolated H atoms. Thus the wavefunction of the electrons in the ground state will have the form,

$$\psi = a\psi(1)_{1s} + a\psi(2)_{1s} \quad \dots(2)$$

The Schrödinger equation for the molecular ion is

$$\left[ \left( \frac{-\hbar^2}{2m} \right) \nabla^2 - \frac{e^2}{4\pi\epsilon_0 r_a} - \frac{e^2}{4\pi\epsilon_0 r_b} \right] \psi = E\psi \quad \dots(3)$$

Here,  $r_a$  and  $r_b$  are the distances of the electron from the two nuclei ‘a’ and ‘b’.

The last two terms in the brackets are the electrostatic attraction of the electron to the two positive nuclei.

For the hydrogen molecule, which has two electrons, a term for the electrostatic repulsion of the two electrons would be added. The Schrödinger equation is solved to calculate the actual electronic and geometric structures of small clusters.

When there are many atoms in the molecule and many electrons, the problem becomes complex, and many approximations are used to obtain the solution. Density functional theory represents one approximation. With the development of large fast computer capability and new theoretical approaches, it is possible using molecular orbital theory to determine the geometric and electronic structures of large molecules with a high degree of accuracy. The calculations can find the structure with the lowest energy, which will be the equilibrium geometry. These molecular orbital methods with some modification have been applied to metal nanoparticles.

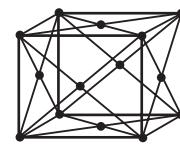


Fig. 32.5.

#### 32.3.3 Geometric Structure

Generally the crystal structure of large nanoparticles is the same as the bulk structure with somewhat different lattice parameters. X-ray diffraction studies of 80-nm aluminum particles have shown that it has the face-centered cubic (FCC) unit cell (Fig. 32.5), which is the structure of the unit cell of bulk aluminum.

But small particles having diameters of  $< 5$  nm may have different structures. For example, it has been shown that 3-5 nm gold particles have an icosahedral structure rather than the bulk FCC structure.

Indium clusters undergo a change of structure when the size is smaller than 5.5 nm. Above 6.5 nm, a diameter corresponding to about 6000 atoms, the clusters have a face-centered tetragonal structure with a  $c/a$  ratio of 1.075. In a tetragonal unit cell the edges of the cell are perpendicular, the long axis is denoted by  $c$ , and the two short axes by  $a$ . Below  $\sim 6.5$  nm the  $c/a$  ratio begins to decrease. At 5 nm  $c/a = 1$ , meaning that the structure is face-centered cubic. Fig. 32.6 is a plot of  $c/a$  versus the diameter of indium nanoparticles. A different structure can result in a change in many properties.

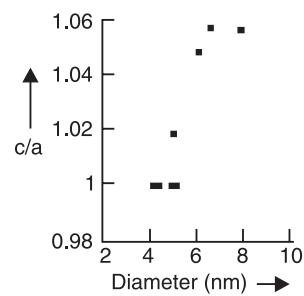


Fig. 32.6.

### 32.3.4 Electronic Structure

When a metal particle having bulk properties is reduced in size to a few hundred atoms, the density of states in the conduction band changes dramatically. Fig. 32.7 shows the changes in the electronic structure during the transition of a bulk metal to a large cluster, and then down to a small cluster of less than 15 atoms. The figure illustrates how energy levels of a metal change when the number of atoms of the material is reduced.

(i) At a size when a particle has a few hundred atoms or less, the continuous density of states changes over to a set of discrete energy levels. If the spacing between discrete energy levels turns out to be greater than the thermal energy  $k_B T$ , a gap opens up.

(ii) The small cluster is analogous to a molecule having discrete energy levels with bonding and antibonding orbitals.

Eventually a size is reached where the surfaces of the particles are separated by distances which are in the order of the wavelengths of the electrons. In this situation, the energy levels can be modeled by the quantum-mechanical treatment of a particle in a box. This is referred to as the *quantum size effect*.

The emergence of new electronic properties can be understood in terms of the Heisenberg uncertainty principle, which states that the more an electron is spatially confined the broader will be its range of momentum. The average energy will not be determined so much by the chemical nature of the atoms, but mainly by the dimension of the particle.

**Reactivity:** Since the electronic structure of nanoparticles depends on the size of the particle, the ability of the cluster to react with other species should depend on cluster size. This has important implications for the design of catalytic agents.

### 32.3.5 Magnetic Clusters

In this section we discuss the magnetic properties of nanoclusters of metal atoms that have a net magnet moment. In a cluster the magnetic moment of each atom will interact with the moments of the other atoms, and can force all the moments to align in one direction with respect to some symmetry axis of the cluster. The cluster will have a net moment, and is said to be magnetized. The

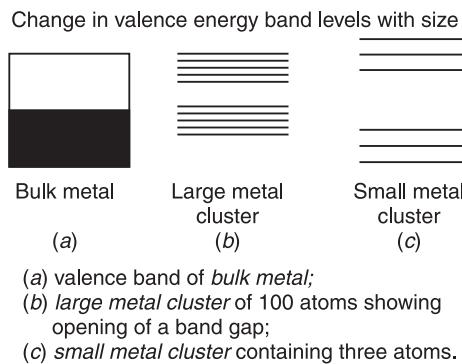


Fig. 32.7.

magnetic moment of such clusters is measured by Stern-Gerlach experiment.

Clusters made up of nonmagnetic atoms can have a net magnetic moment. For example, clusters of rhenium show a pronounced increase in their magnetic moment when they contain less than 20 atoms. Fig. 32.8 is a plot of the magnetic moment versus the size of the rhenium cluster. The magnetic moment is large when  $n$  is less than 15.

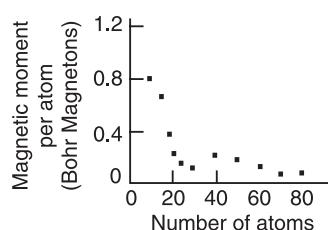


Fig. 32.8.

## 32.4 SEMICONDUCTING NANOPARTICLES

### 32.4.1 Optical Properties

In this section we discuss nanoparticles made of the elements, which are normal constituents of semiconductors. Nanoparticles made of cadmium, germanium, or silicon are not themselves semiconductors.

**Synthesis of  $\text{Si}_n$  nanoparticles.** A nanoparticle of  $\text{Si}_n$  can be made by laser evaporation of a Si substrate in the region of a helium gas pulse. The beam of neutral clusters is photolyzed by a UV laser producing ionized clusters whose mass to charge ratio is then measured in a mass spectrometer.

**Optical absorption spectrum:** In a bulk semiconductor a bound electron-hole pair, called an *exciton*, can be produced by a photon having an energy greater than that of the band gap of the material. The photon excites an electron from the filled band to the unfilled band above. The result is a hole in the valence band. Because of the Coulomb attraction between the positive hole and the negative electron, a bound pair, called an *exciton*, is formed.

The exciton can be modeled as a hydrogen-like atom. Light-induced transitions between these hydrogen-like energy levels produce a series of optical absorptions. Fig. 32.9 shows the optical absorption spectrum of excitons in cuprous oxide ( $\text{Cu}_2\text{O}$ ).

There is a significant shift in the optical absorption spectra toward the blue as the particle size is reduced. There are two situations, called the *weak-confinement* and the *strong-confinement regimes*.

(i) In the weak regime the particle radius is larger than the radius of the electron-hole pair, but the range of motion of the exciton is limited, which causes a blue shift of the absorption spectrum.

(ii) When the radius of the particle is smaller than the orbital radius of the electron-hole pair, the motion of the electron and the hole become independent. The exciton does not exist. The hole and the electron have their own set of energy levels. So, new lines appear in the absorption spectrum, showing again a blue shift.

Fig. 32.10 shows the optical absorption spectrum of CdSe for two nanoparticles having sizes  $20 \text{ \AA}$  and  $40 \text{ \AA}$ , respectively measured at  $10 \text{ K}$ . The lowest energy absorption region, referred to as the *absorption edge*, is shifted to higher energy as the particle size decreases. Since the absorption edge is due

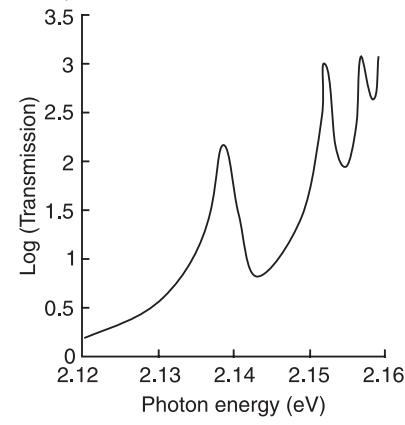


Fig. 32.9.

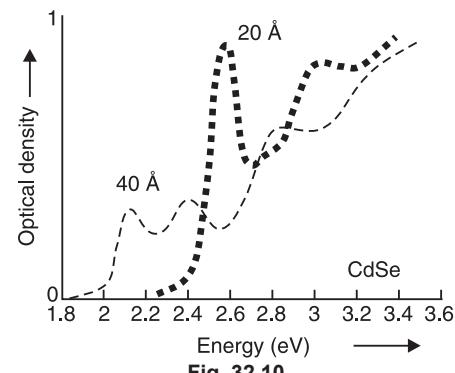


Fig. 32.10.

to the band gap, this means that the band gap increases as particle size decreases. The intensity of the absorption increases as the particle size is reduced. The higher energy peaks are associated with the exciton. They shift to higher energies with the decrease in particle size. These effects are a result of the confinement of the exciton. As the particle size is reduced, the hole and the electron are forced closer together, and the separation between the energy levels changes.

### 32.4.2 Photofragmentation

Nanoparticles of Ge and Si undergo fragmentation when subjected to laser light from a Q-switched Nd:YAG laser. Some of the fissions are



Here,  $h\nu$  is a photon of light energy.

The emerging products depend on the cluster size, the wavelength and the intensity of laser beam. When the cluster size is greater than 30 atoms, the fission occurs explosively.

### 32.4.3 Coulombic Explosion

Multiple ionizations of a cluster cause a rapid redistribution of the charges on the atoms of the cluster, making each atom more positive. When the Coulombic repulsion between the atoms in a cluster grows beyond the binding energy between the atoms, the cluster explodes and the atoms fly apart with great velocities. The phenomenon is called *Coulombic explosion*.

## 32.5 RARE GAS AND MOLECULAR CLUSTERS

Nanoparticles can be assembled from rare gases such as krypton and xenon, and molecules such as water and carbon dioxide. Examples of such largest clusters are:  $(\text{Kr}_{73})^{2+}$ ,  $(\text{Xe}_{206})^{4+}$  and  $(\text{CO}_2)_{216}^{4+}$ .

**1. Inert-Gas Clusters:** Xenon clusters are formed by adiabatic expansion of a supersonic jet of the gas through a small capillary into a vacuum. The gas is then collected by a mass spectrometer, where it is ionized by an electron beam, and its mass: charge ratio measured. The most stable clusters occur at particles having 13, 19, 25, 55, 71, 87, and 147 atoms.

These numbers are called *structural magic numbers*. It means that clusters having a certain number of atoms are more stable than others. The forces that bond inert-gas atoms into clusters are weaker than those that bond metals and semiconducting atoms.

**2. Superfluid Clusters:** Clusters of  ${}^4\text{He}$  and  ${}^3\text{He}$  atoms formed by supersonic free-jet expansion of helium gas have been studied by mass spectrometry. Magic numbers are found at cluster sizes of  $N = 7, 10, 14, 23, 30$  for  ${}^4\text{He}$ , and  $N = 7, 10, 14, 21, 30$  for  ${}^3\text{He}$ . Superfluidity has been observed in He clusters having 64 and 128 atoms. In the state of *Bose-Einstein condensation* (the state where all the bosons occupy the lowest energy level), the liquid  $\text{He}^4$  at 2.2 K becomes a superfluid. The viscosity of liquid  $\text{He}^4$  drops to zero.

**3. Molecular Clusters:** Individual molecules can form clusters. One of the most common examples of this is the water molecule. At ambient conditions 80% of water molecules are bonded into clusters. Other examples of molecular clusters are  $(\text{NH}_3)_n^+$ ,  $(\text{CO}_2)_{44}$  and  $(\text{C}_4\text{H}_8)_{30}$ .

## CARBON NANOSTRUCTURES

### 32.6 CARBON CLUSTERS

#### 32.6.1 Small Carbon Clusters

Laser evaporation of a carbon substrate using the apparatus shown in Fig. 32.1 in a pulse of He gas is used to make carbon clusters. The neutral cluster beam is photoionized by a UV laser and analyzed by a mass spectrometer. Fig. 32.11 shows the mass spectrum of carbon clusters.

For the number of atoms  $N$  less than 30, there are clusters of every  $N$ , although some are more prominent than others. Calculations of the structure of small clusters by molecular orbital theory show the clusters have linear or closed nonplanar monocyclic geometries.

(i) The linear clusters have  $sp$  hybridization with carbon bond angle as  $180^\circ$ . The linear structures occur when  $N$  is odd. The open structures with 3, 11, 15, 19, and 23 carbons are more prominent and more stable.

The two other forms of hybridization occurring in carbon molecules are  $sp^2$  and  $sp^3$ , with carbon bond angles measuring  $120^\circ$  and  $109^\circ 28'$  respectively.

(ii) Closed structures form when  $N$  is even. The closed structures have angles between the carbon bonds that differ from those predicted by the conventional hybridization concept. One such cluster composed of 60 atoms shows the highest intensity in the mass spectrum and hence the most stable.

#### 32.6.2 Discovery of $C_{60}$

The evidence for the existence of a molecule containing 60 carbon atoms was first found while measuring the intensity of light coming from stars at different wavelengths. A pronounced reduction in the intensity of light at 220 nm falling on Earth was detected. This is referred to as *optical extinction*. It was attributed to light scattered from small particles of graphite that were believed to be present in the regions between the stars. This explanation for the optical extinction in the 220-nm region has been widely accepted by astronomers.

The IR studies involving the use of graphite arcs were carried out for further clarification and confirmation. Huffman and Kratschmer observed four IR lines that correspond closely to the lines of a molecule  $C_{60}$ , predicted theoretically to exist several years earlier. Other investigations using techniques including the mass spectrometry were then made that led to the confirmation of the existence of  $C_{60}$  molecule in 1990.

#### 32.6.3 Structure of $C_{60}$ and its Crystal

The  $C_{60}$  molecule has been named fullerene.

Fig. 32.12 shows the structure of the  $C_{60}$  fullerene molecule. It has 12 pentagonal (5 sided) and 20 hexagonal (6 sided) faces symmetrically arrayed to form a molecular ball. A soccer ball has the

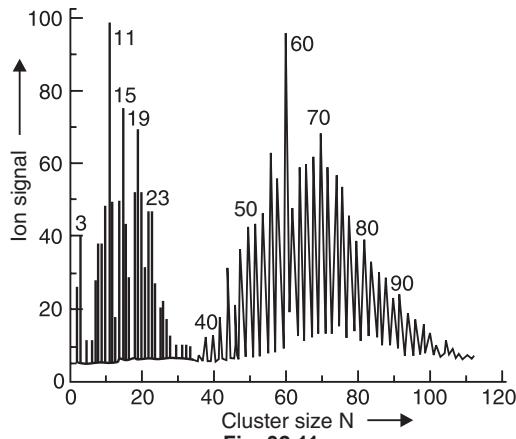


Fig. 32.11.

same geometric configuration as fullerene. These ball-like molecules bind with each other in the solid state to form a crystal lattice having a face centered cubic structure (Fig. 32.13).

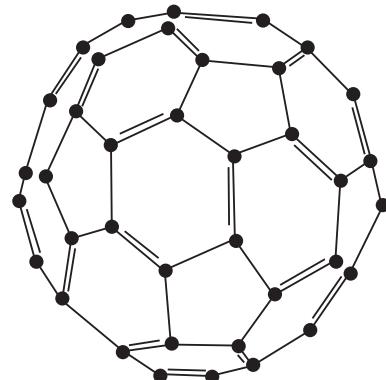


Fig. 32.12.

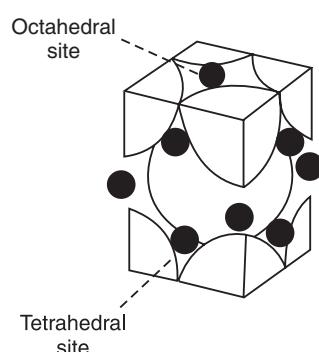


Fig. 32.13.

In the lattice each  $C_{60}$  molecule is separated from its nearest neighbor by 1 nm. The molecules are held together by weak forces called van der Waals forces. Because  $C_{60}$  is soluble in benzene, single crystals of it can be grown by slow evaporation from benzene solutions.

#### 32.6.4 Alkali-Doped $C_{60}$

In the FCC fullerene structure, 26% of the volume of the unit cell is empty. So alkali atoms can easily fit into the empty spaces between the molecular balls of the material. Fig. 32.13 shows the crystal lattice unit cell of  $C_{60}$  molecules (large spheres) doped with alkali atoms (dark circles).

The  $C_{60}$  crystal is an insulator, but when doped with an alkali atom it becomes electrically conducting.

#### 32.6.5 Superconductivity in $C_{60}$

Some fullerenes have been derived from  $C_{60}$  by crystallizing  $C_{60}$  along with alkali metals whose atoms are placed in gaps between the  $C_{60}$  spheres. These are found to show superconductivity. The transition temperatures of materials of this class range from 15 K to about 48 K. The transition temperature  $T_c$  of the fullerene  $Cs_2RbC_{60}$  is 33 K.

## 32.7 CARBON NANOTUBES

### 32.7.1 Fabrication of Carbon Nanotubes

There are three techniques of synthesizing carbon nanotubes: laser evaporation, carbon arc and chemical vapor deposition method.

#### 1. Laser Evaporation Method

Fig. 32.14 shows the experimental arrangement for synthesizing carbon nanotubes by laser evaporation.

A quartz tube contains argon gas and a graphite target. The tube is centred inside a horizontal furnace maintained at 1200°C with portions toward the ends falling outside the furnace. A water-cooled copper collector is mounted

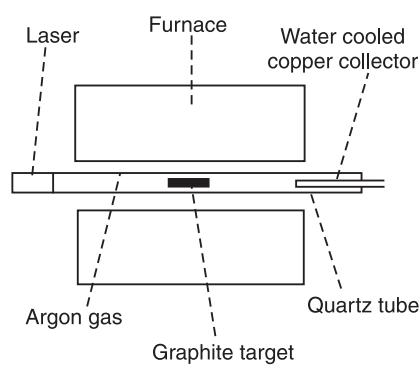


Fig. 32.14.

within the tube at one of its ends, lying outside the furnace. The graphite target contains small amounts of cobalt and nickel that act as catalytic nucleation sites for the formation of the nanotubes.

An intense pulsed laser beam is incident on the target, evaporating carbon from the graphite. The argon then sweeps the carbon atoms from the high-temperature zone to the colder copper collector on which they condense into nanotubes. Tubes 10-20 nm in diameter and 100  $\mu\text{m}$  long can be made by this method. This method produces tubes with closed ends.

## 2. Carbon Arc Method

A potential of 20-25V is applied across carbon electrodes of 5-20  $\mu\text{m}$  diameter and separated by 1 mm at 500 torr pressure of flowing helium. Carbon atoms are ejected from the positive electrode and form nanotubes on the negative electrode. As the tubes form, the length of the positive electrode decreases, and a carbon deposit forms on the negative electrode.

This method produces tubes with closed ends.

(i) **SWNT:** To produce single-walled nanotubes, a small amount of cobalt, nickel, or iron is incorporated as a catalyst in the central region of the positive electrode. This method can produce single-walled nanotubes of diameters 1–5 nm with a length of 1  $\mu\text{m}$ .

(ii) **MWNT:** If no catalysts are used, multi-walled nanotubes are produced. Fig. 32.15 shows a MWNT in which one tube is inside the other.

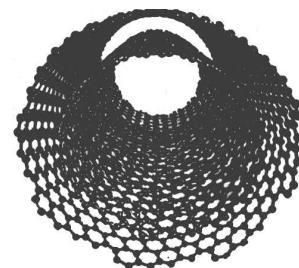


Fig. 32.15.

## 3. Chemical Vapor Deposition Method

This method involves decomposing a hydrocarbon gas such as methane ( $\text{CH}_4$ ) at 1100°C. As the gas decomposes, carbon atoms are produced. Carbon atoms then condense on a cooler substrate that contains various catalysts such as iron. This method produced tubes with open ends.

### 32.7.2 Structure of Carbon Nanotubes

A nanotube can be formed when a graphite sheet is rolled up about the axis  $T$  (Fig. 32.16).  $a_1$  and  $a_2$  are the basis vectors of the two-dimensional unit cell. The  $C_h$  vector is called the *circumferential vector*. Three possible structures of carbon nanotubes constructed by rolling the graphite sheet about the  $T$  vector having different orientations in the graphite sheet are shown in Fig. 32.17.

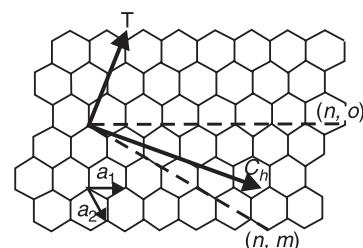
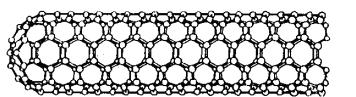
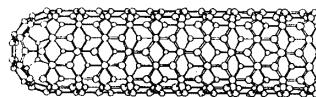


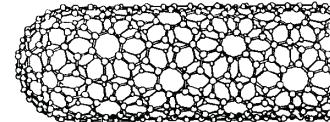
Fig. 32.16.



(a) Armchair structure



(b) Zigzag structure



(c) Chiral structure

Fig. 32.17

(i) When the axis vector  $T$  is parallel to the C—C bonds of the carbon hexagons, the “*armchair*” structure is obtained.

(ii) The *zigzag* and the *chiral structures* are formed by rolling about a  $T$  vector having different orientations in the graphite plane, but not parallel to C—C bonds.

### 32.7.3 Electrical Properties of Carbon Nanotubes

1. Carbon nanotubes are metallic or semiconducting, depending on the diameter and chirality of the tube. *Chirality* refers to how the tubes are rolled with respect to the direction of the axis vector  $T$ . Synthesis generally results in a mixture of tubes two-thirds of which are semiconducting and one-third metallic. The metallic nanotubes have the armchair structure.

Fig. 32.18 is a plot of the energy bandgap of semiconducting chiral carbon nanotubes versus the reciprocal of the diameter. As the diameter of the tube increases, the bandgap decreases.

2. The electronic structure of carbon nanotubes is studied using scanning tunnelling microscopy (STM). The data show that the electronic states of the tubes do not form a single wide electronic energy band, but instead split into one-dimensional subbands. These states can be modeled by a potential well having a depth equal to the length of the nanotube.

3. The voltage-current characteristics of a single metallic nanotube placed across two metal electrodes show step-like features. This resembles a field effect transistor (FET) constructed from a carbon nanotube.

4. In the metallic state the conductivity of the nanotubes is very high. They can carry extremely high current with density  $10^9 \text{ A/cm}^2$ . Copper wire fails at  $10^6 \text{ A/cm}^2$  because resistive heating melts the wire. Carbon nanotubes have very few defects to scatter electrons, and thus a very low resistance. This is the reason for the high conductivity of the carbon nanotubes. High currents do not heat the carbon nanotubes in the same way that they heat copper wires.

5. Carbon nanotubes are very good conductors of heat, the thermal conductivity being almost twice as large as that of diamond.

6. *Magnetoresistance* is a phenomenon whereby the resistance of a material is changed by the application of a DC magnetic field. Carbon nanotubes display magnetoresistive effects at low temperature.

Fig. 32.19 shows the effect of a DC magnetic field on the resistance of nanotubes at the temperatures of 0.35 and 2.3 K.  $\Delta R$  is the change in resistance of nanotubes at 2.3 and 0.35 K compared to their resistance  $R$  in zero magnetic field. This is a negative magnetoresistance effect because the resistance decreases with increasing DC magnetic field.

So the conductance  $G = 1/R$ , increases. This occurs because when a DC magnetic field is applied to the nanotubes, the conduction electrons acquire new energy levels associated with their spiraling motion about the field. For nanotubes, these levels, called *Landau levels*, lie very close to the topmost filled energy levels (the Fermi level). Thus there are more available states for the electrons to increase their energy, and the material is more conducting.

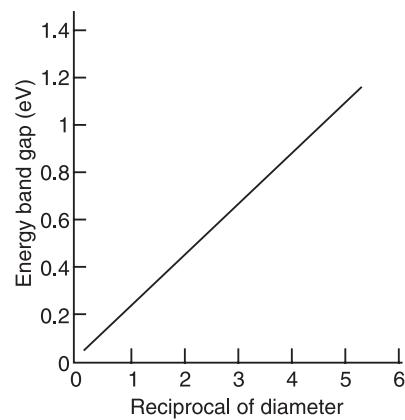


Fig. 32.18.

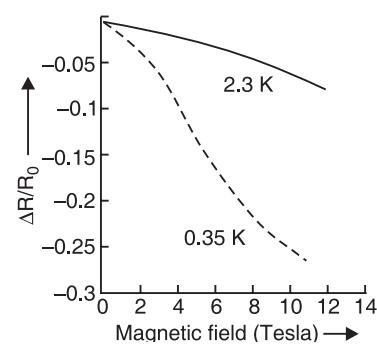


Fig. 32.19.

### 32.7.4 Vibrational Properties of Carbon Nanotubes

Fig. 32.20 (a) and (c) show two normal modes of vibration of carbon nanotubes.

(i)  **$A_{1g}$  Mode:**  $A_{1g}$  mode involves an ‘in and out’ oscillations of the diameter of the tube.

(ii)  **$E_{2g}$  Mode:** The  $E_{2g}$  mode involves a squashing of the tube where it squeezes down in one direction and expands in the perpendicular direction essentially oscillating between a sphere and an ellipse.

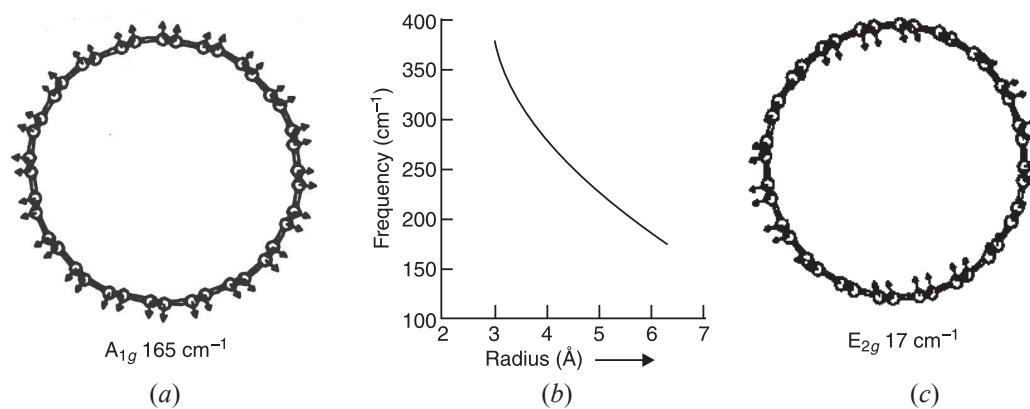


Fig. 32.20

The frequencies of these two modes are Raman-active and depend on the radius of the tube.

Fig. 32.20 (b) is a plot of the frequency of the Raman  $A_{1g}$  vibrational normal mode versus the radius of the nanotube.

It is now a standard practice to exploit this property for measuring the radius of carbon nanotubes.

### 32.7.5 Mechanical Properties of Carbon Nanotubes

(i) Carbon nanotubes are very strong.

(ii) The Young’s modulus of carbon nanotubes ranges from 1.28 to 1.8 TPa. Young’s modulus of steel is 0.21 TPa, which means that Young’s modulus of carbon nanotubes is almost 10 times that of steel.

(iii) When carbon nanotubes are bent, they are very resilient. They buckle like straws but do not break, and can be straightened back without any damage. Carbon nanotubes have a very low density of defects in the structure of their walls. So they do not fracture when bent severely.

(iv) The tensile strength of carbon nanotubes is about 45 billion pascals. High-strength steel alloys break at about 2 billion pascals. Thus carbon nanotubes are about 20 times stronger than steel.

(v) Multi-walled nanotubes also have improved mechanical properties, but they are not as good as their single-walled counterparts. For example, multi-walled nanotubes of 200 nm diameter have a tensile strength of 0.007 TPa and a modulus of 0.6 TPa.

## 32.8 APPLICATIONS OF CARBON NANOTUBES

### 1. Field Emission and Shielding

(i) When a small electric field is applied parallel to the axis of a nanotube, electrons are emitted at a very high rate from the ends of the tube. This effect, called *field emission*, has contributed to the development of flat panel displays.

(ii) The high electrical conductivity of carbon nanotubes means that they will be poor transmitters of electromagnetic energy. A plastic composite of carbon nanotubes could provide lightweight shielding material for electromagnetic radiation.

**2. Computers:** Carbon nanotubes with diameters of 2 nm have extremely low resistance. They can carry large currents without heating. So they could be used as interconnects.

Their very high thermal conductivity means that they can also serve as heat sinks, allowing heat to be rapidly transferred away from the chip.

**3. Chemical Sensors:** A field-effect transistor made of the chiral semiconducting carbon nanotubes is a sensitive detector of various gases.

**4. Catalysis:** Carbon nanotubes serve as catalysts for some chemical reactions.

**5. Mechanical Reinforcement:** Carbon nanotubes are used to increase the strength of plastic composites.

## QUANTUM WELLS, WIRES, AND DOTS

### 32.9 QUANTUM NANOSTRUCTURES

When the size or dimension of a material is continuously reduced from a large or macroscopic size, such as a meter or a centimeter, to a very small size, the properties remain the same at first, then small changes begin to occur, until finally when the size drops below 100 nm, dramatic changes in properties can occur. Three limiting classes of nanostructures emerge on the basis whether only one, only two or all three dimensions fall in nanometre scale.

**1. Quantum Well:** If one dimension is reduced to the nanometre range while the other two dimensions remain large, then we get a structure called *quantum well*.

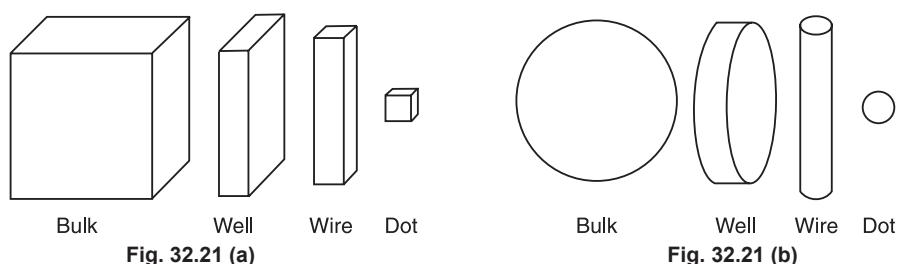
**2. Quantum Wire:** If two dimensions are reduced to the nanometre range and one remains large, the structure is referred to as a *quantum wire*.

**3. Quantum Dot:** The extreme case of this process of size reduction in which all three dimensions reach the low nanometer range is called a *quantum dot*.

The word *quantum* is associated with these three types of nanostructures because the changes in properties arise from the quantum-mechanical nature of physics in the domain of the ultrasmall.

Fig. 32.21 (a) shows formation of rectangular nanostructures by progressive reduction in size.

Fig. 32.21 (b) shows formation of curvilinear nanostructures by progressive reduction in size.



### 32.10 PREPARATION OF QUANTUM NANOSTRUCTURES

There are two approaches to the preparation of a nanostructure, called (i) *bottom-up* approach, and (ii) *top-down* approach.

**(i) Bottom-up method:** The procedure is to collect, consolidate, and fashion individual atoms and molecules into the structure. This is carried out by a sequence of chemical reactions controlled by catalysts.

**(ii) Top-down method:** The *top-down* method starts with a large-scale object or pattern and gradually reduces its dimension or dimensions. This is accomplished by a technique called *lithography* which shines radiation through a template on to a surface coated with a radiation-sensitive resist. The resist is then removed and the surface is chemically treated to produce the nanostructure. A typical resist material is the polymer polymethyl methacrylate  $[C_5O_2H_8]_n$ .

The lithographic process is illustrated by starting with a Gallium arsenide (GaAs) square quantum well located on a substrate [Fig. 32.22 (a)]. The final product to be produced from the material (GaAs) of the quantum well is either a quantum wire or a quantum dot [Fig. 32.22 (b)].

Fig. 32.23 shows the steps in the formation of a quantum wire or quantum dot by *electron-beam lithography*.

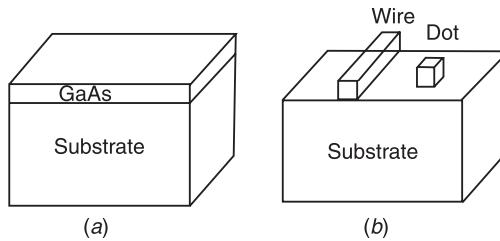


Fig. 32.22

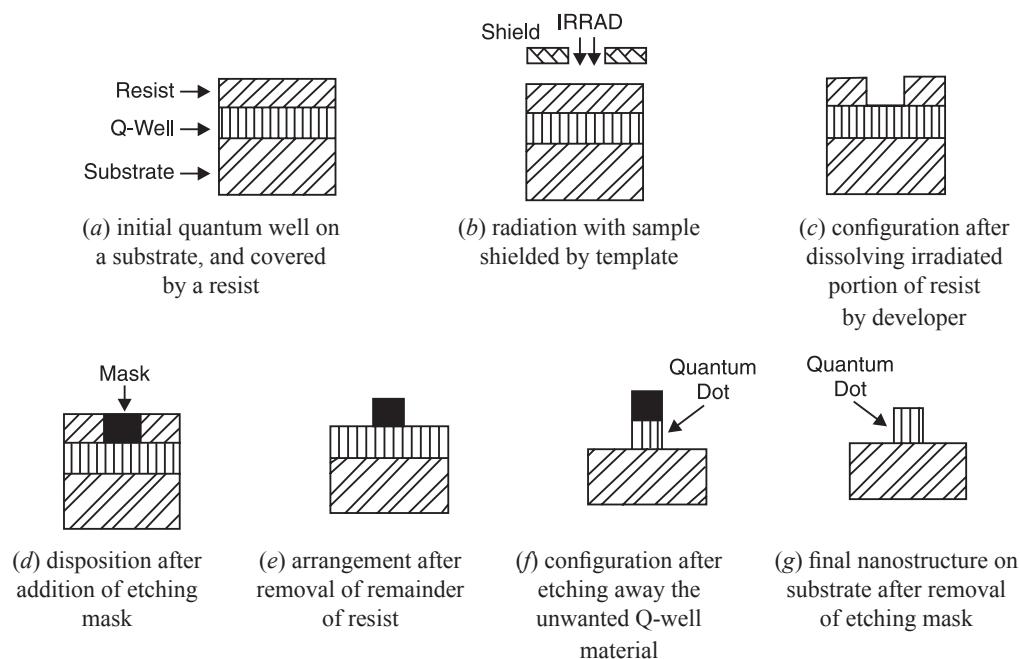


Fig. 32.23

- (a) The sample quantum well is coated with a radiation-sensitive resist.
- (b) The sample is then irradiated by an electron beam in the region where the nanostructure will be located. The radiation chemically modifies the exposed area of the resist so that it becomes soluble in a developer.
- (c) The irradiated portions of the resist are removed, using the developer.
- (d) An etching mask is inserted into the hole or groove in the resist.
- (e) The remaining parts of the resist are now taken out.
- (f) The areas of the quantum well not covered by the etching mask are chemically etched away to produce the quantum structure shown in Fig. 32.23 (f) covered by the etching mask.

- (g) Finally the etching mask is removed, if necessary, to provide the desired quantum structure. The resulting structure may be a quantum wire or a quantum dot, depending on the shape and the size of the groove.

**Quantum-Dot Array:** The lithographic technique is used to make more complex quantum structures than the quantum wire and quantum dot. Starting with a multiple quantum-well structure, we can produce a quantum dot array. A quantum-dot array produces a greatly enhanced photoluminescent output of light.

### 32.11 SIZE AND DIMENSIONALITY EFFECTS

**1. Size Effects:** The genesis of the remarkably different properties of quantum nanostructures lies in their drastically reduced dimensions. As the size on nanoscale reduces, the fraction of atoms on the surface of the nanocrystal grows larger and larger. Consequently, the number of conduction electrons in metals and semiconductors drops considerably and the electron mean free path becomes size-dependent. The resistivity of a polycrystalline conductor or semiconductor composed of microcrystallites with diameters significantly greater than the mean free path resembles that of a network of interconnected resistors. But when the microcrystallite dimensions approach or become less than the mean free path, the resistivity depends mainly on scattering off boundaries between crystallites. Both types of metallic nanostructures are common.

Another size effect arises from the level of doping of a semiconductor. For typical doping levels of  $10^{14}$  to  $10^{18}$  donors/cm<sup>3</sup> a quantum-dot cube of side 100nm would have, on the average, from  $10^{-1}$  to  $10^3$  conduction electrons. A smaller quantum-dot cube of side 10 nm would have, on the average,  $10^{-4}$  conduction electrons for the  $10^{14}$  doping level and 1 electron for the  $10^{18}$  doping level. Similar estimates have been made for quantum wells and quantum wires.

The following Table gives conduction electron content of smaller size (on left) and larger size (on right) quantum structures containing donor concentrations of  $10^{14}$  –  $10^{18}$  cm<sup>-3</sup>.

<i>Quantum Structure</i>	<i>Size</i>	<i>Electron Content</i>	<i>Size</i>	<i>Electron Content</i>
Bulk material	—	$10^{14}$ – $10^{18}$ cm <sup>-3</sup>	—	$10^{14}$ – $10^{18}$ cm <sup>-3</sup>
Quantum well	10 nm thick	$1 - 10^4$ $\mu\text{m}^{-2}$	100 nm thick	$10 - 10^5$ $\mu\text{m}^{-2}$
Quantum wire	$10 \times 10$ -nm cross section	$10^{-2} - 10^2$ $\mu\text{m}^{-1}$	$100 \text{ nm} \times 100 \text{ nm}$ cross section	$1 - 10^4$ $\mu\text{m}^{-1}$
Quantum dot	10 nm on a side	$10^{-4} - 1$	100 nm on a side	$10^{-1} - 10^3$

#### 2. Conduction Electrons and Dimensionality

<b>Quantum Structure</b>	<b>Delocalization Dimensions</b>	<b>Confinement Dimensions</b>
Bulk conductor	$3(x, y, z)$	0
Quantum well	$2(x, y)$	$1(z)$
Quantum wire	$1(z)$	$2(x, y)$
Quantum dot	0	$3(x, y, z)$

- (i) In a copper wire (bulk conductor) the conduction electrons are delocalized and move freely throughout the entire conducting wire. All the wire dimensions are very large compared to the distances between atoms.

The situation changes when one or more dimensions of the copper becomes so small that it approaches several times the spacings between the atoms in the lattice. When this occurs, the delocalization is impeded, and the electrons experience confinement.

(ii) Consider a flat plate of copper that is 10 cm long, 10 cm wide, and only 3.6 nm thick. This thickness corresponds to the length of only 10 unit cells. This means that 20% of the atoms are in unit cells at the surface of the copper. The conduction electrons would be delocalized in the plane of the plate, but confined in the narrow dimension. This configuration is called a *quantum well*.

(iii) A *quantum wire* is a structure such as a copper wire that is long in one dimension, but has a nanometer size as its diameter. The electrons are delocalized and move freely along the wire, but are confined in the transverse directions.

(iv) A *quantum dot* may have the shape of a tiny cube, a short cylinder, or a sphere with low nanometer dimensions. It exhibits confinement in all three spatial dimensions. So there is no delocalization.

The above four cases are summarized in the Table.

### 3. Partial Confinement

The conduction electrons in nanostructures can be partially confined and partially delocalized, depending on the shape and the dimensions of the structure.

One limiting case is a *quantum dot* in which they are totally confined.

The other limiting case is a *bulk material*, in which they are all delocalized.

The intermediate cases are a quantum wire and a quantum well which involve partial confinement.

A *quantum wire* is long in one dimension and very small in its transverse directions. The quantum wire exhibits electron confinement in two dimensions and delocalization in one dimension.

A *quantum well* is a flat plate nanosized in thickness and much larger in length and width. The quantum well exhibits electron confinement in one dimension and delocalization in two dimensions.

The following Table lists the number of electrons  $N(E)$  and density of states  $D(E) = dN(E)/dE$  as a function of the energy  $E$  for these four cases.

<b>Type</b>	<b>Number of electrons, <math>N(E)</math></b>	<b>Density of states, <math>D(E)</math></b>	<b>Dimensions</b>	
			<b>Delocalized</b>	<b>Confined</b>
Q-dot	$N(E) = K_0 \sum d_i \Theta(E - E_{iw})$	$D(E) = K_0 \sum d_i \delta(E - E_{iw})^2$	0	3
Q-wire	$N(E) = K_1 \sum d_i (E - E_{iw})^{1/2}$	$D(E) = \frac{1}{2} K_1 \sum d_i (E - E_{iw})^{-1/2}$	1	2
Q-well	$N(E) = K_2 \sum d_i (E - E_{iw})$	$D(E) = K_2 \sum d_i$	2	1
Bulk Conductor	$N(E) = K_3 (E)^{3/2}$	$D(E) = \frac{3}{2} K_3 (E)^{1/2}$	3	0

The summations in these expressions are over the various confinement well levels  $i$ .

$K_0, K_1, K_2$  and  $K_3$  are constants.

The degeneracies  $d_i$  refer to potential well energy levels.

$E_{iw}$  refers to the energy of the level  $i$  in the potential well.

The Heaviside step function  $\Theta(x) = 0$  for  $x < 0$  and 1 for  $x > 0$ .

The delta function  $\delta(x) = 0$  for  $x \neq 0$ , and  $\infty$  for  $x = 0$ , and integrates to a unit area.

### Energy Dependence Curves of Number of Electrons and Density of States

Fig. 32.24 shows number of electrons  $N(E)$  (left side) and density of states  $D(E)$  (right side) plotted against the energy  $E$  for three quantum structures and their bulk counterpart in the square well-Fermi gas approximations.

(i) The number of electrons  $N(E)$  increases with the energy  $E$  for all the four structures. So the four types vary only qualitatively from each other.

(ii) The density of states  $D(E)$  differ dramatically for each of the three nanostructure types. This means that the nature of the dimensionality and of the confinement associated with a particular nanostructure have a pronounced effect on its properties. These considerations can be used to predict properties of nanostructures. We can also identify types of nanostructures from their properties.

Some of the properties of solids that depend on the density of states are: electron heat capacity, Pauli susceptibility, the intensity of x-ray emission, electron and hole concentrations in semiconductors, the superconducting energy gap, and Josephson junction tunneling in superconductors. Experimental measurements of these various properties permit us to determine the form of the density of states  $D(E)$ , both at the Fermi level  $E_F$  and over a broad range of temperature.

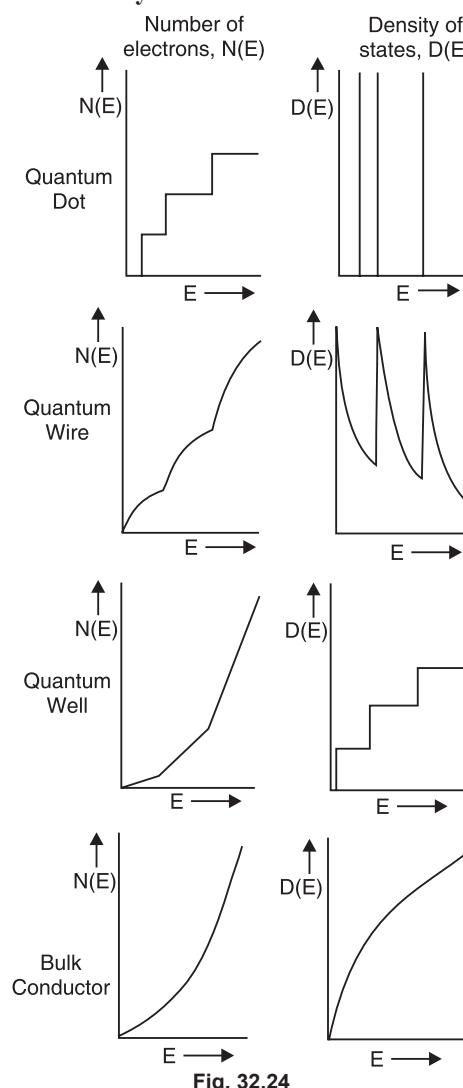


Fig. 32.24

## 32.12 APPLICATIONS OF QUANTUM NANOSTRUCTURES

### 1. Infrared Detectors

Infrared transitions involving energy levels of quantum wells are used for the operation of infrared photodetectors. The infrared detectors depend on the presence of discrete energy levels in a quantum well between which transitions in the infrared spectral region can be induced.

### 2. Quantum Dot Laser

Fig. 32.25 shows the structure of a quantum dot near-infrared laser fabricated by Park *et al.* In the quantum-dot laser, the quantum dots play the role of the active atoms.

The quantum-dot laser diode is grown on an n-doped GaAs substrate (not shown). The top p-metal layer has a GaAs contact layer immediately below it. Between this contact layer above and the GaAs substrate (not shown) below the diagram, there are a pair of 2- $\mu\text{m}$ -thick  $\text{Al}_{0.85}\text{Ga}_{0.15}\text{As}$  cladding or bounding layers that surround a 190-nm-thick waveguide made of  $\text{Al}_{0.05}\text{Ga}_{0.95}\text{As}$ . The waveguide plays the role of conducting the emitted light to the exit ports at the edges of the structure.

### Details of the waveguide region

The inset at the bottom of the figure shows details of the 190-nm-wide ( $\text{Al}_{0.85}\text{Ga}_{0.15}\text{As}$  cladded) waveguide region. Centered in the waveguide (dark horizontal stripe on the figure labeled QD) is a 30-nm-thick GaAs region, and centered in this region are 12 monolayers of  $\text{In}_{0.5}\text{Ga}_{0.5}\text{As}$  quantum dots with a density of  $1.5 \times 10^{10}/\text{cm}^2$ . The quantum dots do the lasing.

The cavity is typically 1 to 5 mm long and 4 to 60  $\mu\text{m}$  wide. The faces of the laser are coated with  $\text{ZnSe}/\text{MgF}_2$  high-reflectivity (> 95%) coatings. The coatings reflect light back and forth inside to augment the stimulated emission. The threshold current for CW operation at room temperature is 4.1 mA. The output signal has the wavelength of 1.32  $\mu\text{m}$ . The laser light exits through the lateral edge of the laser.

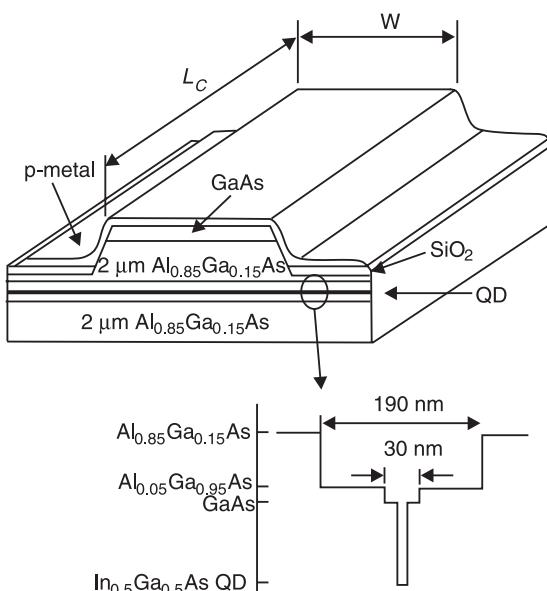


Fig. 32.25

## NANOMATERIALS: FABRICATION BY PHYSICAL TECHNIQUES

### 32.13 PHOTOLITHOGRAPHY

Photolithography is the most commonly applied technique in the mass production of integrated circuits.

Fig. 32.26 outlines the basic steps in the photolithographic process of transferring images in the mask to the substrate surface.

- The resist material is applied as a thin coating over some base.
- Selected areas of the resist are then exposed to light through a mask in an image-wise fashion.
- The developer is now brought in contact with the exposed resist. A positive tone or a negative tone image of the mask is obtained after the developing process, depending on whether the exposed areas of the resist are more or less soluble in the developer as compared to the unexposed areas. The three-dimensional relief image thus produced in the resist material is a replica of the opaque and transparent areas of the mask.
- The areas of resist that remain following the imaging and developing processes are used to mask the underlying substrate for subsequent etching or other image transfer steps. The resist material resists the etchant and prevents it from attacking the underlying substrate in those areas where it remains in place after development.

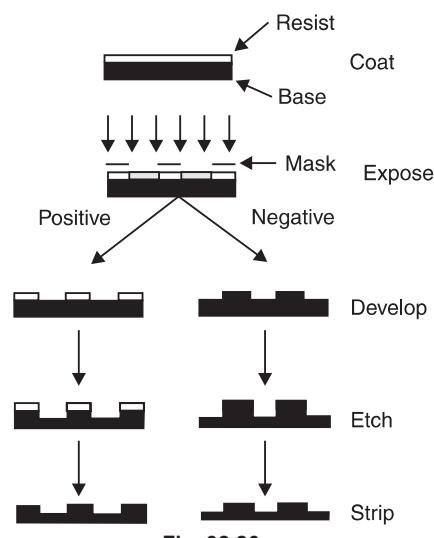


Fig. 32.26

- The resist is removed by stripping to produce a positive or negative tone relief image in the underlying substrate.

There are two basic photolithographic approaches:

- (1) Shadow printing, which can be further divided into contact printing (or contact-mode printing) and proximity printing, and
- (2) Projection printing.

#### **The resolution or minimum size of elements**

The *resolution* or minimum size of elements is set by diffraction effects (for photons).

Diffraction sets the limit of the maximum resolution or the minimum size of the individual elements by photolithography, which can be obtained.

- The theoretical resolution capability of shadow photolithography with a mask consisting of equal lines and spaces of width  $b$  is given by:

$$2b_{\min} = 3\sqrt{\lambda \left( s + \frac{d}{2} \right)} \quad \dots(1)$$

Here,  $2b$  = grating period ( $1/2b$  is the fundamental spatial frequency  $v$ )

$s$  = gap width maintained between the mask and the photoresist surface

$\lambda$  = wavelength of the exposing radiation and

$d$  = photoresist thickness

For hard contact printing,  $s = 0$ , and from the equation, the maximum resolution for 400 nm wavelength light and a 1  $\mu\text{m}$  thick resist film is slightly less than 1  $\mu\text{m}$ .

- In contact-mode photolithography, the mask and wafer are in intimate contact. Thus this method can transfer a mask pattern into a photoresist with almost 100% accuracy and provides the highest resolution. Other photolithographic techniques can approach but not exceed its resolution capabilities. However, the maximum resolution is seldom achieved because of dust on substrates and non-uniformity of the thickness of the photoresist and the substrate. Such problems can be avoided in proximity printing.
- In proximity printing, a gap between the mask and the wafer is introduced. However, increasing the gap degrades the resolution by expanding the penumbral region caused by diffraction. The difficulties in proximity printing include the control of a small and very constant space between the mask and wafer, which can be achieved only with extremely flat wafers and masks.
- In projection printing, the mask and the wafer substrate are separated by several centimetres. Lens elements are used to focus the mask image onto the wafer substrate. Because of lens imperfections and diffraction considerations, projection techniques have lower resolution capability than that provided by shadow printing. The resolution limit in projection photolithography is determined by the Rayleigh's equation.

The minimum resolvable feature,  $R$ , and the corresponding depth of focus (DOF) are given by the following :

$$R = \frac{k_1 \lambda}{NA} \quad \dots(2)$$

$$\text{DOF} = \frac{k_2 \lambda}{(NA)^2} \quad \dots(3)$$

Here  $\lambda$  is the exposure wavelength,  $k_1$  and  $k_2$  are constants that depend on the specific resist material, process technology and image-formation techniques used.

$NA$  is the numerical aperture of the optical system and is defined as

$$NA = n \sin \theta \quad \dots(4)$$

Here,

$n$  = index of refraction in image space and

$\theta$  = maximum cone angle of the exposure light beam.

Conventional photolithography is capable of fabricating features of 200 nm and above.

### 32.14 SCANNING TUNNELLING MICROSCOPY (STM)

**Principle.** STM relies on electron tunnelling, which is a quantum-mechanical phenomenon. Consider two flat surfaces of a metal or semiconductor, separated by an insulator or vacuum (Fig. 32.27).

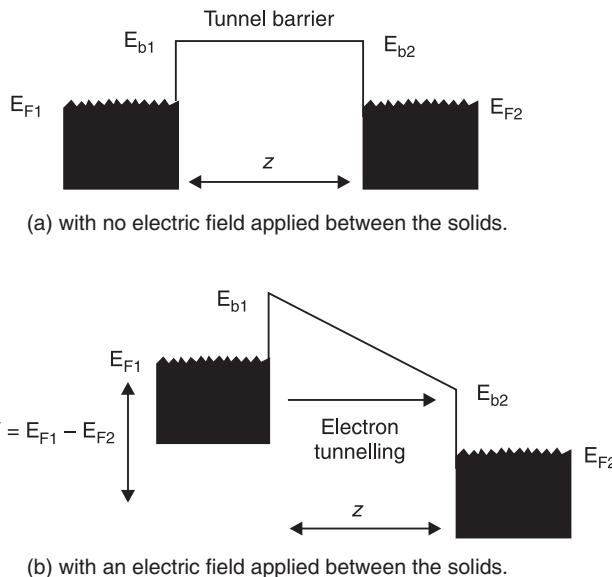


Fig. 32.27

$E_{F1}$  and  $E_{F2}$  are the Fermi levels of the respective solids.

Energies of the electrons in the solids are indicated by the shaded areas up to  $E_{F1}$  and  $E_{F2}$ .

The applied bias  $V$  is  $E_{F1} - E_{F2}$ .

$z$  is the distance between the two solids or the thickness of the insulator.

(a) Electrons in the material cannot transfer from one surface to another through the insulator, since there is an energy barrier.

(b) When a voltage is imposed between the two surfaces, the shape of the energy barrier changes [see Fig. 32.27 (b)]. There is a driving force for electrons to move across the barrier by tunnelling. This results in a small current when the distance is sufficiently small so that the electron wave functions extended from the two surfaces overlap.

The tunnelling current,  $I$ , is given by:

$$I \propto e^{-2kz} \quad \dots(1)$$

$k$  is given by:

$$k = \frac{\sqrt{2m(V-E)}}{h} \quad \dots(2)$$

Here,  $m$  is the mass of an electron,  $h$  is Planck's constant,  $E$  is the energy of electron, and  $V$  is the potential in the insulator.

### Block diagram of scanning tunnelling microscope (STM)

Fig. 32.28 shows the main components of STM.

A conducting wire having a very sharp tip is positioned above the surface of a sample. When the tip moves back and forth across the sample surface at very small intervals, the height of the tip is continually adjusted to keep the tunnelling current constant. The tip positions are used to construct a topographic map of the surface.

- The extremely sharp tip is usually made of metals or metal alloys, such as tungsten or PtIr alloy. The tip is mounted on to a three-dimensional positioning stage made of an array of piezoelectrics. The tip moves above the sample surface in three dimensions accurately controlled by the piezoelectric arrays. The distance between the tip and the sample surface falls between 0.2 and 0.6 nm. A tunneling current in the scale of 0.1 – 10 nA is generated.
- The scanning resolution is about 0.01 nm in XY direction and 0.002 nm in Z direction. This results in atomic scale resolution in the three-dimensional image. The STM provides atomic-scale resolution of surfaces. Its inventors, G. Binnig and H. Rohrer of IBM Zurich, won the 1986 Nobel Prize for this work.

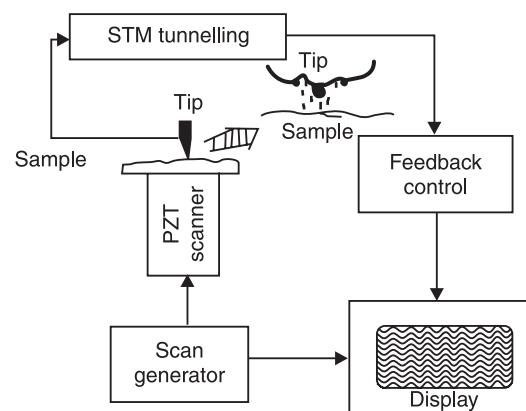


Fig. 32.28

### The Tunnelling Current

The configuration of a STM is a tip-planar surface geometry.

The tunnelling current is given by:

$$I = C \rho_t \rho_s e^{-\frac{1}{z^2}} \quad \dots(3)$$

Here,  $z$  is the distance between the tip and the planar surface or sample,

$\rho_t$  is the tip electronic structure,

$\rho_s$  is the sample electronic structure, and

$C$  is a constant dependent on the voltage applied between the tip and the sample surface.

- The tunnelling current decays exponentially with the tip-sample distance. For example, a 0.1 nm decrease in the distance will increase the tunneling current by one order of magnitude. This quantum mechanical property is used in the STM.

**Operating modes.** STM is operated in two modes.

(i) **Constant current mode:** In constant current imaging, a feed-back mechanism is enabled that a constant current is maintained while a constant bias is applied between the sample and tip. As the tip scans over the sample, the vertical position of the tip is altered to maintain the constant separation.

(ii) **Constant height mode:** In this imaging mode, constant height and bias are simultaneously maintained. A variation in current results as the tip scans the sample surface because a topographic structure varies the tip-sample separation.

- The constant current mode produces a contrast directly related to electron charge density profiles, whereas the constant height mode permits faster scan rates.
- A limitation of the STM is the requirement for a conducting sample. A variation on the

STM for insulators is the atomic-force microscope (AFM) which operates on the atomic force between the sample atoms and the probe atoms (in contrast to the tunneling current for the STM).

### 32.15 ATOMIC FORCE MICROSCOPY (AFM)

AFM is a modification of STM for dielectric materials. AFM monitors the force between the sample surface and the probe tip. A variety of tip-sample interactions may be measured by an AFM, depending on the separation.

- (i) At short distances, the van der Waals interactions are predominant.

Van der Waals force consists of interactions of three components: permanent dipoles, induced dipoles and electronic polarization.

- (ii) Long-range forces become significant when the tip-sample distance is large. Examples of such forces include electrostatic attraction or repulsion, current-induced or static-magnetic interactions, and capillary forces due to the condensation of water between the sample and tip.

#### Working of an atomic force microscope.

In AFM, the motion of a cantilever beam with an ultra small mass is measured. The force required to move this beam through measurable distance ( $10^{-5}$  nm) can be as small as  $10^{-18}$  N.

Fig. 32.29 shows how the AFM works.

The instrument consists of a cantilever with a nanoscale tip, a laser pointing at the end of a cantilever, a mirror and a photodiode collecting the reflected laser beam, and a three-dimensional positioning sample stage which is made of an array of piezoelectrics.

The images are generated by scanning the tip across the surface. The AFM measures the minute upward and downward deflections of the tip cantilever while maintaining a constant force of contact.

The sample is mounted on a scanner. The cantilever and tip are positioned near the surface with a macroscopic positioning device. A photo diode records the position of a laser beam, reflected off the top of the cantilever.

AFM's can achieve a resolution of 10 pm.

#### Difference between STM and AFM:

**STM:** monitors the tunnelling current between the surface and the probe tip.

**AFM:** monitors the force between the surface and the probe tip.

### EXERCISE

1. What are nanomaterials? Describe briefly different types of nanoparticles and discuss their structures.
2. Describe the structure and properties of various nanoparticles.
3. Discuss in brief the essentials of jellium model and its success.

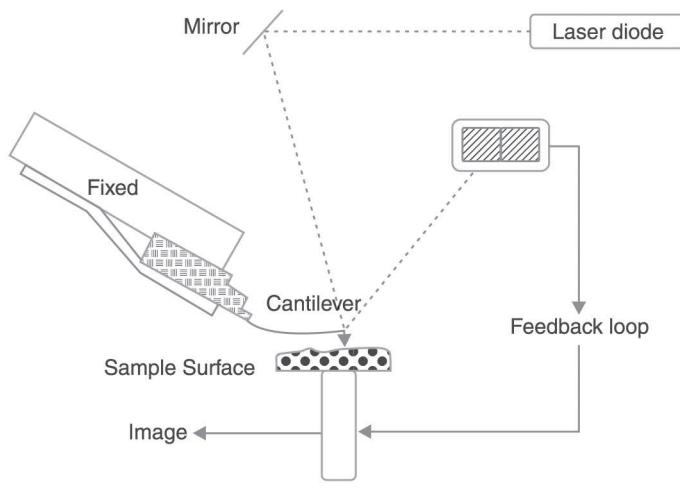


Fig. 32.29

4. Describe the structure and properties of carbon clusters.
5. What are fullerenes? Describe in detail the structure and properties of fullerite. Mention present and projected applications of fullerenes.
6. Describe in brief how carbon nanotubes are formed. Discuss their structure, characteristic properties and applications.
7. What are quantum nanostructures? Give a brief description of their fabrication.
8. Discuss the salient features of quantum nanostructures and describe their potential applications.
9. Discuss the basic principles of photolithography. Give experimental and theoretical bases of the technique.
10. Describe the principle and working of a scanning tunnelling microscope. Discuss briefly its application and the resolution achieved.
11. Describe the principle and working of an atomic force microscope. Give a comparative discussion on its application with reference to STM.

## PART-VIII

# ELECTRONICS AND SOLID STATE DEVICES

---

**33. Semiconductor Diode and Diode Applications**

---

**34. Special-Purpose Diodes**

---

**35. Bipolar Junction Transistor (BJT)**

---

**36. BJT Amplifiers**

---

**37. Oscillators**

---

**38. JFET, MOSFET, SCR and UJT**

---

**39. Operational Amplifiers**

---

**40. Communications Electronics**

---

**41. Number Systems and Logic Circuits**

## SEMICONDUCTOR DIODE AND DIODE APPLICATIONS

**After reading this chapter, you should be able to**

- ◆ Demonstrate that if a junction is formed between a sample of p-type and one of n-type semiconductor, this combination possesses the properties of a rectifier
- ◆ Analyze the voltage-current (V-I) characteristic curve of a diode
- ◆ Explain and analyze the operation of half-wave rectifiers, full-wave rectifiers and power supply filters
- ◆ Explain and analyze the operation of diode clipping and clamping circuits
- ◆ Discuss how diode clippers and diode champers differ in terms of their function

### 33.1 FORMATION OF P-N JUNCTION DIODE

Junction diode is formed by placing a *p*-type crystal in contact with *n*-type crystal and subjecting to high pressure so that it becomes a single piece. The assembly so obtained is called *p-n* junction or junction diode or crystal diode. The surface of contact of *p* and *n*-type crystals is called junction. A *p-n* junction is shown in Fig. 33.1. The *p*-type region has (positive) holes as majority charge-carriers. Similarly, the *n*-type region has (negative) electrons as majority charge carriers.

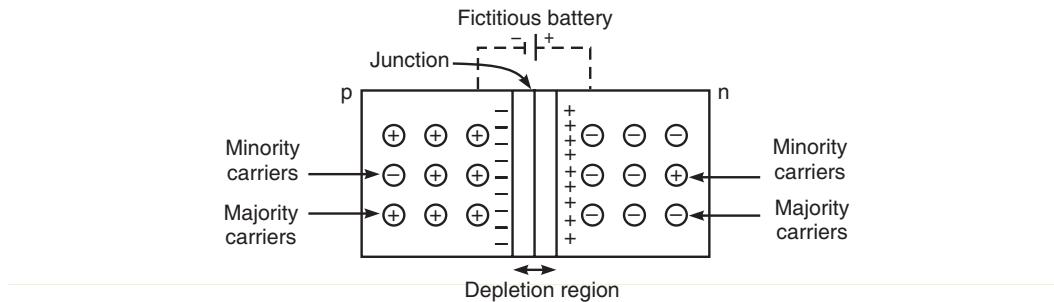
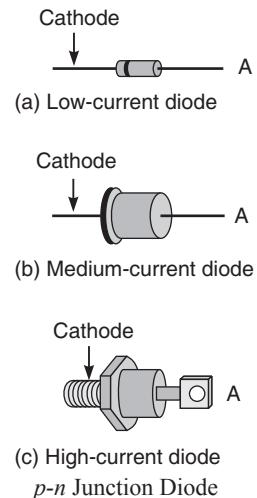


Fig. 33.1

In addition to these majority charge-carriers, there are a few minority charge-carriers in each region. The *p*-region contains a few electrons while the *n*-region contains a few holes. Thus we find that at the junction there is a decreasing hole concentration from left to right which makes the holes diffuse from *p*-side to *n*-side. Similarly electrons diffuse from right to left across the junction. Holes leaving and electrons entering the *p*-side make it negative. Similarly holes entering and electrons leaving the *n*-region make it positive. Thus, there is net negative charge on the *p*-side of the junction and net positive charge on the *n*-side. This produces an electric field across the junction as if a fictitious battery is connected with its positive terminal to *n* and negative terminal to *p* section. Equilibrium is established when the field becomes large enough to stop further diffusion of the majority charge-carriers. The field, however, helps the minority carriers to move across the junction.

The region on either side of the junction which becomes depleted (free) of the mobile charge-carriers is called the 'depletion region'. The thickness of this region is of the order of  $10^{-6}$  m. The potential difference across the depletion region is called the 'potential barrier'.



### 33.2 FORWARD AND REVERSE BIASING OF A JUNCTION DIODE

**1. Forward Bias.** When a battery is connected to the diode with *p*-section connected to positive pole and *n*-section to the negative pole, the junction diode is said to be forward biased (Fig. 33.2). If the forward bias is greater than the potential barrier, the majority carriers move towards the junction and cross it. The current which flows due to majority carriers is called forward current. It increases with forward bias.

**2. Reverse Bias.** When a battery is connected to junction diode with *p*-section connected to negative pole and *n*-section connected to the positive pole, the junction is said to be reverse biased (Fig. 33.3). When reverse bias is applied, the majority carriers do not cross the junction. However a very little amount of current flows due to the motion of minority carriers. This current is called reverse current. This current increases with increasing temperature.

Thus we see that the junction diode offers a low resistance for the current to flow in one direction (under forward bias) but a very high resistance in the opposite direction (under reverse bias). It thus acts as a rectifier.

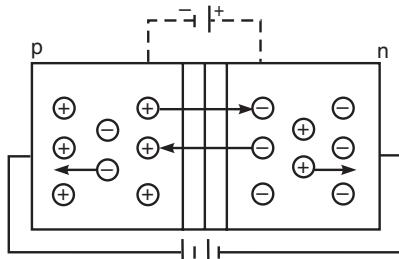


Fig. 33.2

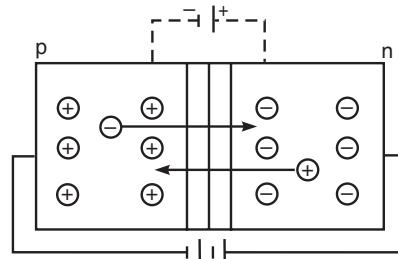


Fig. 33.3

### 33.2.1. V-I Characteristics of a Junction Diode

Graphs drawn between bias voltage and circuit current of a junction diode are called characteristics of the diode.

**1. Forward Bias Characteristic.** This is obtained by plotting a graph between forward bias voltage and circuit current. The circuit used is shown in Fig. 33.4. The milliammeter mA measures the current. The voltmeter  $V$  measures the p.d. across the diode. The forward voltage is gradually increased in steps and corresponding milliammeter readings are noted. A graph is then plotted between voltage and current (Fig. 33.5). Practically no current flows until the barrier potential is overcome. Once the external voltage exceeds the barrier potential, the current increases rapidly, approximately exponentially.

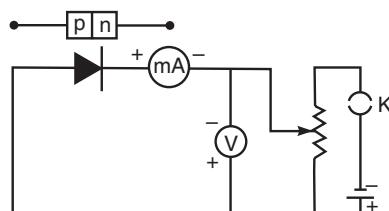


Fig. 33.4

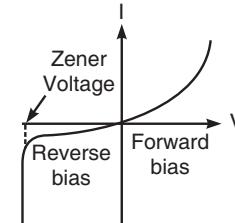


Fig. 33.5

**2. Reverse Bias Characteristic.** Connections are made as shown in Fig. 33.6. The reverse voltage is gradually increased in steps and corresponding microammeter readings are noted. A graph is then plotted between voltage and current (Fig. 33.5). With reverse bias the reverse current remains very small over a long range, increasing very slightly with increasing bias.

**(a) Avalanche breakdown.** If the reverse bias is made very high, the covalent bonds near the junction break down and a large number of electron-hole pairs are liberated. The reverse current then increases abruptly to a relatively large value. This is known as “*Avalanche breakdown*”, and may damage the junction by excessive heat generated unless the current is limited by external circuit. This phenomenon is used in making **zener diodes**.

*The maximum voltage that a junction diode can bear without breakdown is called zener voltage or reverse breakdown voltage.*

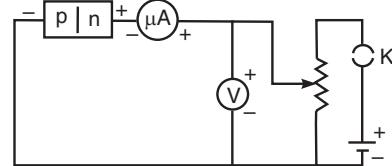


Fig. 33.6

**(b) Dynamic resistance.** Both the forward bias and reverse bias characteristics of the *p-n* junction do not obey Ohm's law. Therefore, the resistance offered by junction diode depends upon the applied voltage.

The dynamic resistance of junction diode is defined as the ratio of small change in voltage to the small change in current produced. It is also called a.c. resistance of the junction diode and is denoted by  $r_d$ .

Thus,

$$r_d = \frac{\Delta V}{\Delta I}.$$

The region of the characteristic curve, where dynamic resistance is almost independent of the applied voltage, is called the linear region of junction diode.

**Example 1.** Can a *p-n* junction be formed for a given breakdown voltage ? Explain.

**Ans.** Yes. It can be formed for a given breakdown voltage by suitably adding the doping agent.

### 33.3 DERIVATION OF EXPRESSIONS FOR THE WIDTH OF DEPLETION LAYER AND HEIGHT OF POTENTIAL BARRIER

Fig. 33.7 shows a *P-N* junction.

Fig. 33.7 (a) shows *P*-type and *N*-type semiconductors before they are joined. *P*-type semiconductor has negative acceptor ions (shown by encircled minus sign) and positively charged free holes which move about on *P*-side. Similarly, *N*-type semiconductor has positive donor ions (shown by encircled positive sign) and negatively charged free electrons which move about *N* side.

Fig. 33.7 (b) shows the *P-N* junction after the two pieces are joined.

*P*-type material has a high concentration of holes. *N*-type material has high concentration of free electrons. Hence there is a tendency of holes to diffuse over to *N*-side and electrons to *P*-side. The process is known as *diffusion*. So due to diffusion, some of the holes from *P*-side cross over to *N*-side where they combine with electrons and become neutral. Similarly, some of the electrons from *N*-side cross over to *P*-side where they combine with holes and become neutral. Thus a region is formed which is known as *depletion layer* or *space charge region* because there is no charge available for conduction. The thickness of the depletion layer is usually of the order of  $10^{-6}$  m.

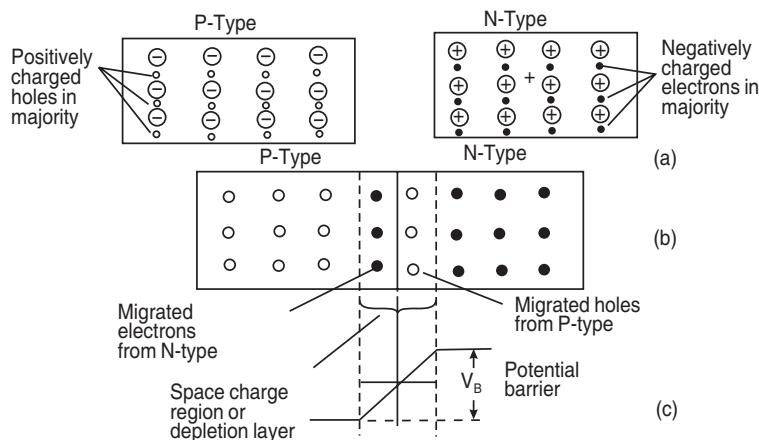


Fig. 33.7

The diffusion of holes and electrons continues till a potential barrier [Fig. 33.7 (c)] is developed in space charge region which prevents further diffusion or neutralization.

**Junction Voltage.** When the depletion layer is formed, there are negative immobile ions in *P*-type semiconductor and positive immobile ions in *N*-type semiconductor (Fig. 33.8).

Now due to charge separation, a voltage  $V_B$  is developed across the junction under equilibrium condition. This voltage is called *junction voltage* or *internal voltage*. The potential barrier  $V_B$  set up in this manner gives rise to electric field. This electric field prevents the respective majority carriers from crossing the barrier region. The potential barrier is about 0.3 V for Ge and 0.7 V for Si at room temperature (300 K).

As soon as the *P-N* junction is formed, the diffusion of holes (from *P* region to *N* region) and electrons (from *N* region to *P* region) takes place for a short time. Then, a force (called a potential barrier) is automatically developed across the junction. The *potential barrier* stops the further diffusion of holes and electrons from one side to the other. The difference of potential from one side of the barrier to the other side is called the *height of the barrier*.

Fig. 33.9 shows the variation of potential with distance along the junction.

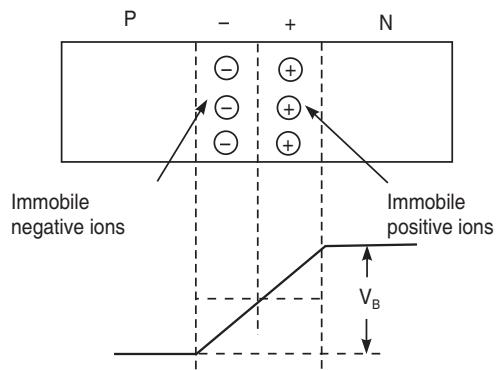


Fig. 33.8

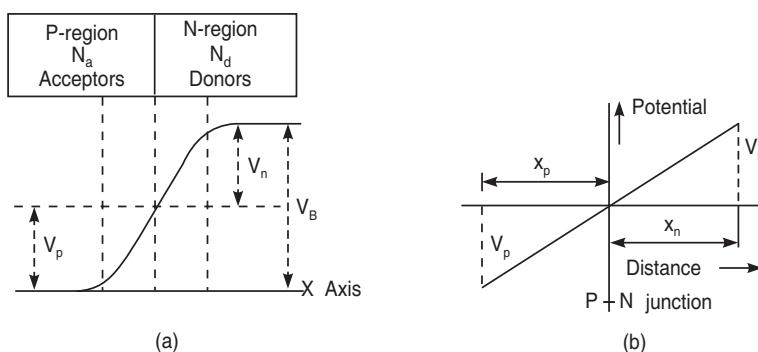


Fig. 33.9

The barrier potential  $V_B$  is given by

$$V_B = |V_p| + |V_n| \quad \dots(1)$$

Here,

$V_p$  = magnitude of the potential fall in *P* region,

$V_n$  = magnitude of the potential rise in *N* region.

The width of the depletion region is given by

$$x = x_p + x_n \quad \dots(2)$$

Here,  $x_p$  and  $x_n$  are the widths of depletion layer in *P* and *N* regions respectively.

To find the distribution of barrier potential in the space charge region, we use Poisson's equation

$$\frac{d^2V}{dx^2} = -\frac{\rho}{\epsilon} \quad \dots(3)$$

Here  $\rho$  is the volume density of charge and  $\epsilon$  is the permittivity of the medium.

The charge density in *P*-side of depletion layer is given by

$$\rho = -eN_a$$

Here,  $N_a$  is the density of completely ionized acceptor atoms.

Substituting the value of  $\rho$  in Eq. (3), we get

$$\frac{d^2V}{dx^2} = \frac{eN_a}{\epsilon}$$

Integrating it, we get

$$\frac{dV}{dx} = \frac{eN_a \cdot x}{\epsilon} + C_1 \quad \dots(4)$$

Here,  $C_1$  is a constant of integration.

Applying boundary condition :

at  $x = -x_p, \frac{dV}{dx} = 0$

we get  $C_1 = \frac{eN_a}{\epsilon} \cdot x_p$

Substituting the value of  $C_1$  in Eq. (4), we get

$$\frac{dV}{dx} = \frac{eN_a}{\epsilon} \cdot x + \frac{eN_a}{\epsilon} \cdot x_p$$

or  $\frac{dV}{dx} = \frac{eN_a}{\epsilon} \cdot (x + x_p) \quad \dots(5)$

Integrating Eq. (5), we get

$$V = \frac{eN_a}{\epsilon} \left( \frac{x^2}{2} + x_p x \right) + C_2 \quad \dots(6)$$

Here,  $C_2$  is another constant of integration.

Its value is obtained by applying the condition that at  $x = 0, V = 0$ .

or  $C_2 = 0$ .

$\therefore V = \frac{eN_a}{\epsilon} \left( \frac{x^2}{2} + x_p x \right) \quad \dots(7)$

For  $P$ -region, at  $x = -x_p, V = V_p$

$$V_p = \frac{eN_a}{\epsilon} \left( \frac{x_p^2}{2} - x_p^2 \right) = -\frac{eN_a x_p^2}{2\epsilon}$$

$\therefore |V_p| = \frac{eN_a}{2\epsilon} x_p^2 \quad \dots(8)$

Similarly, the Poisson's equation for  $N$ -side of depletion layer is

$$\frac{d^2V}{dx^2} = -\frac{\rho}{\epsilon} = \frac{eN_d}{\epsilon}.$$

Here,  $N_d$  is the density of completely ionized donor atoms.

Proceeding as above and applying boundary conditions :

at  $x = 0, V = 0$   
and at  $x = x_n, dV/dx = 0$  and  $V = V_n$

we get  $V_n = -\frac{eN_d}{2\epsilon} x_n^2$

$$|V_n| = \frac{eN_d}{2\epsilon}x_n^2 \quad \dots(9)$$

Therefore, the height of potential barrier is given by

$$\begin{aligned} V_B &= |V_p| + |V_n| \\ &= \frac{eN_a}{2\epsilon}x_p^2 + \frac{eN_d}{2\epsilon}x_n^2 \\ \therefore V_B &= \frac{e}{2\epsilon} \left[ N_a x_p^2 + N_d x_n^2 \right] \end{aligned} \quad \dots(10)$$

Since the crystal as a whole is electrically neutral, the number of charge carriers on both sides must be equal, i.e.,

$$\begin{aligned} N_a x_p &= N_d x_n \\ \text{or} \quad x_n &= \frac{N_a}{N_d} \cdot x_p \end{aligned}$$

Substituting the value of  $x_n$  in Eq. (10), we get

$$\begin{aligned} V_B &= \frac{e}{2\epsilon} \left[ N_a x_p^2 + N_d \frac{N_a^2}{N_d^2} \times x_p^2 \right] \\ \text{or} \quad V_B &= \frac{e}{2\epsilon} N_a \cdot x_p^2 \left[ 1 + \frac{N_a}{N_d} \right] \end{aligned} \quad \dots(11)$$

$$\begin{aligned} \text{or} \quad x_p &= \left[ \frac{2\epsilon V_B}{e N_a \{1 + (N_a / N_d)\}} \right]^{1/2} \\ \therefore x_p &= \left[ \frac{2\epsilon V_B}{e} \cdot \frac{N_d / N_a}{N_a + N_d} \right]^{1/2} \end{aligned} \quad \dots(12)$$

$$\begin{aligned} \text{Similarly,} \quad x_n &= \left[ \frac{2\epsilon V_B}{e N_d \{1 + (N_d / N_a)\}} \right]^{1/2} \\ \therefore x_n &= \left[ \frac{2\epsilon V_B}{e} \cdot \left( \frac{N_a / N_d}{N_a + N_d} \right) \right]^{1/2} \end{aligned} \quad \dots(13)$$

Therefore, total width of depletion layer

$$\begin{aligned} x &= x_p + x_n \\ x &= \left\{ \frac{2\epsilon V_B}{e(N_a + N_d)} \right\}^{1/2} \left[ \left( \frac{N_d}{N_a} \right)^{1/2} + \left( \frac{N_a}{N_d} \right)^{1/2} \right] \end{aligned} \quad \dots(14)$$

If  $N_d \gg N_a$  then  $(N_a + N_d) \rightarrow N_d$  and  $N_d/N_d \rightarrow 0$ .

Eq. (14) becomes

$$x = \left( \frac{2\epsilon V_B}{e N_d} \right)^{1/2} \left( \frac{N_d}{N_a} \right)^{1/2} = \left( \frac{2\epsilon V_B}{e N_a} \right)^{1/2} \quad \dots(15)$$

Eq. (15) gives the width of depletion layer.

The width of depletion layer decreases with increase in impurity concentration.

### Expression for the height of barrier potential in terms of impurity densities

The density of electrons in conduction band in an intrinsic semiconductor is given by

$$n = n_c = 2 \left( \frac{2\pi m_e kT}{h^2} \right)^{3/2} e^{(E_F - E_c)/kT} \quad \dots(16)$$

Addition of donor impurity raises the Fermi level by  $\Delta E$  on  $N$ -side.

The new Fermi level is

$$E'_F = E_F + \Delta E$$

Expression for density of electrons on  $N$ -side modifies to

$$(n_c)_N = 2 \left( \frac{2\pi m_e kT}{h^2} \right)^{3/2} e^{(E_F + \Delta E - E_c)/kT}$$

or

$$(n_c)_N = n e^{\Delta E/kT} \quad \dots(17)$$

Similarly, the density of electrons on  $P$ -side is given by

$$(n_c)_P = n e^{(-\Delta E/kT)} \quad \dots(18)$$

because addition of donor impurities brings down the Fermi level on  $P$ -side by  $\Delta E'$ .

From Eqs. (17) and (18), we get

$$\begin{aligned} \frac{(n_c)_N}{(n_c)_P} &= e^{(\Delta E + \Delta E')/kT} \\ &= e^{(eV_B/kT)} \end{aligned} \quad \dots(19)$$

Here,  $V_B$  is the height of potential barrier and  $eV_B = \Delta E + \Delta E'$ .

From Eq. (19), we get

$$V_B = \frac{kT}{e} \log_e \left[ \frac{(n_c)_N}{(n_c)_P} \right] \quad \dots(20)$$

By the *law of mass action*,

$$n_c n_h = np = n_i^2.$$

Here  $n_i$  is called the intrinsic density of either carrier.

For a semiconductor having  $N_d$  completely ionised donor atoms, it becomes

$$(n'_c + N_d) n'_h = n_i^2.$$

Here,  $n'_c$  and  $n'_h$  are the number of thermally generated electron-holes.

For an  $N$ -type semiconductor,  $N_d \gg n'_c$  and we have

$$N_d N'_h = n_i^2.$$

Therefore, the hole concentration in  $N$ -type semiconductor, i.e., minority carrier concentration

$$(n'_h)_N = \frac{n_i^2}{N_d} \quad \dots(21)$$

Similarly, for a semiconductor having  $N_a$  completely ionised acceptor atoms,

$$n'_c (N_a + n'_h) = n_i^2.$$

But for a  $P$ -type semiconductor  $N_a \gg n'_h$  and we have

$$n'_c N_a = n_i^2.$$

Therefore, electron concentration in  $P$ -type material

$$(n'_c)_P = \frac{n_i^2}{N_a} \quad \dots(22)$$

If we assume that minority carriers on *P*-side are thermally generated, then

$$(n'_c)_P = \frac{(n_c)_P}{N_a} \quad \dots(23)$$

Eq.(20) for  $V_B$  becomes

$$\begin{aligned} V_B &= \frac{kT}{e} \log_e \left[ \frac{(n_c)_N}{n_i^2 / N_a} \right] \\ \text{or} \quad V_B &= \frac{kT}{e} \log_e \left[ \frac{(n_c)_N N_a}{n_i^2} \right] \end{aligned} \quad \dots(24)$$

Since majority carriers (electrons) in *N*-type semiconductor are largely contributed by donor atoms, we take

$$(n_c)_N = N_d$$

Therefore, Eq. (24) becomes

$$V_B = \frac{kT}{e} \log_e \left[ \frac{N_d N_a}{n_i^2} \right] \quad \dots(25)$$

This gives the height of *barrier potential in terms of impurity densities that create it.*

### 33.4

### DERIVATION OF VOLTAGE CURRENT (V-I) EQUATION FOR A P-N JUNCTION DIODE

The density or the concentration of electrons in the *N*-region is

$$n_n = 2 \left( \frac{2\pi m_n^* kT}{h^2} \right)^{3/2} e^{-(E_{cn} - E_{fn})/kT} \quad \dots(1)$$

Here,

$m_n^*$  = effective mass of the electron in the conduction band,

$E_{cn}$  = energy at the bottom of conduction band,

and

$E_{fn}$  = Fermi level.

The *P*-region also contains some thermally generated electrons of concentration

$$n_p = 2 \left( \frac{2\pi m_n^* kT}{h^2} \right)^{3/2} e^{-(E_{cp} - E_{fp})/kT} \quad \dots(2)$$

∴

$$\frac{n_n}{n_p} = \frac{e^{-(E_{cn} - E_{fn})/kT}}{e^{-(E_{cp} - E_{fp})/kT}} \quad \dots(3)$$

For the same semiconducting material, the positions of the conduction and valence band edges do not change with doping. But the position of the Fermi level changes with both concentration and type of doping. Thus for *N* and *P* regions,

$$E_{cn} = E_{cp} \text{ and } E_{fn} \neq E_{fp}$$

Therefore, Eq. (3) reduces to

$$\frac{n_n}{n_p} = e^{(E_{fn} - E_{fp})/kT} = e^{(eV_B/kT)} \quad \dots(4)$$

Here,

$$\begin{aligned} E_{fn} - E_{fp} &= eV_B = E_B = \text{barrier energy.} \\ V_B &= \text{barrier potential.} \end{aligned}$$

Thus at equilibrium, the minority carrier electron concentration in the *P*-region is related to the majority carrier electron concentration in the *N*-region by

$$n_p = n_n e^{(-eV_B/kT)} \quad \dots(5)$$

On applying a small forward voltage *V*, the effective barrier potential decreases to  $(V_B - V)$  and some of the majority carriers diffuse across the junction. Thus holes diffuse to *N*-side and electrons to *P*-side. Consequently, minority carrier electron concentration at the edge of transition region on the *P*-side becomes greater than its equilibrium value by a small amount  $\Delta n_p$ . The majority carrier electron concentration in the *N*-region, however, remains nearly constant.

Thus under the forward bias condition, Eq. (5) becomes

$$\begin{aligned} n_p + \Delta n_p &= n_n e^{-e(V_B - V)/kT} \\ &= n_p e^{e(V_B)/kT} \cdot e^{-e(V_B - V)/kT} \quad [\text{using Eq. (5)}] \\ &= n_p e^{(eV/kT)} \\ \text{or} \quad \Delta n_p &= n_p [e^{eV/kT} - 1] \end{aligned} \quad \dots(6)$$

Similarly, for excess holes on the *N*-side

$$\Delta p_n = p_n [e^{eV/kT} - 1] \quad \dots(7)$$

The diffusion currents produced by these charges can be determined by using diffusion equations and Fick's first law.

The electron diffusion current injected into the *P*-region at the junction is given by

$$I_n = \frac{eAD_n}{L_n} \Delta n_p = \frac{eAD_n}{L_n} \cdot n_p [e^{eV/kT} - 1] \quad \dots(8)$$

Here,

*A* = area of cross-section of the junction,

*D<sub>n</sub>* = diffusion coefficient of electron,

*L<sub>n</sub>* = electron diffusion length.

Similarly, the hole diffusion current injected into the *N*-region at the junction is

$$I_p = \frac{eAD_p}{L_p} p_n [e^{eV/kT} - 1] \quad \dots(9)$$

Here, *D<sub>p</sub>* and *L<sub>p</sub>* are respectively the diffusion coefficient and diffusion length for holes.

Since the direction of both the currents is the same, the total diode current

$$\begin{aligned} I &= I_p + I_n = eA \left[ \frac{D_p}{L_p} p_n + \frac{D_n}{L_n} n_p \right] [e^{eV/kT} - 1] \\ \therefore I &= I_0 (e^{eV/kT} - 1) \end{aligned} \quad \dots(10)$$

$$\text{Here, } I_0 = eA \left[ \frac{D_p}{L_p} P_n + \frac{D_n}{L_n} N_p \right] \quad \dots(11)$$

Eq.(10) is called the *diode equation*.

$$I_0 = \text{reverse saturation current at temperature } T \text{ K.}$$

The empirical form of Eq. (10) is

$$I = I_0 [e^{eV/\eta kT} - 1] \quad \dots(12)$$

Here,  $\eta$  is a numerical constant which depends upon the material of the diode.

For *Ge*,  $\eta = 1$ . For *Si*,  $\eta = 2$ .

(i) For large forward bias voltage,  $e^{eV/\eta kT}$  is large when compared to 1.

$$\therefore I_f = I_0 e^{eV/\eta kT} \quad \dots(13)$$

Thus forward current  $I_f$  increases exponentially with voltage  $V$  except for very small values of  $V$  (Fig. 33.10).

(ii) For reverse bias voltage,  $V$  is negative and Eq. (12) becomes

$$I_R = I_0 \left[ \frac{1}{e^{V_e/\eta kT}} - 1 \right] \quad \dots(14)$$

For large reverse bias voltage, exponential term quickly becomes negligible with respect to unity. Then

$$I_R \approx -I_0 \quad \dots(15)$$

Thus reverse current  $I_R$  is constant independent of applied reverse bias.

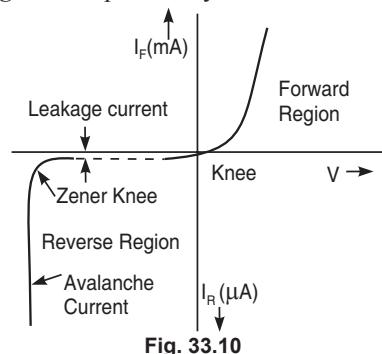


Fig. 33.10

## DIODE APPLICATIONS

### RECTIFIERS AND FILTERS

## 33.5 P-N JUNCTION DIODE AS RECTIFIER

Rectification is the process in which ac is converted into dc. The device which is used for rectification is called a rectifier. A junction diode allows a current to flow through it when it is forward biased. This property of diode is used for rectification. A half-wave rectifier is one which converts a.c. voltage into a pulsating voltage using only one half cycle of the applied a.c. voltage. A full wave rectifier is one which converts a.c. voltage into a pulsating voltage using full cycle of the applied a.c. voltage.

### 33.5.1. Diode as a Half-Wave Rectifier

**Construction.** Fig. 33.11 shows the circuit for a half-wave rectifier.  $T$  is a transformer. The primary of the transformer is connected to the ac mains. The diode  $D$  is connected across the secondary in series with a load resistance  $R_L$ .

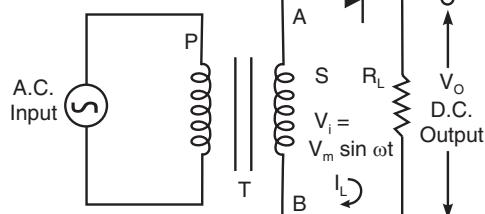


Fig. 33.11

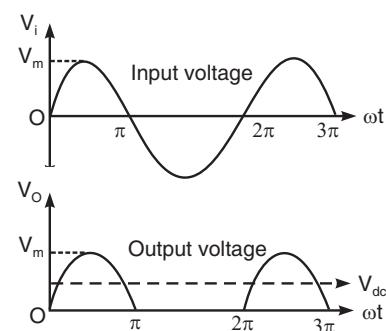


Fig. 33.12

**Working.** The primary of the transformer is connected to the ac mains. An ac voltage will be induced across the secondary. This voltage can be represented by

$$V_i = V_m \sin \omega t.$$

Fig. 33.12 shows the variation of this input voltage with time.  $V_m$  is the peak value.

During the first half cycle of a.c., one end of the secondary, say  $A$ , becomes positive. Then the diode is forward biased. Hence current flows through the load  $R_L$  in the direction of arrows (Fig. 33.11). The diode offers very little resistance when forward biased. Hence the p.d. across it is very

small. The voltage across the load  $R_L$  is therefore practically the same as that across the secondary of the transformer, i.e.,  $V_i$ . During the next half cycle, the end A becomes negative. The diode is now reverse biased. Therefore, no current flows through the load  $R_L$ . The voltage across the load is zero. The shape of the output voltage is shown in Fig. 33.12. This voltage is not a perfect dc. But it is unidirectional.

### Mathematical Analysis

The input voltage applied to the diode is given by

$$V_i = V_m \sin \omega t \quad \dots(1)$$

Then, the instantaneous output current through the load resistance  $R_L$  is given by

$$I_L = I_m \sin \omega t \quad \text{when } 0 \leq \omega t \leq \pi \quad \dots(2)$$

and

$$I_L = 0 \quad \text{when } \pi \leq \omega t \leq 2\pi \quad \dots(3)$$

$$I_m = \frac{V_m}{R_L + R_f}$$

Here,

$I_m$  = peak value of the current

$R_f$  = dynamic forward resistance of the diode.

#### (i) D.C. (average) value of output current

The average d.c. current over one complete cycle is given by

$$\begin{aligned} I_{av} &= I_{dc} = \frac{1}{2\pi} \int_0^{2\pi} I_L d(\omega t) \\ &= \frac{1}{2\pi} \left[ \int_0^{\pi} I_m \sin(\omega t) d(\omega t) + \int_0^{2\pi} 0 \times d(\omega t) \right] \\ &= \frac{I_m}{2\pi} [-\cos \omega t]_0^{\pi} \\ \therefore I_{dc} &= \frac{I_m}{\pi} = \frac{1}{\pi} \left[ \frac{V_m}{R_f + R_L} \right] \end{aligned} \quad \dots(4)$$

The dc voltage developed across the load  $R_L$  is given by

$$V_{dc} = I_{dc} \times R_L = \frac{I_m}{\pi} R_L \quad \dots(5)$$

#### (ii) D.C. power output.

The dc power output across the load  $R_L$  is

$$\text{D.C. power output} = P_{dc} = I_{dc}^2 \cdot R_L = \frac{I_m^2}{\pi^2} \cdot R_L \quad \dots(6)$$

#### (iii) R.M.S. (effective) value of output current

The root mean square value of the current, by definition, is given by

$$\begin{aligned} I_{rms} &= \left[ \frac{1}{2\pi} \int_0^{2\pi} i_L^2 d(\omega t) \right]^{\frac{1}{2}} \\ &= \left[ \frac{1}{2\pi} \left\{ \int_0^{\pi} i_m^2 \sin^2 \omega t d(\omega t) + \int_{\pi}^{2\pi} 0 \cdot d(\omega t) \right\} \right]^{\frac{1}{2}} \\ &= \left[ \frac{1}{2\pi} \int_0^{\pi} i_m^2 \sin^2 \omega t d(\omega t) \right]^{\frac{1}{2}} \end{aligned}$$

$$\therefore I_{rms} = \frac{I_m}{2} \quad \dots(7)$$

**(iv) A.C. power input :** The power supplied to the circuit from the ac source is given by

$$P_{AC} = I_{rms}^2(R_f + R_L) = \frac{I_m^2}{4}(R_f + R_L) \quad \dots(8)$$

**(v) Rectifier efficiency :** It is defined as the ratio of dc output power to the total ac power supplied to the rectifier.

$$\begin{aligned} \eta &= \frac{\text{D.C. power output}}{\text{A.C. power input}} \\ &= \frac{I_m^2 R_L / \pi^2}{I_m^2 (R_f + R_L) / 4} = \frac{4}{\pi^2} \left[ \frac{R_L}{R_L + R_f} \right] \\ \eta &= \frac{0.406}{1 + R_f / R_L} \end{aligned}$$

Efficiency is maximum when  $R_f \ll R_L$ .

Theoretical maximum efficiency  $\eta = 0.406 = 40.6\%$ .

**(vi) Ripple factor :** The ripple factor is the ratio of r.m.s. value of A.C. component to the D.C. component in the rectifier output, i.e.,

$$\gamma = \frac{\text{r.m.s. value of A.C. component of output voltage}}{\text{D.C. component of output voltage}}$$

$$\gamma = \frac{V_{ac}}{V_{dc}} = \frac{I_{ac}}{I_{dc}}$$

The effective (r.m.s.) value of total load current is given by,

$$\begin{aligned} I_{rms} &= \sqrt{I_{dc}^2 + I_{ac}^2} \\ I_{ac} &= \sqrt{I_{rms}^2 - I_{dc}^2} \\ \frac{I_{ac}}{I_{dc}} &= \frac{1}{I_{dc}} \sqrt{I_{rms}^2 - I_{dc}^2} \\ \gamma &= \sqrt{\left( \frac{I_{rms}}{I_{dc}} \right)^2 - 1} \end{aligned}$$

For a half-wave rectifier,

$$\begin{aligned} \frac{I_{rms}}{I_{dc}} &= \frac{I_m/2}{I_m/\pi} = \frac{\pi}{2} \\ \gamma &= \sqrt{\frac{\pi^2}{4} - 1} = 1.21. \end{aligned}$$

This indicates that the amount of A.C. component present in the output of a half-wave rectifier is 121% of D.C. output voltage. The half-wave rectifier is therefore a poor converter of ac into dc.

**(vii) Peak inverse voltage.** Peak inverse voltage (PIV) is defined as the *maximum reverse voltage which the rectifier has to withstand during the non-conducting period*. Thus for a half-wave rectifier,

$$PIV = V_m.$$

**(viii) Voltage regulation.** Voltage regulation is the ability of a rectifier to *maintain a specific output voltage irrespective of the variation in the load resistance*. In a half-wave rectifier,

$$I_{dc} = \frac{I_m}{\pi} = \frac{V_m}{\pi(R_L + R_f)}$$

$$\text{or} \quad I_{dc} R_L = \frac{V_m}{\pi} - I_{dc} R_f$$

$$\therefore \quad V_{dc} = \frac{V_m}{\pi} - I_{dc} R_f.$$

When  $I_{dc} = 0$ ,  $V_{dc}$  has its maximum value ( $V_m / \pi$ ). As  $I_{dc}$  increases,  $V_{dc}$  decreases linearly. Therefore, *voltage regulation of a half-wave rectifier is poor*.

**Example 1.** The applied input a.c. power to a half-wave rectifier is 100 watt. The d.c. output power obtained is 40 watt. Calculate the rectification and power efficiency of the above circuit. What happens to the remaining 60 watt?

**Sol.** The rectification efficiency is given by

$$\eta = \frac{\text{d.c. output}}{\text{a.c. input}} \times 100 = \frac{40}{100} \times 100 = 40\%$$

Half-wave rectifier rectifies only half the cycle of sinusoidal wave. The input power is only 50 W.

$$\text{Power efficiency} = \frac{\text{output power}}{\text{input power}} \times 100 = \frac{40}{50} \times 100 = 80\%.$$

**Example 2.** The turns ratio of a transformer used in a half-wave rectifier is 12 : 1. The primary of the transformer is connected to 220 V, 60 Hz mains. If the diode resistance on forward bias is zero, calculate the d.c. voltage across the load. Also find the Peak Inverse Voltage.

**Sol.** Turns ratio of a transformer is given by

$$\frac{N_p}{N_s} = \frac{V_p}{V_s}$$

$$\therefore \quad V_s = \frac{N_s}{N_p} \cdot V_p = \frac{1}{12} \times 220 = 18.33 \text{ V}$$

$$\therefore \text{Maximum voltage across secondary} \\ = \sqrt{2} V_s = \sqrt{2} \times 18.33 = 25.92 \text{ V}$$

$$\therefore \quad \text{D.C. output voltage} = \frac{V_m}{\pi} = \frac{25.92}{3.14} = 8.252 \text{ V}$$

$$\text{Peak inverse voltage} = 25.92 \text{ V.}$$

### 33.5.2. Full-Wave Rectifier

In a full-wave rectifier both halves of the input-cycle are used. There are two types of full-wave rectifiers: (1) Centre tapped full-wave rectifier, and (2) Bridge rectifier.

**Centre tapped full-wave rectifier.** A full wave rectifier circuit consists of two diodes  $D_1$  and  $D_2$  connected to the secondary of the step-down transformer. The input A.C. signal is fed to the primary of the transformer (Fig. 33.13).

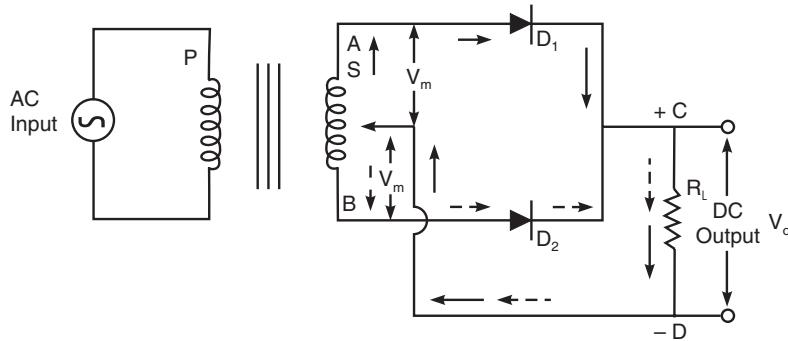


Fig. 33.13

**Working.** During the positive half-cycle of the secondary voltage, one end of the secondary, say *A*, becomes positive and end *B* becomes negative. So the diode  $D_1$  is forward biased, and diode  $D_2$  is reverse biased. As a result of this, the diode  $D_1$  conducts current whereas the diode  $D_2$  does not conduct. Current through the load resistance flows from *C* to *D* producing output voltage  $V_o$ . The current is shown by solid arrows.

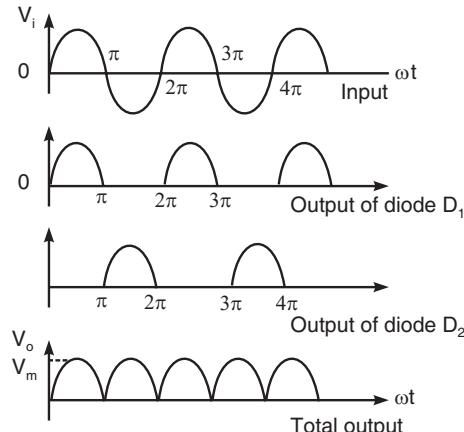


Fig. 33.14

During the negative half cycle of *AC* input, end *A* becomes negative and end *B* positive. So the diode  $D_1$  is reverse biased and the diode  $D_2$  is forward biased. As a result, the diode  $D_1$  does not conduct and  $D_2$  conducts current. Again current flows from *C* to *D* through the load resistance  $R_L$  producing output voltage  $V_o$ . The current is shown by the dotted arrows.

Thus, during both the half cycles, current flows through the load in the same direction. The output voltage is developed across the load  $R_L$  during the entire cycle. It is a pulsating D.C. voltage containing both A.C. and D.C. components. The input and the rectified output wave-forms are shown in Fig. 33.14.

### Mathematical Analysis

Let the diodes  $D_1$  and  $D_2$  be identical and have the same dynamic resistance  $R_f$ . At any instant, let the magnitudes of AC voltages applied to the diodes be each equal to  $V_i = V_m \sin \omega t$ .  $V_m$  is the peak input voltage.

Let

$R_f$  = dynamic forward resistance of the diode.

The current pulses in the two diodes are given by

$$i = \begin{cases} I_m \sin \omega t & \text{for } 0 < \omega t < \pi \\ -I_m \sin \omega t & \text{for } \pi < \omega t < 2\pi \end{cases} \quad \dots(1)$$

Here,

$$I_m = \frac{V_m}{R_f + R_L}.$$

(i) **D.C. (average) value of output current.** The output dc current  $I_{dc}$  is given by

$$\begin{aligned} I_{dc} &= \frac{1}{2\pi} \int_0^{2\pi} i d(\omega t) \\ &= \frac{1}{2\pi} \left[ \int_0^\pi I_m \sin \omega t d(\omega t) + \int_\pi^{2\pi} -I_m \sin \omega t d(\omega t) \right] \\ &= \frac{I_m}{2\pi} \left[ [-\cos \omega t]_0^\pi + [\cos \omega t]_\pi^{2\pi} \right] = \frac{I_m}{2\pi} [2 + 2] \\ \therefore I_{dc} &= \frac{2I_m}{\pi}. \end{aligned} \quad \dots(2)$$

(ii) **R.M.S. (effective) value of load current.** The r.m.s. value of total output current is given by

$$\begin{aligned} I_{rms} &= \sqrt{\frac{1}{2\pi} \int_0^{2\pi} i^2 d(\omega t)} \\ &= \left[ \frac{1}{2\pi} \left\{ \int_0^\pi I_m^2 \sin^2 \omega t d(\omega t) + \int_\pi^{2\pi} I_m^2 \sin^2 \omega t d(\omega t) \right\} \right]^{1/2} \\ \therefore I_{rms} &= \frac{I_m}{\sqrt{2}}. \end{aligned} \quad \dots(3)$$

(iii) **Power supplied to the circuit.** The a.c. power input to the rectifier from the supply is given by

$$P_{ac} = I_{rms}^2 (R_f + R_L) = \frac{(R_f + R_L) I_m^2}{2}. \quad \dots(4)$$

(iv) **Average power supplied to the load  $R_L$ .** The d.c. power output across the load  $R_L$  is given by

$$P_{dc} = I_{dc}^2 R_L = \frac{4I_m^2 R_L}{\pi^2}. \quad \dots(5)$$

(v) **Rectifier efficiency.** In a rectifier, the useful power output is the d.c. power which is developed across the load  $R_L$ . Therefore, efficiency

$$\begin{aligned} \eta &= \frac{\text{d.c. power supplied to the load}}{\text{Total input A.C. power}} \times 100\% \\ &= \frac{P_{dc}}{P_{ac}} \times 100\% = \frac{4I_m^2 R_L / \pi^2}{(R_f + R_L) I_m^2 / 2} \times 100\% \end{aligned}$$

[Using Eqs. (4) and (5)]

$$\therefore \eta = \frac{81.2}{1 + \frac{R_f}{R_L}} \% \quad \dots(6)$$

Thus, the rectification efficiency of a full-wave rectifier is double that of a half-wave rectifier under identical conditions.

The maximum possible efficiency of a full-wave rectifier is 81.2 % when  $R_f \ll R_L$ .

(vi) **Ripple factor.** The ripple factor  $\gamma$  is given by

$$\gamma = \sqrt{\left(\frac{I_{rms}}{I_{dc}}\right)^2 - 1} = \sqrt{\left(\frac{I_m / \sqrt{2}}{2I_m / \pi}\right)^2 - 1}$$

$$\therefore \gamma = 0.482$$

The ripple factor of a full-wave rectifier is 0.482 and is much smaller than that of half-wave rectifier. Hence, in actual practice, a full-wave rectifier is preferred to a half-wave rectifier.

(vii) **Peak inverse voltage.** PIV is the maximum reverse voltage which the rectifier has to withstand during the non-conduction period. Suppose the diode  $D_1$  is conducting and  $D_2$  non-conducting.

The reverse voltage across diode  $D_2$

$$\text{PIV} = \text{PD across } R_L - (-V_m) = V_m + V_m = 2V_m$$

Thus in a full-wave rectifier, PIV across each diode is two times the maximum transformer voltage measured from the centre-tap to either end.

**EXAMPLE 1.** A full-wave rectifier supplies d.c. to a load of  $1k\Omega$ . If the a. c. voltage applied to the diodes is 200-0-200 volts (rms), calculate: (a) Average d.c. voltage, (b) Average d.c. current, (c) Ripple voltage (rms), neglecting the diode resistance.

**SOL.** The peak value of dc current is given by

$$I_m = \frac{V_m}{R_L} = \frac{V_{rms} \sqrt{2}}{R_L} = \frac{200\sqrt{2}}{10^3} = 0.2828 \text{ A}$$

The average d.c. current is given by

$$I_{dc} = \frac{2I_m}{\pi} = \frac{2 \times 0.2828}{\pi} = 0.18 \text{ A}$$

Average d.c. voltage is  $V_{dc} = I_{dc} \times R_L = 0.18 \times 10^3 = 180 \text{ V}$

$$\text{Ripple factor, } \gamma = \sqrt{\frac{V_{rms}^2}{V_{dc}^2} - 1} = \sqrt{\left(\frac{200}{180}\right)^2 - 1} = 0.4843$$

$$\text{Ripple voltage } V_{rms} = \gamma \times V_{dc} = 0.4843 \times 180 = 87.17 \text{ V.}$$

**EXAMPLE 2.** In a full-wave rectifier the load resistance is  $1k\Omega$ . The forward dynamic resistance of each diode is  $10 \Omega$ . The voltage across half of the secondary winding is  $220 \sin 200 t$ . Find the (a) Peak value of the current. (b) Average dc value of the current. (c) The rms value of the current. (d) The rectification efficiency.

**SOL.** The instantaneous value of the A.C. input voltage is given by

$$V = V_m \sin \omega t$$

where  $V_m$  is peak value of the A.C. input voltage.

$$\text{Peak value of current, } I_m = \frac{V_m}{(R_f + R_L)} = \frac{220}{(10 + 1000)} = 0.2178 \text{ A}$$

The average d.c. current of full-wave rectifier is given by

$$I_{dc} = \frac{2I_m}{\pi} = \frac{2 \times 0.2178}{\pi} = 0.1387 \text{ A}$$

$$\text{r.m.s. value of current, } I_{rms} = \frac{I_m}{\sqrt{2}} = \frac{0.2178}{\sqrt{2}} = 0.154 \text{ A}$$

$$\text{Rectifier efficiency, } \eta = \frac{\text{d.c. power output}}{\text{a.c. power input}}$$

$$\eta = \frac{I_{dc}^2 R_L}{I_{rms}^2 (R_f + R_L)} = \frac{(0.1387)^2 \times 1000}{(0.154)^2 \times (10 + 1000)} = 0.803$$

Rectifier efficiency is 80.3 %.

**EXAMPLE 3.** Is the frequency content of the output of a half-wave rectifier and a full-wave rectifier the same? Explain.

**SOL.** No. Full-wave rectifier rectifies both the half-cycles. Hence the frequency of the output current is double that of the input current. But in a half-wave rectifier only one-half of the input signal is rectified. Hence the frequency of the output current is the same as that of the input current.

### 33.5.3. Bridge Rectifier

**Construction.** Fig. 33.15 shows a **bridge rectifier** circuit.

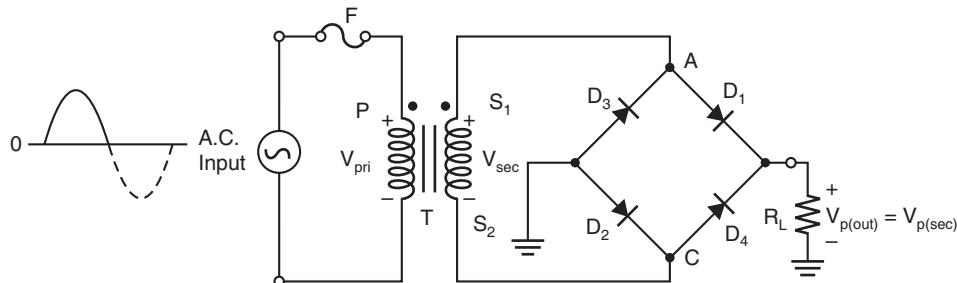


Fig. 33.15

- The diodes  $D_1, D_2, D_3$  and  $D_4$  are arranged in the form of Wheatstone Bridge network.
- The two opposite ends  $A$  and  $C$  of the network are connected to the ends  $S_1$  and  $S_2$  of the secondary of transformer  $T$ .
- The primary  $P$  of the transformer is connected to the ac mains.
- Diodes  $D_1$  and  $D_2$  conduct on the positive half cycle. Diodes  $D_3$  and  $D_4$  conduct on the negative half cycle. As a result, the rectified load current flows during both the half cycles. The secondary voltage appears across the load resistor  $R_L$ .

$$V_{p(\text{out})} = V_{p(\text{sec})} \quad \dots(1)$$

#### Working

- During positive half-cycle of the input,  $D_1$  and  $D_2$  are forward-biased and conduct current.  $D_3$  and  $D_4$  are reverse-biased.

The direction of flow of current is shown in Fig. 33.16.

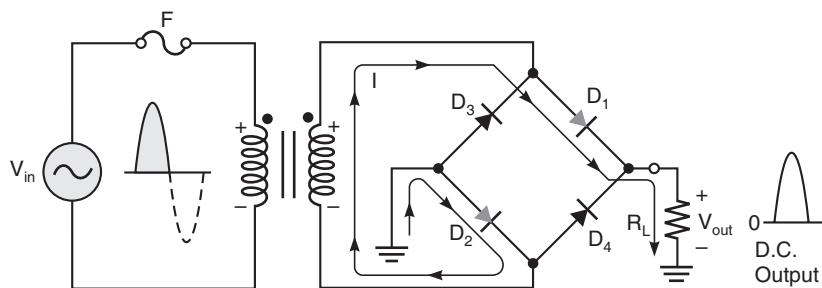


Fig. 33.16

(ii) During negative half-cycle of the input,  $D_3$  and  $D_4$  are forward biased and conduct current.  $D_1$  and  $D_2$  are reverse-biased.

The direction of flow of current is shown in Fig. 33.17.

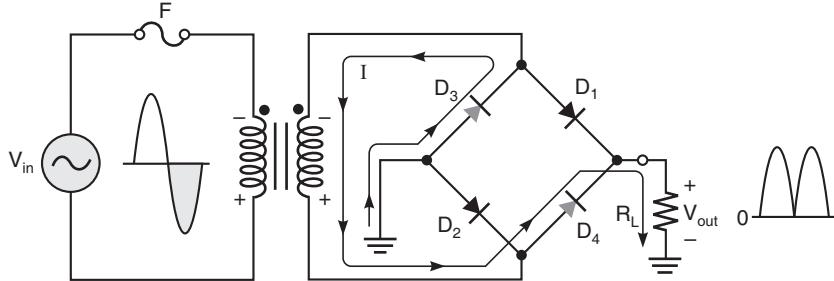


Fig. 33.17

- Therefore during both the halves of the input cycle, current flows through the load  $R_L$  in the same direction. Thus a DC output is developed across  $R_L$  and we have full wave rectification.

#### Practical diodes (Diode drops included)

Two diodes are always in series with the load resistor during both the positive and negative half-cycles. If these diode drops are taken into account, the output voltage is

$$V_{p(\text{out})} = V_{p(\text{sec})} - 1.4 \text{ V} \quad \dots (2)$$

**Peak Inverse Voltage.** Let's assume that  $D_1$  and  $D_2$  are forward-biased and examine the reverse voltage across  $D_3$  and  $D_4$ . Since the diodes are considered ideal, the diodes  $D_1$  and  $D_2$  have negligible forward resistance and can be taken to be simple connecting wires. Diodes  $D_3$  and  $D_4$  have a peak inverse voltage equal to the peak secondary voltage. Since the output voltage is ideally equal to the secondary voltage,

$$PIV = V_{p(\text{out})} \quad \dots (3)$$

The PIV rating of the bridge diodes is less than that required for the center-tapped configuration. If the diode drop is neglected, the bridge rectifier requires diodes with half the PIV rating of those in a center-tapped rectifier for the same output voltage.

#### Average value and output frequency

Because a bridge rectifier produces a full-wave output, the equations for average value and output frequency are the same as given for a full-wave rectifier:

$$V_{dc} = \frac{2V_p}{\pi} \quad \dots (4)$$

$$f_{\text{out}} = 2f_{\text{in}} \quad \dots (5)$$

Table 33.1 compares the three rectifiers and their properties.

Table 33.1. Unfiltered Rectifiers\*

	Half-wave	Full-wave	Bridge
Number of diodes	1	2	4
Rectifier input	$V_{p(2)}$	$0.5 V_{p(2)}$	$V_{p(2)}$
Peak output (ideal)	$V_{p(2)}$	$0.5 V_{p(2)}$	$V_{p(2)}$
Peak output (2d)	$V_{p(2)} - 0.7 \text{ V}$	$0.5 V_{p(2)} - 0.7 \text{ V}$	$V_{p(2)} - 1.4 \text{ V}$
DC output	$V_{p(\text{out})}/\pi$	$2 V_{p(\text{out})}/\pi$	$2 V_{p(\text{out})}/\pi$
Ripple frequency	$f_{\text{in}}$	$2f_{\text{in}}$	$2f_{\text{in}}$

\* $V_{p(2)}$  = peak secondary voltage;  $V_{p(\text{out})}$  = peak output voltage

### 33.6 FILTER CIRCUITS

The output of a rectifier is a unidirectional voltage (or current), which is pulsating. It can be considered as consisting of a pure d.c. voltage with superposed a.c. components, known as "ripples". In electronic circuits we require a constant d.c. supply. Therefore, the ripples must be removed or minimised from the rectifier output. This can be done by introducing an electrical circuit consisting of choke coils and capacitors between the rectifier and the load. Such a circuit is known as a 'filter circuit'. There are several types of filters. Here we study a few of them.

(i) **Shunt capacitor filter.** The simplest type of filter is a shunt capacitor connected across the rectifier output and the load (Fig. 33.18 a). It is an effective way of filtering the a.c. components from the output of the rectifier. The capacitor  $C$  is so chosen that its reactance ( $1/\omega C$ ) at the frequency of a.c. mains is very small as compared to the load  $R_L$ . Then the a.c. components find a low reactance shunt path through the capacitor and are mostly bypassed. Thus ripples are filtered from the output voltage.

The pulsating output voltage of the rectifier, in the absence of capacitor  $C$ , is shown in Fig. 33.18 (b). The filtering action of the capacitor is indicated in Fig. 33.18 (c). The capacitor is charged almost equal to the peak voltage of the rectifier output at the end of the first quarter-cycle (point  $p$ ).

During the subsequent half-cycle, when the rectifier output drops to zero and then rises again, the capacitor discharges slowly (exponentially) through the load. This discharge is represented by  $pq$ . At  $q$  the capacitor begins to be recharged to the peak voltage of the rectifier output (point  $r$ ), and then discharges again to the point  $s$ . This cycle is repeated. Thus the voltage across the capacitor (and hence across the load) varies as shown by the solid curve in Fig. 33.18 (c). The variations are now much less than in the rectifier output (Fig. 33.18 b). If the time-constant ( $CR$ ) of the filter is large, the capacitor discharges quite slowly and the average dc voltage across the load remains nearly constant.

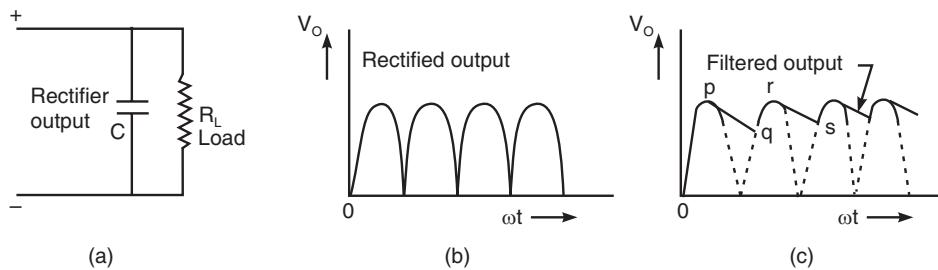


Fig. 33.18

For a full-wave rectifier with capacitor filter, the ripple factor is given by,

$$\gamma = \frac{1}{4\sqrt{3}f CR_L}$$

where  $f$  is the supply frequency.

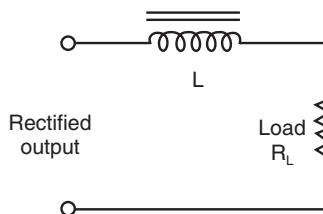


Fig. 33.19

(ii) **Series inductor filter.** In this filter, an inductor  $L$  (or choke) is placed in series with the load resistor  $R_L$  (Fig. 33.19). An inductor opposes any change in the current flowing through it. The output of the rectifier (ac + dc) is fed to the choke. The choke offers a high reactance to the ac component and a small resistance to the dc component. Consequently, the amplitude of the ac component in the output voltage is considerably reduced.

For a full-wave rectifier with series inductor filter, the ripple factor is given by

$$\gamma = 0.236 \left( \frac{R_L + R_C}{\omega L} \right)$$

where  $R_C$  is the resistance of the inductor.

**(iii) The L-section or Inductor Input Filter.** In this type of filter circuit, choke  $L$  is connected in series with the output of a rectifier and capacitor  $C$  is connected parallel to the load  $R_L$  (Fig. 33.20). The series inductor  $L$  readily passes the d.c. components from the rectifier output but offers a high impedance ( $\omega L$ ) to the ac pulsations. Any pulsations in current that remains after it passes through  $L$  are largely bypassed by the capacitor which offers a low impedance ( $1/\omega C$ ) to them, but an infinite resistance to dc. The current flowing in the load is therefore very nearly steady.

In a full-wave rectifier with inductor input filter, ripple factor

$$\gamma = \frac{0.47}{4\omega^2 LC - 1}.$$

Here, ripple factor is independent of load resistance. It is used in large current power supply.

**(iv) The  $\pi$ -section (capacitor-input) filter.** This type of filter is shown in Fig. 33.21. It may be considered to be made up of the following two components :

- (i) Capacitor filter formed by capacitor  $C_1$ .
- (ii) Inductor input filter formed by choke  $L$  and capacitor  $C_2$ .

The capacitor  $C_1$  is periodically charged to almost the peak value of the rectifier output. Between successive charging pulses, the voltage across  $C_1$  falls off somewhat due to its slow discharge through the  $L$ -section filter and the load, but remains very near the peak value. The remaining pulsations in the current are opposed by the series inductor  $L$  and bypassed to ground by capacitor  $C_2$ . In this way the ripple in the output voltage is reduced to a large extent.

In a full-wave rectifier with  $\pi$  filter, ripple factor

$$\gamma = \frac{1}{32\sqrt{2}\pi^3 f^3 C_1 C_2 L R_L}.$$

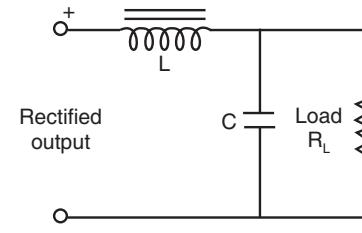


Fig. 33.20

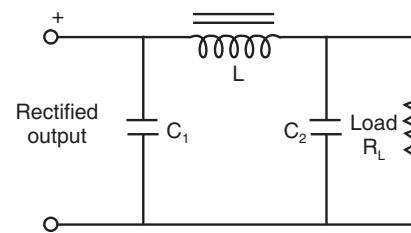


Fig. 33.21

## WAVE SHAPING CIRCUITS

### 33.7 P-N DIODE CLIPPING CIRCUITS

A wave shaping circuit which controls the shape of output waveform by removing or clipping a portion of the applied wave is known as clipping circuit.

“A clipping circuit is one in which a sine wave form is shaped by removing (clipping) a portion of the input sine wave”.

Clippers are used in radar, digital and other electronic devices.

### 1. (i) Positive Clipper :

In a positive clipper, positive half cycle of the input voltage is removed. Fig. 33.22 shows the circuit arrangement of a positive clipper.

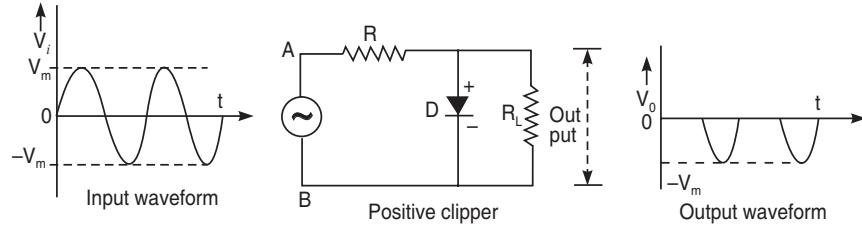


Fig. 33.22

(i) **Working:** When the positive half cycle of the signal voltage is applied to the clipper, i.e., when  $A$  is positive with respect to  $B$ , the diode  $D$  is forward-biased. Hence it acts like a closed switch (or short) and conducts heavily. Therefore the voltage drop across the diode or across the load resistor  $R_L$  is zero. So, the output voltage during the positive half-cycles is zero, i.e., positive half cycle is clipped off (removed).

(ii) In the negative half cycle of the signal voltage, i.e., when  $B$  is positive with respect to  $A$ , the diode  $D$  is reverse-biased. Hence it acts like an open switch. Consequently, the entire input voltage appears across the diode or across the load resistor  $R_L$ . So current flows through the load  $R_L$ . Hence there is an output voltage.

### (ii) Negative Clipper

In a negative clipper, negative half cycle of the input voltage is removed.

Fig. 33.23 shows the circuit of a negative clipper.

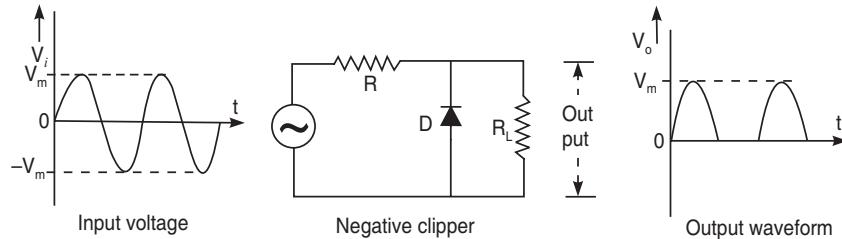


Fig. 33.23

### **Working :**

(i) During the positive half cycle of the input voltage, the diode is reverse-biased. It is not conducting. So current flows through the load  $R_L$ . Hence there is an output voltage.

(ii) During the negative half cycle, the diode is forward-biased and so it is conducting heavily. Hence no current flows through the load. There is no output voltage, i.e., the negative half of the input voltage is removed.

### 2. Biased Clipper

A biased clipper is used when it is desired to remove a small portion of positive or negative half cycle of the signal voltage.

#### (i) Biased Positive Clipper

It clips off a small portion of the positive half cycle of the input voltage at a low voltage level  $V$ .

Fig. 33.24 shows the circuit of a biased positive clipper.

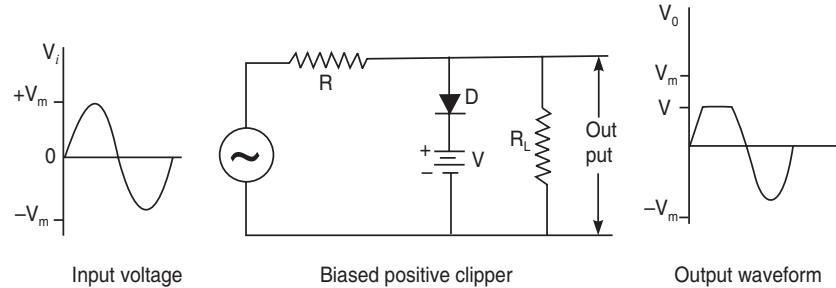


Fig. 33.24

#### **Working:**

(i) When the input signal voltage is less than  $V$  for positive half cycle: In this case, the diode  $D$  is reverse-biased, i.e., acts as an open circuit. Therefore, most of the input voltage appears across the output.

(ii) When the input signal voltage is greater than  $V$  for positive half cycle : In this case, the diode  $D$  is forward-biased, i.e., conducts heavily. The output voltage is equal to  $+V$ . The output voltage will stay at  $+V$  as long as the input voltage is greater than  $+V$ .

(iii) When the input signal voltage is negative: In this case, the diode  $D$  is reverse-biased. Hence entire input voltage appears across the output.

#### **(ii) Biased Negative Clipper**

It removes any desired portion of the negative half cycle of the input voltage.

Fig. 33.25 shows the circuit of a biased negative clipper.

The circuit is same as that for the positive clipper, except that the polarities of the diode and the associated bias supply are reversed.

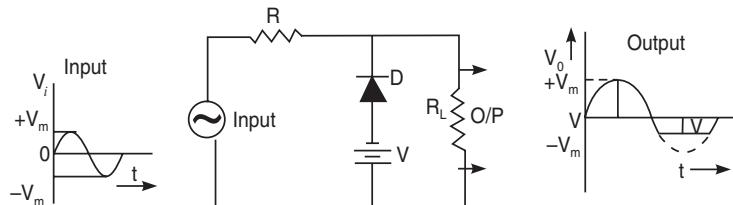


Fig. 33.25

#### **Working:**

During the entire positive half cycle, the diode  $D$  is reverse-biased and so it does not conduct. Hence the entire input voltage appears across the load.

During the negative half cycle, as long as the input voltage is less than  $-V$ , the diode is reverse-biased. So, current flows through  $R_L$  and there is an output voltage. When the input voltage exceeds  $-V$ ,  $D$  becomes forward-biased and so it conducts. Hence output voltage remains at  $-V$  equal to bias battery voltage.

#### **(3) Biased Combination Clipper**

A combined circuit of biased positive and negative clippers is called combination clipper. This is shown in Fig. 33.26.

With the combination clipper, a portion of both positive and negative half cycles of input voltage can be removed.

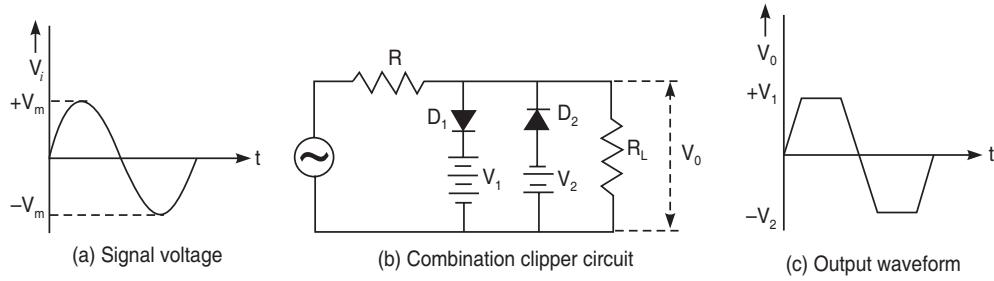


Fig. 33.26

**Working :**

- (i) When the input voltage lies between  $+V_1$  and  $-V_2$ , neither of the diodes conducts and the input voltage appears across the load, without any change.
- (ii) When positive input voltage is greater than  $+V_1$ , diode  $D_1$  is forward-biased while diode  $D_2$  is reverse-biased. Therefore, a voltage  $+V_1$  appears across the load and positive signal voltage above  $+V_1$  is clipped off.
- (iii) During negative half cycle, the diode  $D_2$  is forward-biased and  $D_1$  is reverse-biased so long as the input negative voltage is greater than  $-V_2$ . So the output stays at  $-V_2$ . Hence, signal voltage beyond  $-V_2$  volt is clipped off.

**Applications of Clippers**

Clippers are used to perform one of the following two functions:

- (i) Changing the shape of a waveform
- (ii) Circuit transient protection
- (i) **Changing the shape of waveform**
  - Clippers can alter the shape of a waveform.

**Examples**

- (i) A clipper is used to convert a sine wave into a rectangular wave, square wave etc.
- (ii) They can limit either the negative or positive alternation or both alternations of an a.c. voltage.
- (iii) Series clippers are used as noise limiters in FM transmitters by clipping excessive noise peaks above a specified level.

(ii) **Circuit Transient protection.** A transient is a sudden current or voltage rise that has an extremely short duration. A transient can cause considerable damage to many types of circuits e.g., a digital circuit. A clipper diode is used to prevent the transient from reaching that circuit.

Fig. 33.27 shows the protection of a digital circuit against transients by the diode clipper.

When the transient occurs on the input line, diode  $D_2$  gets forward biased. The diode  $D_2$  will conduct; thus shorting the transient to the ground.

So, the input of the circuit is protected from the transient.

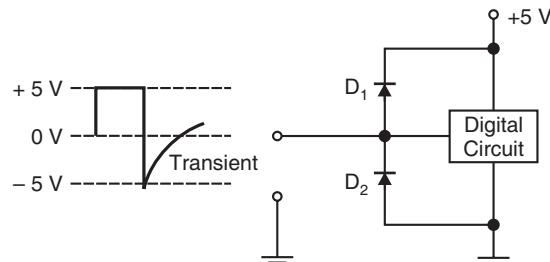


Fig. 33.27

### 33.8 P-N DIODE CLAMPING CIRCUITS

Clamping is a process of introducing a dc level into an ac signal as shown in Fig. 33.28. A clamping circuit is a device that places either the positive or negative peak of a signal at a desired level.

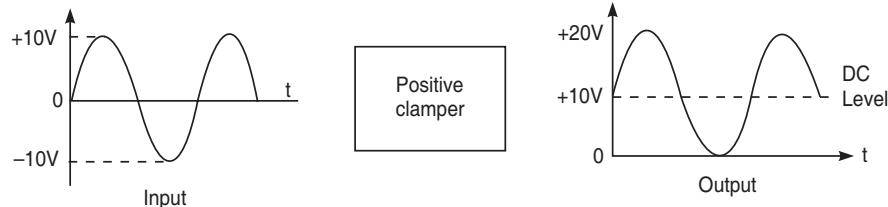


Fig. 33.28

The waveforms of input and output remain the same. But there is a vertical shift either upward or downward.

"A circuit that holds either the positive or negative peak of a signal at a desired dc level is known as clamping circuit".

Clamping circuits are widely used in TV receivers to add a dc voltage to the video signal.

#### (1) Positive Clammer

When a clamer shifts the original signal in vertical upward direction, it is known as positive clamer. Fig. 33.29 shows the circuit of a positive clamer. It consists of a diode  $D$  and a capacitor  $C$ . The clamped output is taken across the load  $R_L$ .

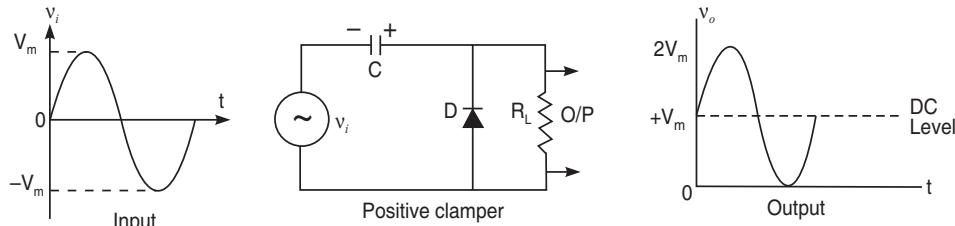


Fig. 33.29

Here we assume that diode is ideal, i.e., it exhibits an arbitrarily sharp break at 0 volt and that its forward resistance is zero. The input signal is sinusoidal which begins at  $t = 0$ . The capacitor is uncharged at  $t = 0$ . Here our aim is to find the waveform of the output  $v_o$ .

During the negative half cycle of the input voltage, the diode  $D$  is forward-biased and conducts heavily. At the maximum negative peak of applied signal, the capacitor is charged to  $V_m$  with polarities as shown in Fig. 33.29.

During the positive half cycle, the diode  $D$  is reverse-biased and does not conduct. The capacitor charged to  $V_m$  behaves as a battery.

Hence the output voltage  $v_o$  is given by

$$v_o = \text{output voltage} = V_m + v_i$$

According to the changes in  $v_i$ , there would be corresponding changes in output voltage.

It is obvious that  $V_m + v_i$  simply shifts the input voltage upward by  $V_m$ . The output voltage is obtained by shifting the input voltage upwards by  $V_m$ . The output is thus positively clamped voltage.

#### (2) Negative Clammer

When a clamer shifts the original signal in vertical downward direction, it is known as negative clamer. The input wave is pushed (i.e., shifted) downwards, so that the positive peak lies on the zero level. Fig. 33.30 shows the circuit of negative clamer.

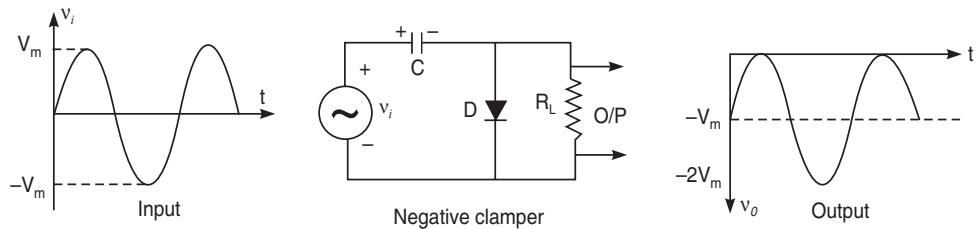


Fig. 33.30

**Working :**

During the positive half cycle, diode  $D$  is forward-biased and conducts heavily. At the maximum positive peak of the applied signal, the capacitor is charged to  $V_m$  with polarities as shown in Fig. 33.30.

The voltage across the capacitor opposes the input voltage  $v_i$ . So the output voltage is given by

$$v_o = \text{output voltage} = v_i - V_m.$$

This is equivalent to subtracting a constant voltage  $V_m$  from every point on the sine curve of  $v_i$  over the full cycle.

**EXERCISE**

1. What is a  $p-n$  junction diode? Explain the formation of a  $p-n$  junction.
2. Obtain an expression for the width of the depletion layer for a  $p-n$  junction diode. (H.P.U. 1994)
3. For a  $p-n$  junction diode derive an expression for barrier potential. (Meerut 2006)
4. Explain the volt-ampere characteristics of a  $p-n$  diode. Show that the total current through a  $p-n$  junction is given by  $I = I_s (e^{eV/kT} - 1)$  where  $I_s$  is the reverse saturation current and  $V$  the applied voltage. (Meerut 2001)
5. Derive an expression for efficiency of half wave rectifier with necessary theory and circuit diagram. (Bangalore 2002)
6. Explain with necessary theory working of a full wave rectifier. (Bharathiar University 2008)
7. Define efficiency of a rectifier. Describe the working of a Bridge rectifier. (Bangalore 2000)
8. Describe the working of a full wave rectifier and discuss the use of filters to avoid ripples. Give the diagram to illustrate your answers. (Garhwal 85; Rohilkhand 88)
9. What is a clipper circuit? Describe (i) positive clipper (ii) biased clipper. (B.U. 2011)
10. Explain the action of a negative peak clipper. (Madras 1995)
11. Mention the applications of clipper circuits. (B.U. 2012)
12. What is a clamping circuit? Explain the action of (a) positive clamper and (b) negative clamper. (Madras 1995)
13. Explain the action of a positive clamper with necessary diagrams. (B.U. 2012)

## SPECIAL-PURPOSE DIODES

**After reading this chapter, you should be able to**

- ◆ Analyze the characteristic (V-I) curve of a zener diode
- ◆ Explain how a zener is used in voltage regulation
- ◆ List several optoelectronic devices and describe how each works
- ◆ Use an LED seven-segment display
- ◆ Explain how a photodiode detects light
- ◆ Discuss the characteristics of the tunnel diode and explain its negative resistance
- ◆ Explain why a reverse-biased varactor exhibits capacitance

### 34.1 THE ZENER DIODE

The diode which operates in the reverse breakdown region with a sharp breakdown voltage is called a *Zener diode* (Fig. 34.1).



Fig. 34.1

It is an ordinary *P-N* junction diode except that it is properly doped to have a very sharp and almost vertical breakdown. It is exclusively operated under reverse bias conditions. It is designed to operate in breakdown region without damage. By adjusting the doping level it is possible to produce zener diodes with a breakdown voltage ranging from 2V to 800 V.

Zener diode primarily depends for its working on *Zener Effect*. In a heavily doped diode, the depletion region is very narrow. When the reverse bias voltage across the diode is increased, the electric field across the depletion region becomes very strong. When this field is  $\approx 3 \times 10^7$  V/m, electrons are pulled out of the covalent bonds. A large number of electron-hole pairs are thereby produced. The reverse current rises steeply. This is *Zener effect*.

The external applied voltage accelerates the minority carriers in the depletion region. These carriers gain sufficient energy to ionise atoms by collision. The electrons produced thereby accelerate to sufficiently large velocities to be able to ionise other atoms. This creates a sort of chain reaction. The cumulative effect of this chain reaction is the *avalanche effect*.

The zener effect is more prominent at breakdown voltages less than 4 V. The avalanche effect is more prominent above 6 V. Between 4 V and 6 V both effects are present. But, ordinarily, all diodes which are operated in the breakdown region of their reverse characteristics are called zener diodes.

#### 34.1.1. Experiment to Study the Characteristics of the Zener Diode

(i) **Forward Bias.** The circuit connections are made as shown in Fig. 34.2. The diode is given a forward biasing. The voltage is varied in regular steps of 0.1 volts by adjusting the rheostat.

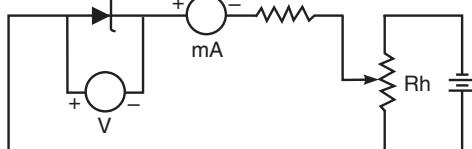


Fig. 34.2

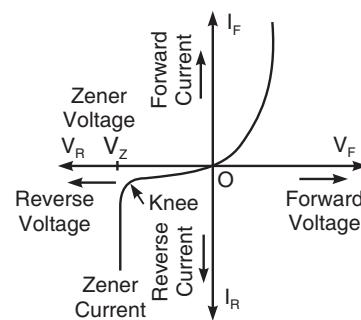


Fig. 34.3

The corresponding current is recorded using a milliammeter. The readings are tabulated. A graph is plotted taking voltage along the  $x$ -axis and current along the  $y$ -axis. Typical current-voltage characteristics for a zener diode are shown in Fig. 34.3. When forward biased, its characteristics are just that of an ordinary semiconductor diode.

(ii) **Reverse Bias.** Connections are made as shown in Fig. 34.4. The voltage is varied in regular steps of 1 volt by adjusting the rheostat. The corresponding current is recorded using milliammeter and tabulated as shown. A graph of voltage against reverse current is drawn (Fig. 34.3).

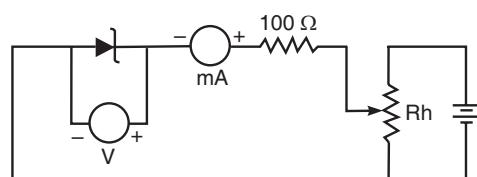


Fig. 34.4

Forward Bias		Reverse Bias	
V (volt)	I (mA)	V (volt)	I (mA)

When reverse biased, a small reverse saturation current flows through it. This current remains approximately constant until a certain critical voltage, called breakdown voltage, is reached. Beyond this voltage, the reverse current  $I_R$  increases sharply to a high value. This breakdown voltage  $V_z$  is called the *Zener voltage* and the reverse current as zener current.

$V_z$  remains constant even when Zener current  $I_z$  increases considerably. This ability of a diode is called regulating ability which enables us to use zener diode for voltage regulation.

### 34.1.2. Zener Diode as Voltage Regulator

In a voltage regulated power supply unit, the output voltage is constant and it is independent of the variations of input supply voltage and load resistance. The circuit diagram of a Zener diode voltage regulator is shown in Fig. 34.5. The unregulated dc is applied across the diode through a series resistor  $R$ , which limits the input current. The value of the series resistor  $R$  is so chosen that initially the diode operates in the breakdown region. The  $P$ -junction of the Zener diode is connected to the negative of the input voltage and  $N$ -junction to the positive. Thus the Zener diode is reverse biased. The output voltage  $V_0$  remains essentially constant (equal to  $V_z$ ) even though the input voltage  $V_i$  and the load resistance  $R_L$  may vary over a wide range.

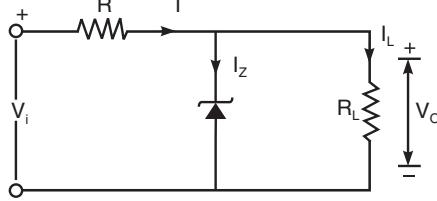


Fig. 34.5

**Working.** Let  $I$  be the current drawn from the supply source,  $I_z$  the current through the Zener diode and  $I_L$  that across the load resistance  $R_L$ .

Applying Kirchhoff's laws, we get

$$I = I_z + I_L \quad \dots(1)$$

$$V_0 = V_i - IR \quad \dots(2)$$

and  $V_0 = I_L R_L \quad \dots(3)$

The variation in the output voltage may be due to two causes. First, the load current may vary. Second, the input voltage may vary.

*Variation of load current :* Suppose the load resistance  $R_L$  varies and the input voltage  $V_i$  remains constant. Since the output voltage  $V_0$  tends to remain constant, Eq. (2) gives

$$\delta I = 0 \quad (\because V_i \text{ and } R \text{ are constant})$$

$$\text{Then Eq. (1) gives} \quad \delta I = \delta I_z + \delta I_L = 0$$

$$\text{or} \quad \delta I_z = -\delta I_L.$$

Thus, if the load resistance increases, when the supply voltage is fixed, the load current  $I_L$  decreases and the Zener diode current  $I_z$  increases by an equal amount. Thus the voltage  $V_0$  across the load will tend to remain constant.

*Variation in input voltage :* Now suppose that the load resistance  $R_L$  remains constant and supply voltage  $V_i$  varies. Since  $V_0$  tends to remain constant, we get from Eq. (2),

$$\delta V_i = R \delta I$$

$$\text{Also Eq. (3) gives,} \quad \delta I_L = 0 \quad (\because R_L \text{ is constant})$$

$$\therefore \text{Eq. (1) gives,} \quad \delta I = \delta I_z$$

Thus when the supply voltage varies but the load resistance remains constant, the *total current I and the Zener current  $I_z$  change equally to keep the load current  $I_L$  constant*. Thus if total current  $I$  decreases by  $\delta I$ , the diode current  $I_z$  also decreases by the same amount, so that load current  $I_L$  remains constant and the *voltage  $V_0$  across the load will tend to remain constant*.

**EXAMPLE 1.** Distinguish a Zener diode from an ordinary junction diode.

- SOL.**
1. Ordinary junction diodes are operated within the breakdown voltage in reverse biased condition. Operation beyond breakdown voltage may damage them.  
But a Zener diode is specially designed for operation beyond breakdown voltage. This breakdown voltage is called the *Zener breakdown potential*.
  2. Ordinary diode is usually used for rectification while Zener diode is used for voltage regulation.

**EXAMPLE 2.** A 9 volt stabilized voltage supply is required to run a car stereo system from the car's 12 volt battery. A Zener diode with  $V_z = 9\text{ V}$  and  $P_{\max} = 0.27\text{ W}$  is used as voltage regulator. Find the value of the series resistor. If the load resistance is  $450\Omega$ , find the diode current.

**SOL.** Let

$I$  – current drawn from the battery

$I_z$  – current through Zener diode

$I_L$  – current through load resistance (Fig. 34.6).

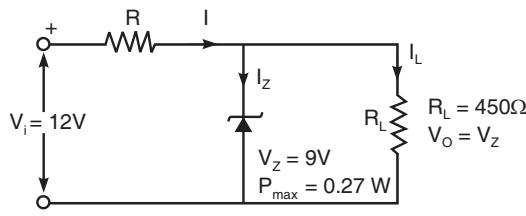


Fig. 34.6

Voltage drop across  $R = V_i - V_L = V_i - V_z = 12 - 9 = 3\text{ V}$ .

Potential difference across  $R_L = V_z = V_0 = 9\text{ V}$ .

$$\therefore \text{Load current } I_L = \frac{\text{p.d. across } R_L}{\text{load resistance}} = \frac{9}{450} = 0.02\text{ A}.$$

$$\text{Zener current } I_z = \frac{P_{\max}}{V_{\max}} = \frac{0.27}{9} = 0.03\text{ A}.$$

$\therefore$  The current drawn from the battery

$$I = I_z + I_L = 0.03\text{ A} + 0.02\text{ A} = 0.05\text{ A}.$$

$\therefore$  The value of series resistor,

$$R = \frac{\text{voltage drop across } R}{\text{current through } R} = \frac{3}{0.05} = 60\Omega.$$

## 34.2 OPTOELECTRONIC DEVICES

- Optoelectronics is the technology that combines optics and electronics. This field includes many devices based on the action of a *pn* junction.

### 34.2.1. Light Emitting Diode (LED)

A light emitting diode (LED) is a specially made forward-biased P-N junction diode which emits visible light when energized.

### Theory

When a junction diode is forward-biased, electrons from *n*-side and holes from *p*-side move towards the depletion region and they recombine. During this process, energy is released because electrons make transition from conduction band (higher energy level) to valence band (lower energy level) [Fig. 34.7].

If  $E_g$  is the semiconductor band gap, then the energy  $E_g = h\nu = \frac{hc}{\lambda}$  is emitted in the form of radiation. The corresponding emission wavelength is given by

$$\lambda = \frac{hc}{E_g}.$$

In case of Gallium Arsenide Phosphide (Ga As P), band gap  $E_g = 1.9$  eV, and we get

$$\lambda = \frac{hc}{E_g} = \frac{6.62 \times 10^{-34} \times 3 \times 10^8}{1.9 \times 1.6 \times 10^{-19}} \text{ m} = 653.3 \text{ nm (red)}.$$

### Construction

At first an *N*-type layer is grown on a substrate and then a *P*-type layer is deposited on it by the process of diffusion [Fig. 34.8]. Metal contacts (Anode) are made at the outer edge of the *P*-layer so that more upper surface is left free for light to escape. For making Cathode connections, a metal film (preferably gold) is coated at the bottom of the substrate. This film also reflects as much light as possible to the surface of the device.

*Ga As P* LED emits red light when forward-biased [Fig. 34.9 (a)]. Fig. 34.9 (b) shows the schematic symbol of LED.

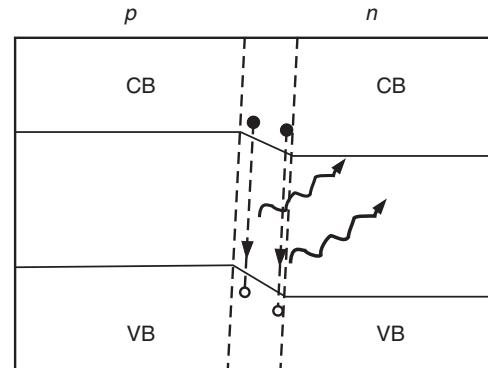


Fig. 34.7

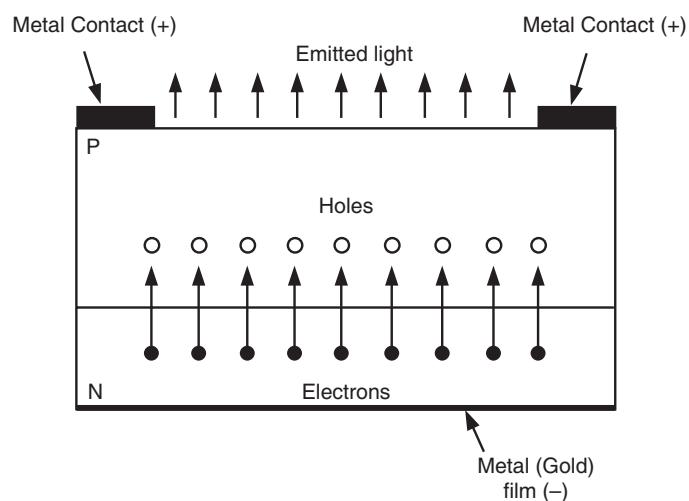


Fig. 34.8

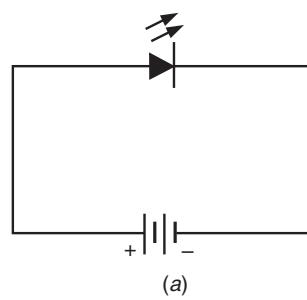
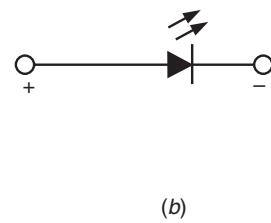


Fig. 34.9



### LED Voltage and current

From 34.10 (a) shows a source connected to a resistor and an LED. The outward arrows symbolize the radiated light.

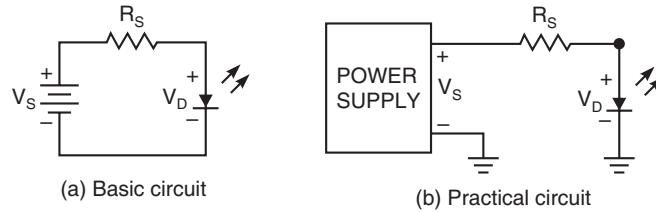


Fig. 34.10

The resistor of Fig. 34.10 (b) is the current-limiting resistor that prevents the current from exceeding the maximum rating of the diode. The resistor has a node voltage of  $V_s$  on the left and a node voltage of  $V_D$  on the right.

So the voltage across the resistance is the difference between the two voltages. With Ohm's law, the series current is

$$I_s = \frac{V_s - V_D}{R_s} \quad \dots (1)$$

For most commercially available LEDs, the typical voltage drop is from 1.5 to 2.5 V for currents between 10 to 50 mA. The exact voltage drop depends on the LED current, color, tolerance, and so on.

- The brightness of an LED depends on the current. When  $V_s$  is much greater than  $V_D$  in Eq. (1), the brightness of the LED is approximately constant.

### Applications of LEDs

- (i) **Seven-segment display.** LEDs can be used in numeric displays such as seven segment display. Fig. 34.11 shows a seven segment display unit.

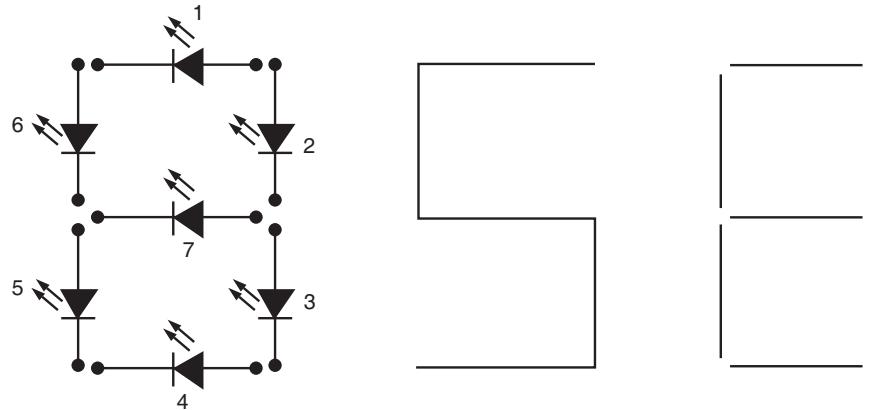


Fig. 34.11

**Working.** It is made by using a number of LEDs. By activating suitable combination of LEDs in this unit, any digit from 0 to 9 can be displayed by it. For example, switching ON the segments 1, 6, 7, 3, 4 displays the numerical 5.

When all the segments are ON, the digit displayed is 8. If only the segment 7 is OFF, the digit displayed is zero.

- (ii) **Diodes constructed of GaAs emit light in the infrared (invisible) zone during the recombination process at the p-n junction.** Even though the light is not visible, infrared LEDs have numerous applications where visible light is not a desirable effect. These include security systems,

industrial processing, optical coupling, safety controls such as on garage door openers, and in home entertainment centers, where the infrared light of the remote control is the controlling element.

**Disadvantage.** LEDs draw considerable current in comparison to other types of visual displays.

#### Multicolour LEDs

- A LED that emits one colour when forward biased and another colour when reverse biased is called a multicolour LED.

Multicolour LEDs are available that will :

- emit one colour when supply voltage is one polarity.
- emit a second colour when polarity is reversed.
- emit a third colour when bias polarity is rapidly switched.

**Construction.** Fig. 34.12 shows the schematic of multicolour LEDs. Multicolour LEDs are usually two LEDs connected in antiparallel. That is, the anode of each diode is connected to the cathode of the other.

**Working.** Each LED can emit light only when forward-biased. Thus, when voltage of either polarity is applied, one LED is forward-biased and emits its native colour.

LEDs most commonly used are red and green in colour. The green LED is normally used to indicate whether something is functioning properly, and the red LED is used to indicate that there is a problem. If a multicolour LED is rapidly switched between two polarities, the red/green LED appears to produce a third colour (yellow).

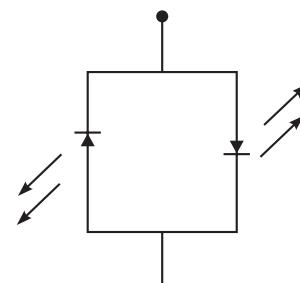


Fig. 34.12

### 34.3 PHOTOCONDUCTORS

If radiation falls upon a semiconductor, its conductivity increases. This photoconductive effect is explained as follows: Radiant energy supplied to the semiconductor ionizes covalent bonds; that is, these bonds are broken, and hole-electron pairs in excess of those generated thermally are created. These increased current carriers decrease the resistance of the material. Hence such a device is called a *photoresistor, or photoconductor*.

Fig. 34.13 shows the energy diagram of a semiconductor having both acceptor and donor impurities. If photons of sufficient energies illuminate this specimen, *photogeneration* takes place. The following three transitions are possible :

- An electron-hole pair can be created by a high-energy photon. This is called *intrinsic excitation*.
- A photon may excite a donor electron into the conduction band.
- A valence electron may go into an acceptor state.

The last two transitions are called *impurity excitations*.

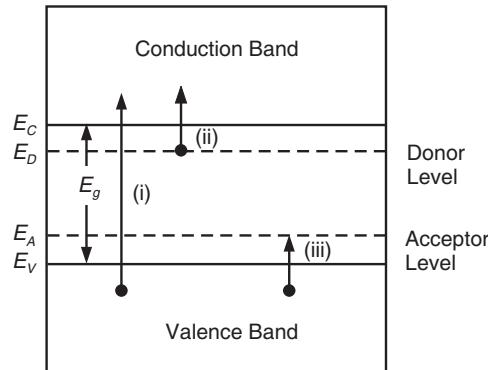


Fig. 34.13

The density of states in the conduction and valence bands is much greater than the density of impurity states. So photoconductivity is due principally to intrinsic excitation.

### Cut-off Wavelength and Spectral Response

The minimum energy of a photon required for intrinsic excitation is the forbidden-gap energy ( $E_g$ ) of the semiconductor material. Hence the minimum frequency  $v_c$  to cause photoconduction is given by

$$h v_c = E_g.$$

Corresponding maximum wavelength or critical wavelength is given by

$$\frac{hc}{\lambda_c} = E_g.$$

If  $\lambda_c$  is expressed in microns and  $E_g$  in electron volts,

$$\lambda_c = \frac{1.24}{E_g}.$$

If the wavelength  $\lambda$  of the radiation exceeds  $\lambda_c$ , then the energy of the photon is less than  $E_g$ . Such a photon cannot cause a valence electron to enter the conduction band. Hence  $\lambda_c$  is called the *critical*, or *cutoff wavelength*, or *long wavelength threshold*, of the material. For Si,  $E_g = 1.1$  eV and  $\lambda_c = 1.13$   $\mu\text{m}$ , whereas for Ge,  $E_g = 0.72$  eV and  $\lambda_c = 1.73$   $\mu\text{m}$  at room temperature.

Fig. 34.14 gives spectral response versus wavelength curve for Ge and Si.

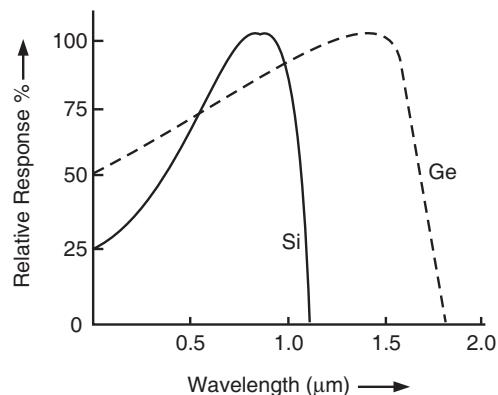


Fig. 34.14

### 34.4 PHOTODIODE

**Principle.** A reverse-biased *pn* junction diode has a reverse saturation current which is mainly due to the flow of the minority carriers. These carriers exist because thermal energy keeps dislodging valence electrons from their orbits, producing free electrons and holes in the process. The lifetime of the minority carriers is short, but while they exist, they can contribute to the reverse current.

When light is incident on the depletion region of the reverse-biased *p-n* junction, the concentration of minority carriers (electrons in *p* type and holes in *n* type) increases to a great extent. But the change in majority carriers is too low. Consequently reverse current increases. The reverse current through the diode varies almost linearly with the intensity of light. The more light striking the junction, the larger the reverse current in a diode. A **photodiode** has been optimized for its sensitivity to light.

A photo diode is essentially a reverse-biased *p-n* junction diode which is designed to respond to photon absorption.

Figure 34.15 shows the schematic symbol of a photodiode. The arrows represent the incoming light.

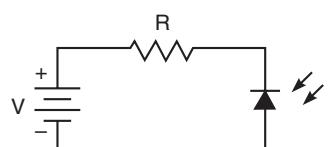


Fig. 34.15

- The source and the series resistor reverse bias the photodiode.
- As the light becomes brighter, the reverse current increases.

**Construction.** A photo diode consists of a *P-N* junction embedded in a clear plastic capsule [Fig. 34.16]. A window lets light pass through the package to the junction. All the sides of the plastic capsule, excepting the illuminated one, are either painted black or enclosed in a metallic case.

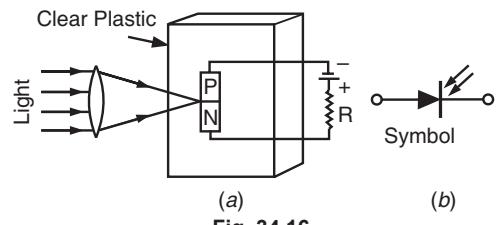


Fig. 34.16

### Working and Characteristics

- The region of operation of a photodiode is limited to the reverse-bias region.

When photo diode is kept under dark condition and a sufficient reverse voltage is applied, then an almost constant current, independent of magnitude of reverse bias, is obtained. This current corresponds to the reverse saturation current due to thermally generated minority carriers. It is called *dark current*. It is proportional to the concentrations of minority carriers and is denoted by  $I_d$ . Majority charge carriers are not allowed to cross the junction by the potential hill under this reverse bias condition.

When light falls on the diode surface, additional electron-hole pairs are formed. But since the concentration of majority carriers is much greater as compared to that of minority carriers, the percentage increase of majority carriers is much smaller than the percentage increase of minority carriers. Hence, we can neglect the increase in majority carrier density and can consider the radiation entirely as a *minority carrier injector*. These injected minority carriers diffuse to the junction, cross it and contribute to the additional current.

Thus under large reverse bias conditions, the total reverse current is given by

$$I = I_s + I_d$$

where  $I_s$  is the short circuit current and is proportional to light intensity.

With any bias  $V$ , the reverse current due to thermal electron-hole pairs, i.e., dark current is, given by

$$I_d = I_0 (1 - e^{V_e / \eta kT})$$

Hence, the volt-ampere characteristic of photo diode is given by

$$I = I_s + I_0 (1 - e^{V_e / \eta kT})$$

where  $\eta$  is equal to 1 for Ge and equals 2 for Si.

The volt-ampere characteristic curve of a photo diode is as shown in Fig. 34.17. From the curve it is seen that

- the current increases with increase in the level of illumination for a given reverse voltage.
- only for the dark current at zero voltage the current is zero.

### Applications

The photodiode finds extensive application in light detection systems, reading of film sound track, light operated switches, high-speed reading of computer punched cards and tapes.

(i) **Using a photodiode in an alarm system.** In Fig. 34.18, the photodiode is employed in an alarm system.

- The reverse current  $I$  will continue to flow as long as the light beam is not broken.
- If the light beam is interrupted, reverse current  $I$  drops to the dark current level and sounds the alarm.

(ii) **Using a photodiode in a counter operation.** In Fig. 34.19, a photodiode is used to count items on a conveyor belt.

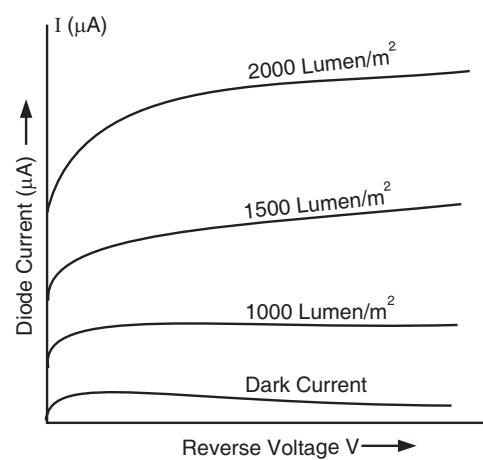


Fig. 34.17

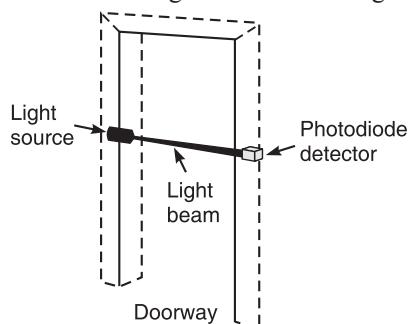


Fig. 34.18

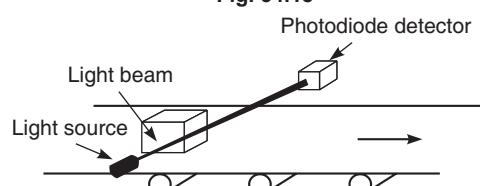


Fig. 34.19

- As each item passes, the light beam is broken, the reverse current  $I$  drops to the dark current level, and the counter is increased by one.

### 34.5 LIQUID CRYSTAL DISPLAY (LCD)

**Principle.** Molecules in an ordinary liquid normally have random orientations. But molecules in liquid crystals are oriented in a definite crystal pattern [Fig. 34.20(a)].

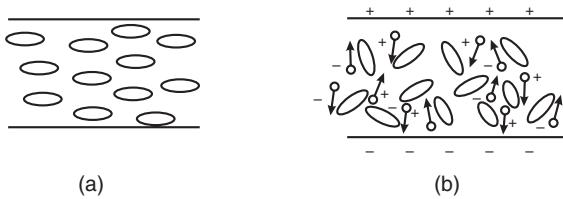


Fig. 34.20

When an electric field is established in a liquid crystal, the molecules which are approximately cigar shaped tend to orient themselves perpendicularly to the electric field. Charge carriers flowing through the liquid disrupt the molecular orientation and cause turbulence in the liquid [Fig. 34.20(b)].

When not activated, the liquid crystal is transparent. When an electric field is applied, molecular turbulence causes light to be scattered in all directions making the activated area appear bright. This phenomenon is called **dynamic scattering**. This principle is used in dynamic scattering LCD. We do not cover this type of dynamic-scattering units.

- Another type of unit of major interest is *field-effect* LCD.

**Principle.** This display works on the principle of polarised light. Two polarisers are arranged at the top and bottom. In between we have liquid crystal molecules capable of rotating polarised light by  $90^\circ$ .

*The field-effect or twisted nematic LCD* has the same thin layer of encapsulated liquid crystal, but its mode of operation is very different. The field-effect LCD can be operated in the reflective or the transmissive mode with an internal source. We shall discuss the reflective type field-effect LCD in some detail in this section.

#### Reflective-type field-effect LCD

**Construction.** Fig. 34.21 shows the reflective-type field effect LCD with no applied bias.

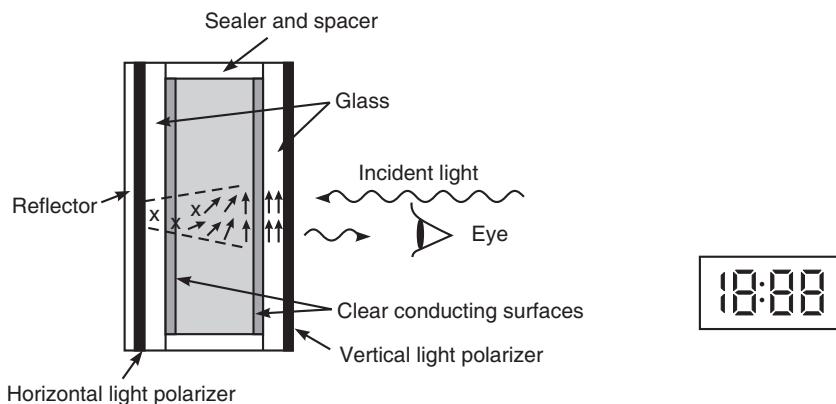


Fig. 34.21

Fig. 34.22

#### Working

(i) **No applied voltage.** Light is incident through a vertical polarizer. Inside the liquid crystal layer, the molecular orientations are graded continuously, the extreme right crystal has vertical

orientaion and the extreme left crystal has horizontal orientation. The change of orientation from vertical to horizontal occurs progressively as shown by arrows.

As the vertical component of light travels through the layer of liquid crystal, the direction of polarization alters gradually. The horizontal component of light after passing through the horizontal polarizer gets reflected by the reflector to the right side. This light beam then traces a reverse sequence of transformation and reappears through the vertical polarizer at some point resulting into a bright spot on the LCD. Thus, in presence of light, the entire LCD panel appears bright when there is no signal (voltage) across the conducting surfaces of the cell.

- There is a uniformly lit display with *no applied bias*.

(ii) **Application of voltage.** Whenever a signal (voltage) is applied across any segment, the resulting electric field disturbs the orientation of molecules of the crystal and light after arriving at the extreme left no longer remains horizontal. It thus fails to pass through the horizontal polarizer and is lost resulting into a dark display of segments. By activating an appropriate combination of segments, a numeric (character) display appears in dark against a bright background.

A dark area results on the crystal, and the pattern shown in Fig. 34.22 appears.

**Advantage.** The liquid-crystal diplay (LCD) has the distinct advantage of having a lower power requirement than the LED, typically on the order of microwatts for the display, compared to the order of milliwatts for LEDs. So LCDs are widely used in digital watches and calculators.

#### Disadvantages

(1) LCDs require an external or internal light source, and is limited to a temperature range of about 0°C to 60°C. Lifetime is an area of concern because LCDs can chemically degrade.

(2) A further consideration in displays is turn-on and turn-off time. LCDs are characteristically much slower than LEDs. LCDs typically have response times in the range 100 ms to 300 ms, whereas LEDs are available with response time below 100 ns.

## 34.6 SOLAR CELL

In recent years, there has been increasing interest in the solar cell as an alternative source of energy. The power density received from the sun at sea level is about  $1 \text{ kW/m}^2$ . It is certainly an energy source that requires further research and development to maximize the conversion efficiency from solar to electrical energy.

**Principle.** A *solar cell* is basically a *p-n* junction diode which converts solar energy (light energy) into electrical energy. In principle, a solar cell is nothing but a light emitting diode (LED) operating in reverse.

**Materials.** Common materials for solar cells include silicon, gallium Arsenide (Ga As), indium Arsenide (In As) and cadmium Arsenide (Cd As). The most common is silicon. For silicon, the band gap (the energy necessary to transfer an electron from the upper valence level to the conduction band) is 1.12 eV. The maximum theoretical efficiency of a solar cell depends on this band gap. For silicon, the maximum efficiency is 22%.

**Construction.** Fig. 34.23 shows the basic construction of a silicon *p-n* junction solar cell.

A solar cell consists of *P-N* junction diode made of Si. The *P-N* diode is packed in a can with glass window on top so that light may fall upon *P* and *N* type materials. As shown in the top view, the surface area perpendicular to the sun is a maximum. The thickness of the *P*-region is kept very small so that electrons generated in this region can diffuse to the junction before recombination takes place. Thickness of *N*-region is also kept small to allow holes generated near the surface to diffuse to the junction before they recombine. A heavy doping of *P* and *N* regions is recommended to obtain a large photo voltage. A nickel plated ring is provided around the *P*-layer which acts as the

positive output terminal. A metal contact at the bottom serves as the negative output terminal. Under open-circuit conditions ( $I_d = 0$ ), the *photovoltaic* voltage  $V_{oc}$  will result.

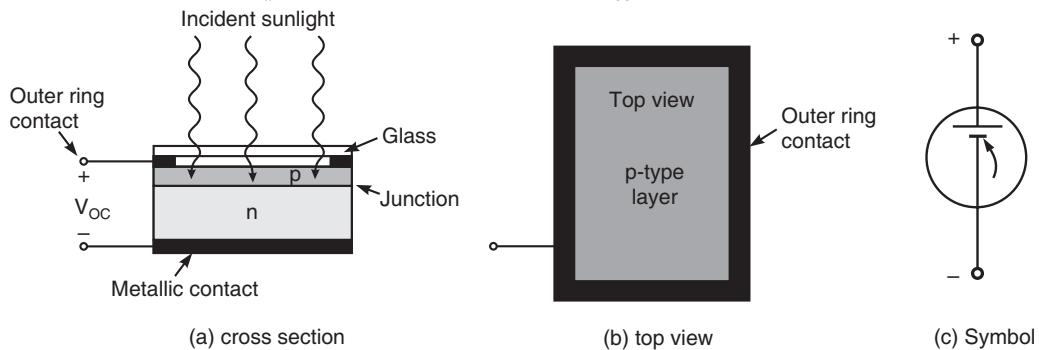


Fig. 34.23

When the *p*-and *n*-type sides are connected externally through an electrical load  $R_L$ , the current flowing across the external load is  $I_L$  (Fig. 34.24).

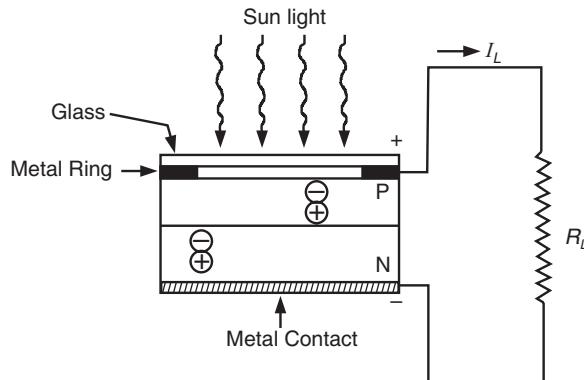


Fig. 34.24

### Working

The working of a solar cell may be understood from Fig. 34.25. When light radiation falls on a *P-N* junction diode, photons collide with valence electrons and impart them sufficient energy enabling them to leave their parent atoms. Thus electron-hole pairs are generated in both the *P* and *N* sides of the junction. These electrons and holes reach the depletion region *W* by diffusion [Fig. 34.25 (a)] and are then separated by the strong barrier field existing there. However, the minority carrier electrons in the *P*-side slide down the barrier potential to reach the *N*-side and the holes in the *N*-side move to the *P*-side [Fig. 34.25 (b)]. Their flow constitutes the minority current which is directly proportional to the illumination and also depends on the surface area being exposed to light.

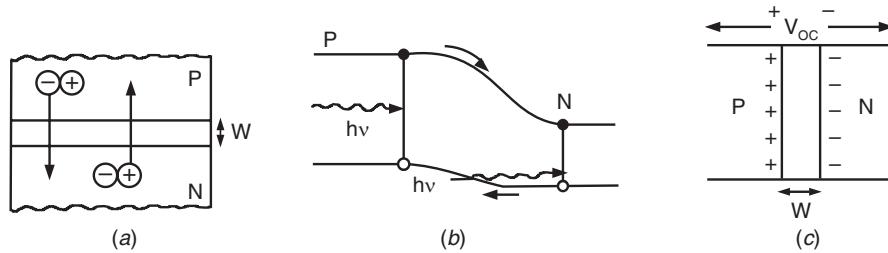


Fig. 34.25

The accumulation of electrons and holes on the two sides of the junction [Fig. 34.25 (c)] gives rise to an *open circuit voltage*  $V_{oc}$  which is a function of illumination. The open-circuit voltage produced for a silicon solar cell is typically 0.6 volt and the short-circuit current is about  $40 \text{ mA/cm}^2$  in bright noon day sun light.

### Characteristics

Typical V-I characteristics of a solar cell, corresponding to different levels of illumination are shown in Fig. 34.26. Maximum power output is obtained when the cell is operated at the knee of the curve.

### Relation between Load Current and Load Voltage

Holes and electrons produced in the junction region are swept to the *p*-type and *n*-type sides, respectively, as shown in Fig. 34.25. This produces a current  $I_s$  across the junction and also acts to charge the *p*-region positively and the *n*-type region negatively. Hence if there are no external connections to the junction, this forward bias causes a forward current  $I$  to flow. Under this condition, the forward current just balances the current  $I_g$ . When the *p*-and *n*-type sides are connected externally through an electrical load (Fig. 34.27), a portion of  $I_g$  flows in the external circuit.

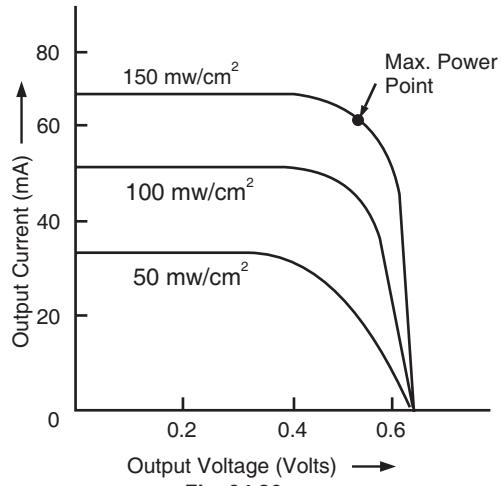


Fig. 34.26

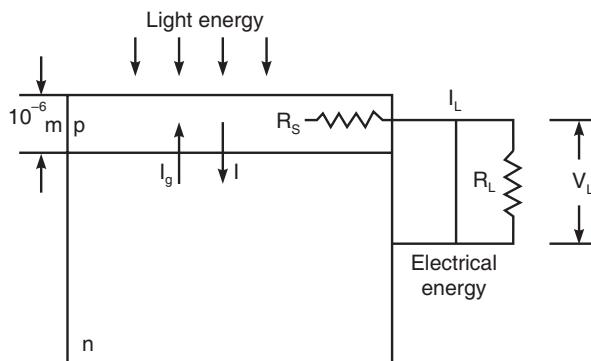


Fig. 34.27

The current  $I_g$  consists of that flowing across the junction and that flowing across the external load  $I_L$ .

$$I_g = I + I_L = I_o [\exp(eV/kT) - 1] + I_L \quad \dots(1)$$

The *p*-type region being made very thin, introduces an internal series resistance  $R_s$  in the solar battery which reduces its efficiency. The voltage drops in the circuit are

$$V = I_L R_S + V_L \quad \dots(2)$$

Eqs. (1) and (2) are combined to give the relation between load current and load voltage for a given  $I_g$ .

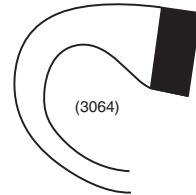
$$V_L = \frac{kT}{e} \log \left( 1 + \frac{I_g - I_L}{I_o} \right) - R_S I_L \quad \dots(3)$$

**Efficiency.** The efficiency of operation of a solar cell is determined by the electrical power output divided by the power provided by the light source.

$$\eta = \frac{P_o(\text{electrical})}{P_i(\text{light energy})} \times 100\% = \frac{P_{\max(\text{device})}}{(\text{area in cm}^2)(100 \text{ mW/cm}^2)} \times 100\%$$

**Typical solar cells and their electrical characteristics.**

Part Number	Active Area	Test Voltage	Minimum Current at Test Voltage
3013	0.032 in <sup>2</sup> (0.21 cm <sup>2</sup> )	0.4V	4.2 mA
3064	0.32 in <sup>2</sup> (2.1 cm <sup>2</sup> )	2V	8.4 mA



### Uses

1. Solar cells are used extensively in satellites and space vehicles to supply power to electronic and other equipments or to charge storage batteries.
2. Solar cells are used to power calculators, watches, and mobile phones.
3. Solar cells are used to provide commercial electricity.

## 34.7 TUNNEL DIODE

It is a *p-n* junction semiconductor diode in which the concentration of impurity atoms is very large in *p* and *n* regions ( $\approx 10^{24}/\text{m}^3$ ). The width of the depletion region is very small ( $\approx 10^{-8}$  m). Since the depletion region is very narrow, electrons are capable of tunneling through from one side of the junction to the other at relatively low forward bias voltage. This phenomenon is called *tunneling*. This type of diode is called Tunnel Diode (Fig. 34.28)



Fig. 34.28

### The Volt-Ampere (V-I) Characteristics

The V-I characteristics of a tunnel diode are shown in Fig. 34.29. When forward voltage increases from zero, the forward current quickly reaches the peak value  $I_p$  (point *A*) at a particular low forward voltage  $V_p$ . When the forward voltage exceeds the value  $V_p$ , the forward current decreases and reaches a minimum value called valley current  $I_v$  at valley voltage  $V_v$  (point *B*). The region between the peak current  $I_p$  and the valley current  $I_v$  is called *negative resistance region*. When the forward voltage is further increased beyond the valley point, the current increases again as in ordinary *p-n* junction diode.

**Explanation.** Forward current in a tunnel diode is the sum of two components due to two different mechanisms:

(i) **Normal injection current :** It arises from an external voltage which reduces the potential barrier across the depletion region and allows current to flow due to majority carriers in the conduction band. It is shown by *dashed curve*.

(ii) **Tunnel current :** In a tunnel diode due to heavy doping, the depletion region is very narrow. It results in a very high electric field across the junction and allows carriers in the valence energy band on one side of the junction to ‘tunnel’ through to the conduction energy band on the other side of the junction without overcoming the potential barrier. The increase in current from 0 to its peak value  $I_p$  (at *A*) at a forward voltage  $V_p$  is due to tunneling phenomenon. The normal injection current is negligible at this value of forward bias.

The application of the *forward bias beyond  $V_p$*  reduces the electric field strength because the depletion region becomes less well defined due to diffusion of carriers across the *p-n* junction.

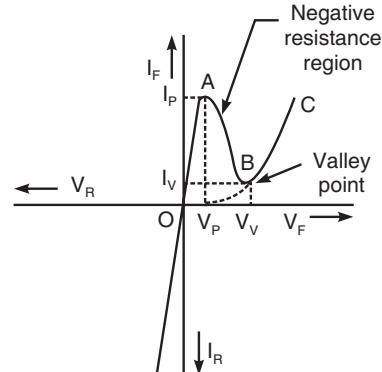


Fig. 34.29

Consequently the *tunneling current decreases*. Thus *current decreases with increase in applied voltage* between the peak point A and valley point B. It causes a negative slope in the  $V$ - $I$  curve and the tunnel diode exhibits a negative resistance property.

If forward voltage is increased beyond  $V_v$ , the tunneling effect ceases and current rises because of injection currents as in any ordinary junction diode.

**Applications.** The greatest advantage of the tunnel diode is its very fast response (change in current) to the voltage-change in the region OAB. In ordinary diodes and transistors, the response time depends on the diffusion speed of the charge-carriers which is low. Hence these devices cannot be used at high frequencies. Tunnel diodes, on the other hand, can be used as oscillators at frequencies above  $10^{11}$  Hz, and in switching circuits operating in time-intervals  $\approx 10^{-9}$  sec.

### 34.7.1. Tunnel Diode Operation

(a) A *p-n* junction between two degenerate semiconductors is illustrated in terms of energy bands in Fig. 34.30 (a).

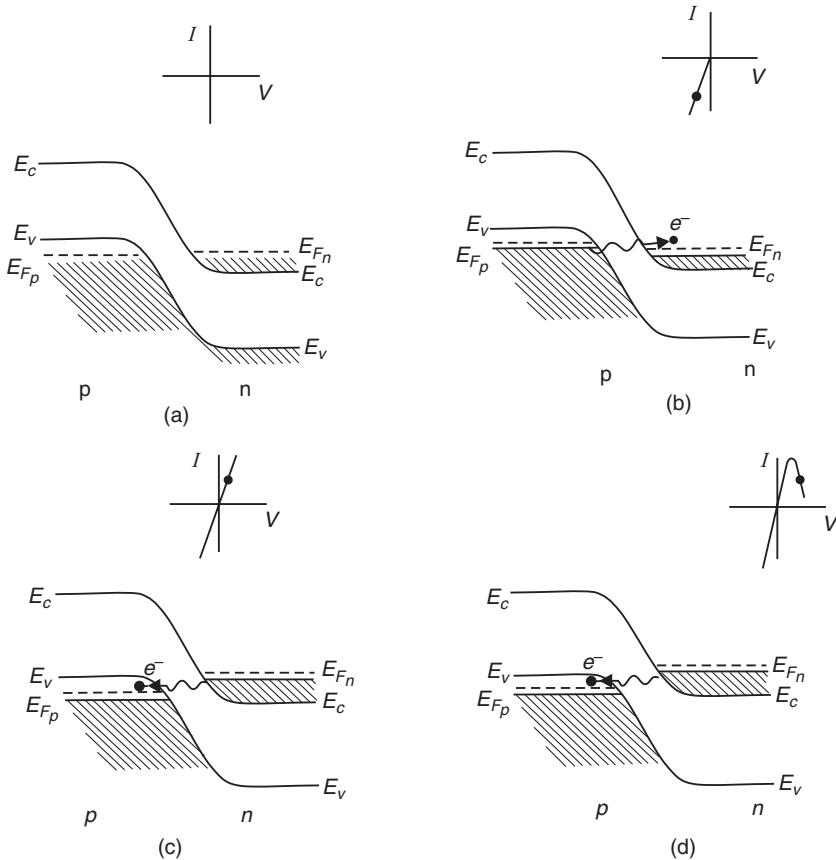


Fig. 34.30

The tunnel diode energy band diagrams: (a) zero-bias condition (no tunneling), (b) small reverse bias (electron tunneling from *p* to *n*), (c) small forward bias (electron tunneling from *n* to *p*), (d) increased forward biased ( $e^-$  tunneling from *n* to *p* decreases as bands pass by each other).

In this equilibrium condition, the Fermi level is constant throughout the junction.  $E_{Fp}$  lies below the valence band edge on the *p* side.  $E_{Fn}$  is above the conduction band edge on the *n* side. Thus the bands must overlap on the energy scale in order for  $E_F$  to be constant.

(b) Suppose, now a small reverse bias is applied [Fig. 34.30 (b)]. It allows electron tunneling from the filled valence band states below  $E_{Fp}$  to the empty conduction band states above  $E_{Fn}$ . As the reverse bias is increased,  $E_{Fn}$  continues to move down the energy scale with respect to  $E_{Fp}$ , placing more filled states on the  $p$  side opposite empty states on the  $n$  side. Thus the tunneling of electrons from  $p$  to  $n$  increases with increasing reverse bias. The resulting conventional current is opposite to the electron flow, that is, from  $n$  to  $p$ .

(c) When a small forward bias  $V$  is applied,  $E_{Fn}$  moves up in energy with respect to  $E_{Fp}$  by the amount  $eV$ . Thus electrons below  $E_{Fn}$  on the  $n$  side are placed opposite empty states above  $E_{Fp}$  on the  $p$  side [Fig. 34.30 (c)]. Electron tunneling occurs from  $n$  to  $p$  with the resulting conventional current from  $p$  to  $n$ . This forward tunneling current continues to increase with increased bias as more filled states are placed opposite empty states.

(d) However, as  $E_{Fn}$  continues to move up with respect to  $E_{Fp}$ , a point is reached at which the bands begin to pass by each other. When this occurs, the number of filled states opposite empty states decreases. The resulting decrease in tunneling current is illustrated in Fig. 34.30 (d). This region of the  $I-V$  characteristic corresponds to the *decrease* of tunneling current with *increased* bias, i.e., it produces a region of negative slope. The *dynamic resistance*  $dV/dI$  is negative.

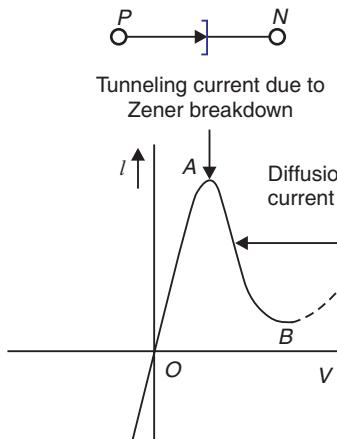


Fig. 34.31

Tunnel diode symbol  
and characteristic

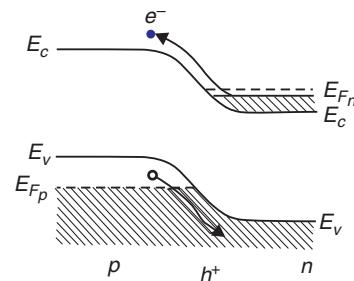


Fig. 34.32

Energy band diagram for the diode beyond  
the tunneling condition (corresponding to  
part BC of the curve in Fig. 34.31)

If the forward bias is increased beyond the negative resistance region, the current begins to increase again (Fig. 34.31). Once the bands have passed each other, the characteristic resembles that of a conventional diode. The forward current is now dominated by the diffusion current. Fig. 34.32 shows the band diagram for the tunnel diode beyond the tunnel current region.

### 34.7.2. Tunnel Diode Oscillator

**Damped oscillations.** Fig. 34.33 (i) shows a parallel resonant circuit.

The resistance  $R_p$  is the parallel equivalent of the series winding resistance of the coil. When the tank circuit is set into oscillations by applying voltage [Fig. 34.33 (ii)], damped oscillations are produced. It is because energy is lost in the resistance  $R_p$  of the tank circuit.

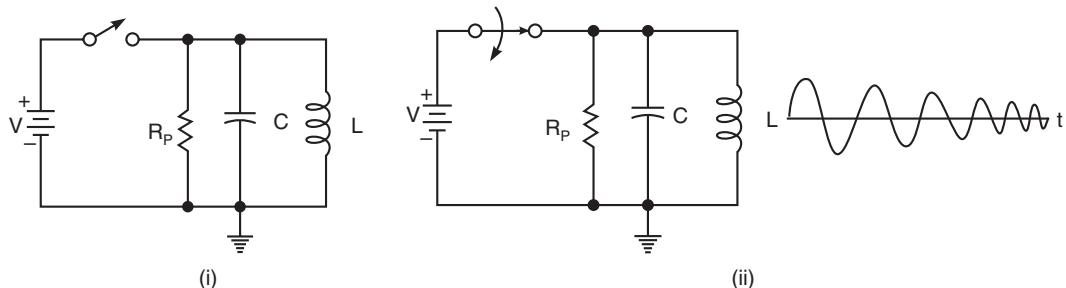


Fig. 34.33

**Undamped oscillations.** A *negative-resistance oscillator* is constructed using a tunnel diode (Fig. 34.34). It is called tunnel diode oscillator.

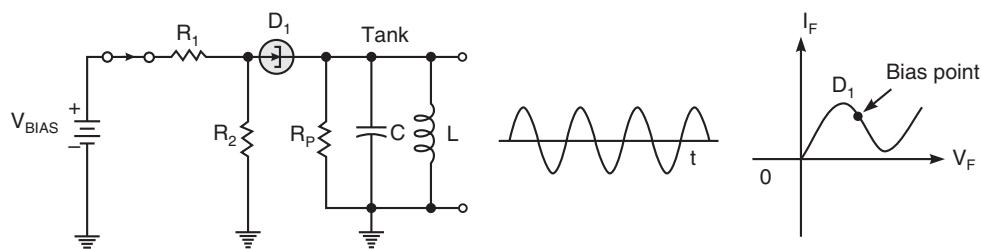


Fig. 34.34

A tunnel diode is placed in series with the tank circuit and biased at the centre of the negative-resistance portion of its characteristic. The negative-resistance characteristic of the tunnel diode counteracts the positive-resistance characteristic of the tank circuit. So *undamped* oscillations are produced at the output.

**Disadvantage.** The circuit works very well at extreme high frequencies (upper mega hertz range). But it cannot be used efficiently at low frequencies.

### 34.8 VARACTOR DIODE

- A junction diode which acts as a variable capacitor under changing reverse bias is called varactor diode.
- It is widely used in television receivers, FM receivers, and other communications equipment because it can be used for electronic tuning.

**Principle.** When a *pn* junction is formed, depletion layer is created in the junction area (Fig. 34.35a).

- The depletion layer is between the *p* region and the *n* region. The *p* and *n* regions are like the plates of a capacitor.
- The depletion layer is like the dielectric.
- When a diode is reverse-biased, the width of the depletion layer increases with the reverse voltage. Since the depletion layer gets wider with more reverse voltage, the capacitance becomes smaller. Capacitance is controlled by reverse voltage.
- A semiconductor diode that is designed to maximise the capacitance variation with applied reverse bias is the varactor.

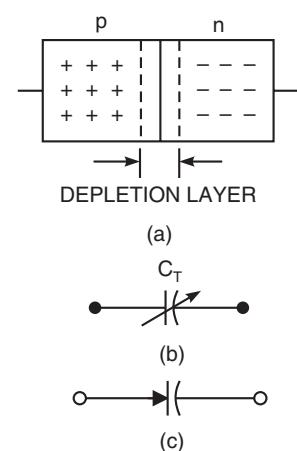


Fig. 34.35

- Fig. 34.35 *b* shows the ac equivalent circuit for a reverse-biased diode. As far as an ac signal is concerned, the varactor acts the same as a variable capacitance.
- Fig. 34.35 *c* shows the schematic symbol for a varactor.
- Fig. 34.36 shows how the capacitance varies with reverse voltage. This graph shows that the capacitance gets smaller when the reverse voltage gets larger.
- Varactor diodes are made with gallium arsenide and optimized for use at microwave frequencies. Variation in capacitance can be utilised to change resonant frequency of a tuned circuit. This is the principle behind electronic tuning.

**Theory.** Capacitance of a varactor diode is given by:

$$C_T = \epsilon \frac{A}{W_d} .$$

Here,  $C_T$  = Total capacitance of the junction,

$\epsilon$  = Permittivity of the semiconductor material,

$A$  = Cross-sectional area of the junction,

$W_d$  = Width of the depletion layer.

- When reverse voltage across a varactor diode is increased, the width  $W_d$  of the depletion layer increases. Therefore, the total junction capacitance  $C_T$  of the junction decreases. The larger the reverse voltage, the smaller the capacitance  $C_T$ .
- When the reverse voltage across the diode is decreased, the width  $W_d$  of the depletion layer decreases. Therefore, the total junction capacitance  $C_T$  increases. Fig. 34.37 shows the graph of total junction capacitance  $C_T$  versus voltage  $V_R$ .

### Applications

- Because the capacitance is voltage-controlled, varactors have replaced mechanically tuned capacitors in many applications such as television receivers and automobile radios.
- The tuning range of a varactor depends on the doping level.

**Abrupt junction.** Fig. 34.38 shows the doping profile for an abrupt-junction diode. The doping is uniform on both sides of the junction.

The tuning range of an abrupt junction diode is between 3 : 1 and 4 : 1.

**Hyperabrupt junction.** To get larger tuning ranges, some varactors have a *hyperabrupt junction*.

Fig. 34.39 shows the doping profile.

- The doping level increases as we approach the junction. The heavier doping produces a narrower depletion layer and a larger capacitance. Furthermore, changes in reverse voltage have more pronounced effects on capacitance. A hyperabrupt varactor has a tuning range of about 10 : 1, enough to tune an AM radio through its frequency range of 535 to 1605 kHz.

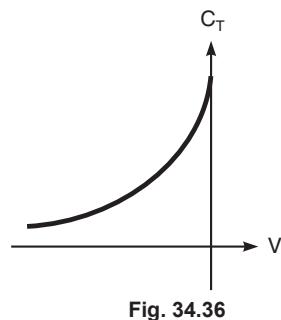


Fig. 34.36

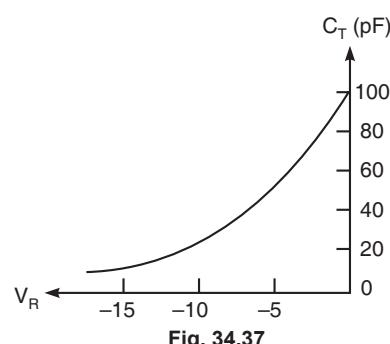


Fig. 34.37

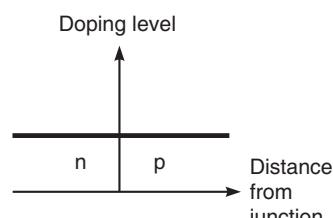


Fig. 34.38

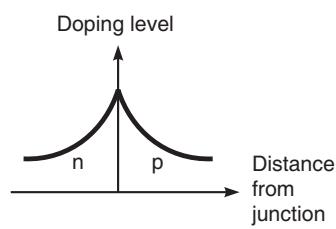


Fig. 34.39

Fig. 34.40 shows the use of varactor diode in a tuned circuit.

- The capacitance of the varactor is in *parallel* with the inductor. The varactor and the inductor form a parallel *LC* circuit.

The resonant frequency  $f_r$  of the *LC* circuit is given by

$$f_r = \frac{1}{2\pi\sqrt{LC}}.$$

**Working.** The resistance  $R_W$  in the circuit is the *winding resistance* of the inductor. This winding resistance is in series with the potentiometer  $R_1$ . Thus  $R_1$  and  $R_W$  form a voltage divider that is used to determine the amount of reverse bias across the varactor diode  $D_1$  and therefore its capacitance. By adjusting the setting of  $R_1$ , we can vary the diode capacitance. This, in turn, varies the resonant frequency of the *LC* circuit.

- Some of the high-frequency (as defined by the small capacitance levels) areas of application include FM modulators, automatic-frequency-control devices, adjustable bandpass filters, and parametric amplifiers.

### 34.9 SHOCKLEY DIODE

**Construction.** Fig. 34.41 shows the basic construction of a Shockley diode and its schematic symbol.

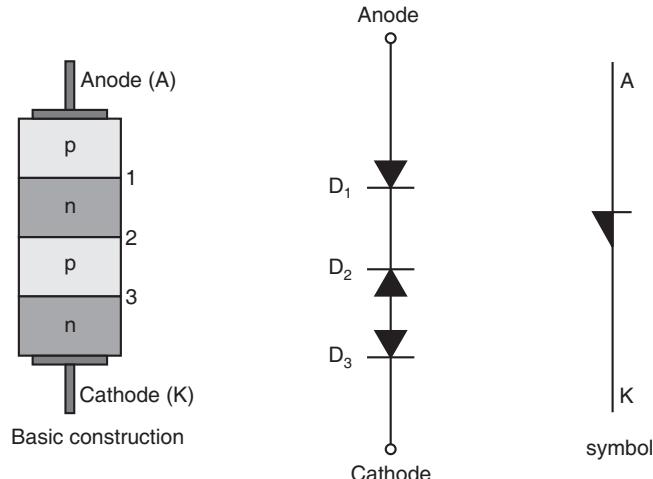


Fig. 34.41

- It is a *pnpn* device having two terminals.
- This device acts as a switch and consists of four alternate *p*-type and *n*-type layers in a single crystal.
- Since a *p*-region adjacent to an *n*-region may be considered a junction diode, the Shockley diode is equivalent to three junction diodes connected in series.

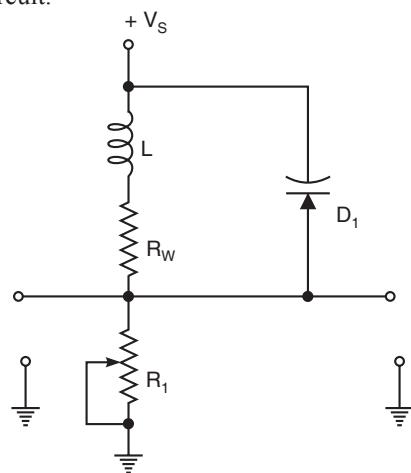


Fig. 34.40

**Equivalent circuit.** The *pnpn* structure can be represented by an equivalent circuit consisting of a *pnp* transistor and an *npn* transistor [Fig. 34.42(a)]. The upper *pnp* layers form  $Q_1$  and the lower *npn* layers form  $Q_2$ , with the two middle layers shared by both equivalent transistors.

- The base-emitter junction of  $Q_1$  corresponds to *pn* junction 1.
- The base-emitter junction of  $Q_2$  corresponds to *pn* junction 3.
- The base-collector junctions of both  $Q_1$  and  $Q_2$  corresponds to *pn* junction 2.

When a positive bias voltage is applied to the anode with respect to the cathode [Fig. 34.42(b)], the base-emitter junctions of  $Q_1$  and  $Q_2$  are forward-biased, and the common base-collector junction is reverse-biased. Therefore, both equivalent transistors are in the linear region.

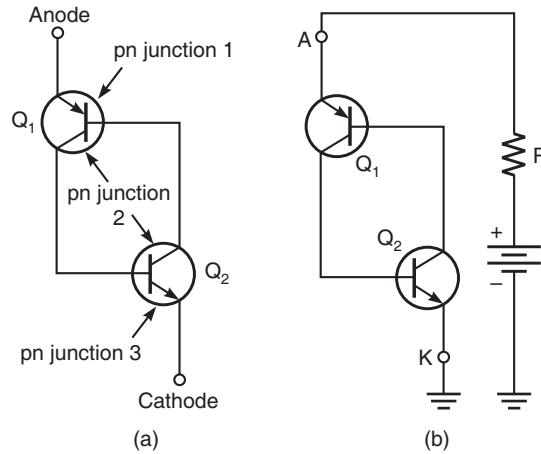


Fig. 34.42

**V-I characteristics of a Shockley diode.** Fig. 34.43 shows the Shockley diode characteristic curve.

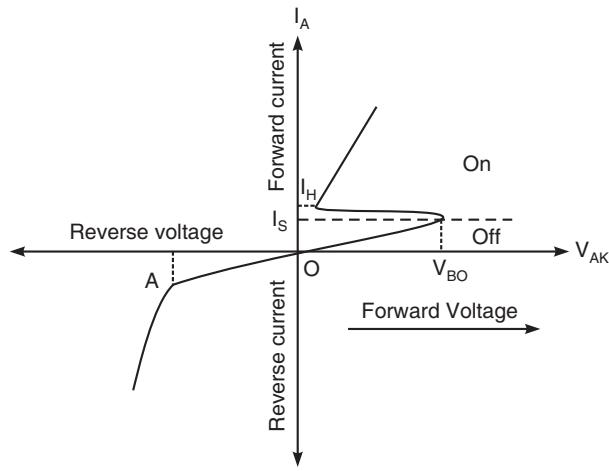


Fig. 34.43

(1) **Forward bias characteristic.** When Shockley diode is forward biased, diodes  $D_1$  and  $D_3$  are forward-biased. But diode  $D_2$  is reverse-biased. Since diode  $D_2$  offers very high resistance and the three diodes are in series, the Shockley diode presents a very high resistance.

As the forward voltage increases, the reverse bias across  $D_2$  also increases

At some forward voltage (called *breakover voltage*  $V_{BO}$ ), reverse breakdown of  $D_2$  occurs. Since this breakdown results in reduced resistance, the Shockley diode presents a very low resistance.

- From now onwards, the Shockley diode behaves as a conventional forward-biased diode. The forward current is determined by the applied voltage and external load resistance.

**Shockley diode behaves like a switch.** There is a region of forward bias, called the *forward-blocking region*, in which the device has a very high forward resistance (ideally an open) and is in the *off* state. The forward-blocking region exists from  $V_{AK} = 0$  V up to a  $V_{AK}$  called the **forward breakover voltage**  $V_{BO}$ .

So long as the forward voltage is less than breakover voltage, Shockley diode offers very high resistance (*i.e.*, switch is open) and practically conducts no current. At voltages above the break-over value, Shockley diode presents a very low resistance (*i.e.*, switch is closed) and Shockley diode conducts heavily.

- When  $V_{AK} = V_{BO}$ ,  $I_A = I_S$ , the device is in the *on* state and acts as a closed switch.
- When the anode current drops back below the holding value,  $I_H$ , the device turns *off*.

**Holding current.** Once the Shockley diode is conducting (in the *on* state), it will continue to conduct until the anode current is reduced below a specified level, called the **holding current**,  $I_H$ . When  $I_A$  falls below  $I_H$ , the device rapidly switches back to the *off* state and enters the forward-blocking region.

**Switching current.** The value of the anode current at the point where the device switches from the forward-blocking region (*off*) to the forward-conduction region (*on*) is called the *switching current*  $I_s$ . This value of current is always less than the holding current,  $I_H$ .

(2) **Reverse bias characteristic.** When Shockley diode is reverse biased, diodes  $D_1$  and  $D_3$  are reverse-biased. But diode  $D_2$  is forward-biased. If reverse voltage is increased sufficiently, the reverse voltage breakdown (point *A* in Fig. 34.43) of Shockley diode is reached. At this point, diodes  $D_1$  and  $D_3$  would go into reverse-voltage breakdown, the reverse current flowing through them would rise rapidly and the heat produced by this current flow could ruin the entire device. For this reason, *Shockley diode should never be operated with a reverse voltage sufficient to reach the reverse voltage breakdown point*.

### 34.10 PIN DIODE

- Fig. 34.44 (a) shows the construction of a PIN diode. Fig. 34.44 (b) shows the schematic symbol.

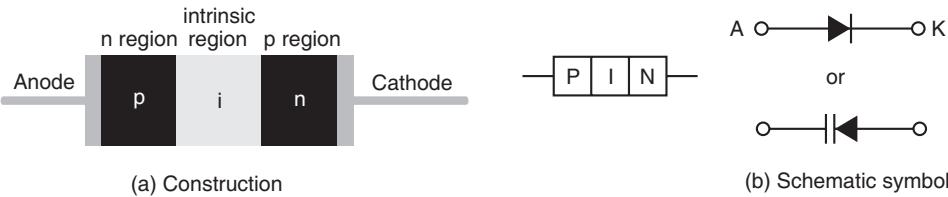


Fig. 34.44

It consists of heavily doped *p* and *n* regions separated by an intrinsic (*i*) region.

- When forward-biased, it acts like a current-controlled variable resistance [Fig. 34.45(a)].

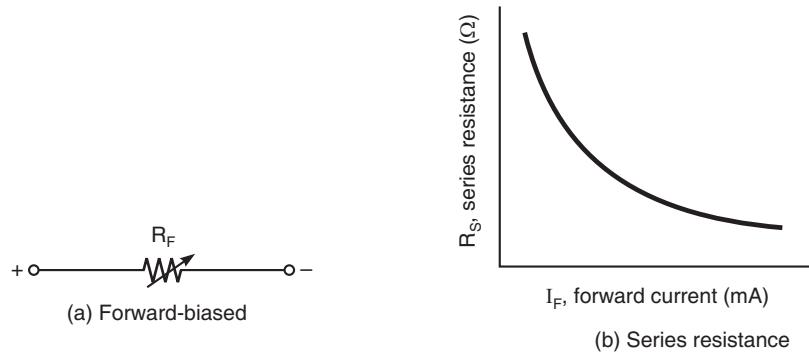


Fig. 34.45

The low forward resistance of the intrinsic region decreases with increasing current.

Fig. 34.45 (b) shows how the PIN diode's series resistance  $R_s$  decreases as its forward current increases.

**Application.** The *pin* diode operates as a variable resistor at RF and microwave frequencies. The PIN diode is widely used in RF and microwave modulator circuits.

- When reverse biased, the PIN diode acts like a fixed capacitor (Fig. 34.46).

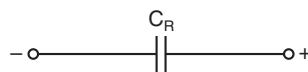


Fig. 34.46. Reverse biased

### EXERCISE

1. Draw characteristics of a Zener diode. Explain with the help of a circuit diagram its use as voltage regulator. *(Rohilkhand 90)*
2. Explain with proper circuit diagram the construction, working and uses of a 'Light Emitting Diode' *(Rohilkhand 90)*
3. Give the structure, symbol, volt-ampere characteristics and use of a photo diode.
4. Describe the action of a LCD. How does it compare with LED?
5. What is solar cell? Discuss the construction, working, I-V characteristics and uses of a solar cell.
6. Describe the construction of a tunnel diode. Also give the theory of tunnel diode.
7. What is a tunnel diode? Explain the V-I characteristics of a tunnel diode. Explain the working of tunnel diode oscillator.
8. What is a varactor diode? List the features of varactor diode.
9. Explain the working of varactor diode. Give one application of varactor diode.
10. Explain the working of Shockley diode. Sketch and discuss the VI characteristics of Shockley diode.
11. Explain the working of a PIN diode.

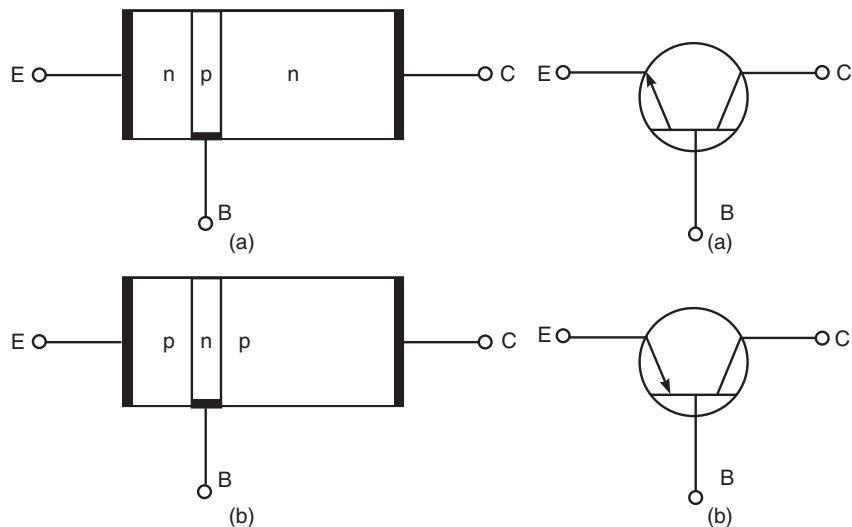
## BIPOLAR JUNCTION TRANSISTOR (BJT)

**After reading this chapter, you should be able to**

- ◆ Explain the physical behaviour of a BJT
- ◆ Discuss the relationships among the base, emitter, and collector currents of a BJT
- ◆ Explain the common base, common emitter, and common collector characteristics of a BJT
- ◆ Explain practical BJT bias circuit design and analysis
- ◆ Analyze dc and ac load line operation
- ◆ Explain the  $h$  parameters

### 35.1 JUNCTION TRANSISTOR

The junction transistors are of two types : *p-n-p* transistor and *n-p-n* transistor. Fig. 35.1 (a) shows an *n-p-n* transistor and its symbol. Fig. 35.1 (b) shows a *p-n-p* transistor and its symbol. We consider only the *n-p-n* transistor. A transistor has the following sections :



**Fig. 35.1**

(i) **Emitter.** This forms the left hand section or region of the transistor. The main function of this region is to supply majority charge carriers (either electrons or holes) to the base. Hence it is more heavily doped in comparison to other regions. Electrons are the charge-carriers within the *n-p-n* transistor, whereas holes are the charge-carriers within a ***p-n-p* transistor**.

(ii) **Base.** The middle section of the transistor is called base. This is very lightly doped and is very thin ( $10^{-6}$  m) as compared to either emitter or collector so that it may pass most of the injected charge carriers to the collector.

(iii) **Collector.** The right hand side of the transistor is called collector. The main function of the collector is to collect majority charge carriers through the base. This is moderately doped.

The junction between the emitter and base is called emitter junction. The junction between the collector and the base is called collector junction. In normal operation of a transistor, the emitter-base junction is forward biased while the collector-base junction is reverse biased.

**Working of an *n-p-n* Transistor.** The emitter-base junction is forward biased by the battery  $V_{EB}$  (Fig. 35.2). The collector-base junction is reverse biased by the battery  $V_{CB}$ . The directions of the emitter, base and collector currents are as shown in the figure. The direction of each current is opposite to the direction of motion of the electrons. The electrons being majority carriers in the emitter are repelled due to forward bias towards the base. The base contains holes as majority carriers and some holes and electrons combine in the base region. Since the base is lightly doped, the probability of electron-hole combination in base region is very small (5%). The remaining electrons cross into collector region, and enter into the positive terminal of the battery  $V_{CB}$  connected to the collector. At the same time an electron enters the emitter from the negative pole of the emitter-base battery  $V_{EB}$ . Thus, in *n-p-n* transistors, the current is carried inside the transistor as well as in the external circuit by the electrons. If  $I_E$ ,  $I_B$  and  $I_C$  are respectively the emitter current, the base current and the collector current, then  $I_E = I_B + I_C$ .

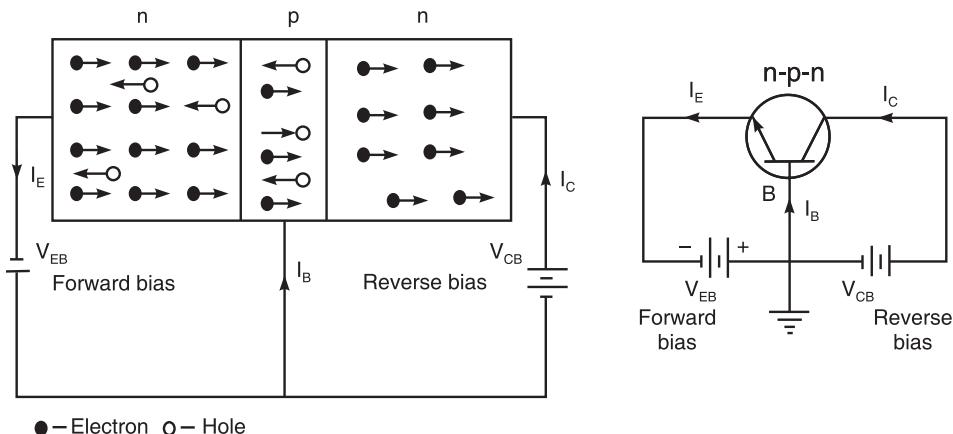


Fig. 35.2

### Transistor currents

The direction of the currents in an *n**p**n* transistor and its schematic symbol are as shown in Fig. 35.3.

- The emitter current ( $I_E$ ) is the sum of the collector current ( $I_C$ ) and the base current ( $I_B$ ), expressed as follows:

$$I_E = I_C + I_B \quad \dots (1)$$

Since the base current is so small, the collector current approximately equals the emitter current:

$$I_C \approx I_E$$

$I_B$  is very small compared to  $I_E$  or  $I_C$ .

The capital-letter subscripts indicate dc values.

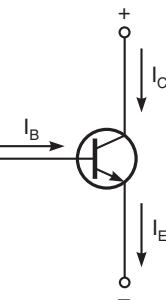


Fig. 35.3

### DC Alpha ( $\alpha_{dc}$ ) and DC Beta ( $\beta_{dc}$ )

**DC Alpha** ( $\alpha_{dc}$ ) is defined as the dc collector current divided by the dc emitter current:

$$\alpha_{dc} = \frac{I_C}{I_E} \quad \dots (2)$$

Since the collector current almost equals the emitter current, the dc alpha is slightly less than 1.

The **dc beta** ( $\beta_{dc}$ ) of a transistor is defined as the ratio of the dc collector current to the dc base current:

$$\beta_{dc} = \frac{I_C}{I_B} \quad \dots (3)$$

The dc beta is also called the **current gain** because a small base current controls a much larger collector current.

The current gain is a major advantage of a transistor and has led to all kinds of applications. For low power transistors (under 1 W), the current gain is typically 100 to 300. High-power transistors (over 1 W) usually have current gains of 20 to 100.

### Relation between $\alpha_{dc}$ and $\beta_{dc}$

We know that  $I_E = I_C + I_B$   $\dots (1)$

$$\text{Dividing Eq. (1) by } I_C, \quad \frac{I_E}{I_C} = 1 + \frac{I_B}{I_C}$$

$$\frac{1}{\alpha} = 1 + \frac{1}{\beta} = \frac{\beta + 1}{\beta} \quad \left( \because \alpha = \frac{I_C}{I_E} \text{ and } \beta = \frac{I_C}{I_B} \right)$$

or

$$\alpha = \frac{\beta}{\beta + 1}$$

... (2)

$$\begin{aligned}\alpha(\beta + 1) &= \beta \\ \alpha\beta + \alpha &= \beta \\ \alpha &= \beta - \alpha\beta = \beta(1 - \alpha)\end{aligned}$$

$$\beta = \frac{\alpha}{1 - \alpha}$$

... (3)

## 35.2 COMMON Emitter CHARACTERISTICS OF A TRANSISTOR

Fig. 35.4 shows the circuit for drawing the common emitter characteristics of *n-p-n* transistor. The base-emitter junction is forward-biased by using base-emitter battery  $V_{BB}$ . The base-emitter voltage  $V_{BE}$  is read by a voltmeter and the base current  $I_B$  by a microammeter ( $\mu A$ ). The base-collector junction is reverse biased by using collector-emitter battery  $V_{CC}$ . The collector-to-emitter voltage  $V_{CE}$  is read by another voltmeter and the collector current  $I_C$  by a milliammeter (mA).

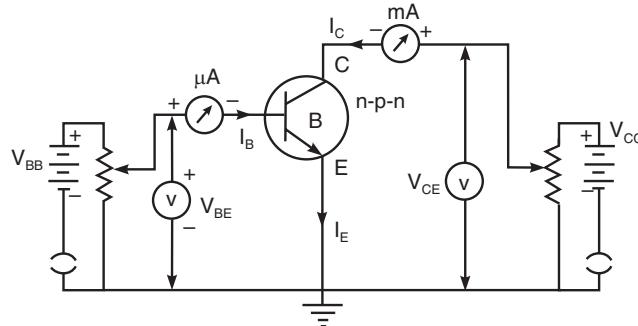


Fig. 35.4

**1. Input characteristics.** The curves showing the variation of the base current  $I_B$  with base-emitter voltage  $V_{BE}$  at constant collector-emitter voltage  $V_{CE}$  are called the input characteristics of a transistor.

To draw the input characteristics for *n-p-n* transistor, the collector-emitter voltage  $V_{CE}$  is maintained at a constant value, say 5V. The base currents  $I_B$  are noted for various values of input potential  $V_{BE}$ . A graph is drawn between  $V_{BE}$  taken along the X-axis and  $I_B$  taken along Y-axis. The experiment is repeated for different values of  $V_{CE}$ . Typical input characteristics are shown in Fig. 35.5 for  $V_{CE} = 5$  and 20 volts. The shape of the graph is the same as that of the forward characteristics of *p-n* junction diode.

$$\text{Input resistance, } r_i = \left( \frac{\Delta V_{BE}}{\Delta I_B} \right)_{V_{CE}}$$

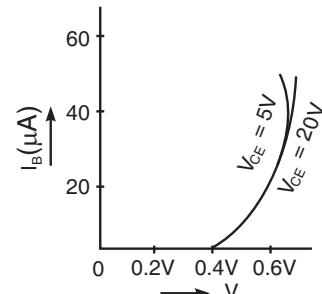


Fig. 35.5

**2. Output characteristics.** Curves showing the variation of the collector current ( $I_C$ ) with the collector-emitter voltage  $V_{CE}$  at constant base current  $I_B$  are called the output characteristics of a transistor.

The collector current  $I_C$  is measured by varying the collector voltage  $V_{CE}$  for a constant value of base current  $I_B$ . The experiment is repeated for different values of  $I_B$  and a family of curves is obtained as shown in Fig. 35.6.

A study of these curves reveals the following facts :

(i) The collector current  $I_C$  varies rapidly with  $V_{CE}$  for very small voltages (say up to  $V_{CE} = 2$  volts). After this collector current becomes almost constant and is decided entirely by base current  $I_B$ . It then becomes independent of  $V_{CE}$ .

(ii) The collector current  $I_C$  is not zero when base current is zero. This is due to the inherent intrinsic conduction in semiconductors.

(iii) **Current gain  $\beta$**  : This is defined as the ratio of change in collector current to the change in base current for constant value of collector voltage  $V_{CE}$  in common emitter configuration, i.e.,

$$\beta = \left( \frac{\Delta I_C}{\Delta I_B} \right)_{V_{CE} = \text{constant}} . \quad \beta \text{ ranges from 20 to 200.}$$

Since the input current  $I_B$  is measured in microampere and output current  $I_C$  is measured in milliampere, the common emitter configuration exhibits a current amplification  $\beta$ . Thus small input current  $I_B$  produces a large output current  $I_C$  or in other words, the collector current is  $\beta$  times greater than the base current.

$$(iv) \text{ Output resistance } r_o = \left( \frac{\delta V_{CE}}{\delta I_C} \right)_{I_B = \text{constant}} .$$

### 35.3 CHARACTERISTICS OF A TRANSISTOR CONNECTED IN COMMON BASE CONFIGURATION

Fig. 35.7 shows the circuit arrangement to draw the characteristics of an NPN transistor connected in Common Base Configuration.

$V_{EB}$  can be changed by varying the position of potentiometer  $R_1$ .  $V_{CB}$  could be varied by adjusting  $R_2$ .

(i) **Input characteristics** : The curve between emitter current  $I_E$  and emitter base voltage  $V_{EB}$  at constant collector base voltage  $V_{CB}$  represents the input characteristic. For plotting the input characteristic, the collector base voltage  $V_{CB}$  is kept fixed. The emitter base voltage  $V_{EB}$  is varied with the help of potential divider  $R_1$  and the emitter current is noted for each value of  $V_{EB}$ . A graph of  $I_E$  against  $V_{EB}$  is drawn. The curve is known as input characteristic. The experiment is repeated for other fixed values of  $V_{CB}$ . The input characteristic is shown in Fig. 35.8.

The following points are noted from the characteristic:

(a) There exists a threshold voltage  $V_{EB}$  (also called the *knee voltage* or *cut-in voltage*), represented by  $V_K$ , which is equal to  $OA$ , below which  $I_E$  is negligibly small.  $V_K \approx 0.5$  V for Si and 0.1 V for Ge transistors.

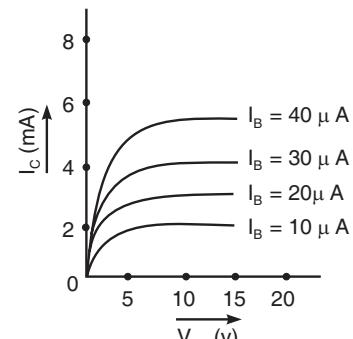


Fig. 35.6

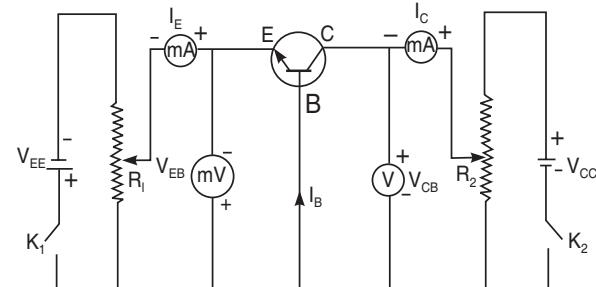


Fig. 35.7

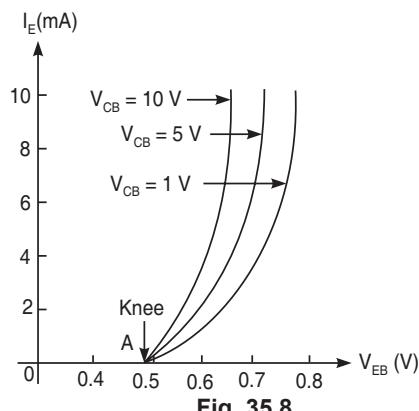


Fig. 35.8

(b) For  $V_{EB} > V_K$ , for a constant  $V_{CB}$ ,  $I_E$  increases sharply with a small increase in  $V_{EB}$ . This shows that the input resistance is very small.

(c) The *dynamic* or *a.c. input resistance* is experimentally obtained from the formula

$$R_i = \left( \frac{\Delta V_{EB}}{\Delta I_E} \right)_{V_{CB}=\text{const.}}$$

$R_i$  may be typically of the order of  $100 \Omega$  in the linear region.

**(ii) Output characteristics.** The curve between collector current  $I_C$  and collector-base voltage  $V_{CB}$  at constant emitter current  $I_E$  represents the output characteristic.

The emitter current  $I_E$  is kept fixed. With the help of potential divider  $R_2$ , the value of  $V_{CB}$  is varied in steps and the collector current  $I_C$  is noted for each value of  $V_{CB}$ . Now a graph is drawn between  $I_C$  and  $V_{CB}$ . The curve so obtained is known as output characteristic. The experiment is repeated for different fixed values of emitter current  $I_E$ . Fig. 35.9 shows the output characteristic.

The output set of characteristics has three regions of interest : the *active*, *cutoff*, and *saturation* regions.

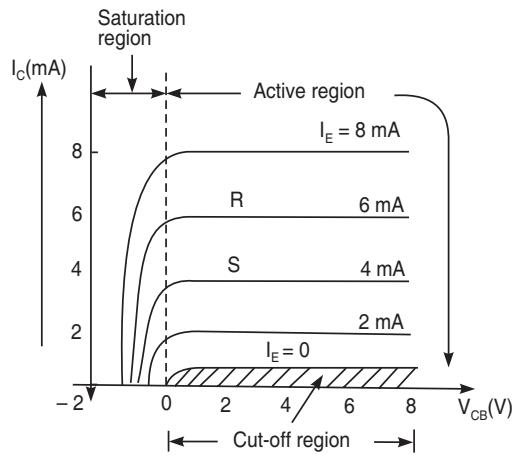


Fig. 35.9

**(i) Active region :** It is the region normally employed for linear (*undistorted*) amplifiers.

In the *active region*, the *base-emitter junction is forward-biased, while the collector-base junction is reverse-biased*.

1. In the active region,  $I_c = \text{constant} \approx I_E$

2.  $I_C$  is practically independent of  $V_{CB}$  in the active region. However, if  $V_{CB}$  is increased beyond a certain large value,  $I_C$  increases rapidly due to *avalanche breakdown* (not shown in the figure) and the transistor action stops.

**(ii) Cut-off region :** In the *cut-off region* (when emitter and collector junctions are both reverse-biased), a small amount of collector current flows even when emitter current  $I_E = 0$ . This is the collector leakage current  $I_{CBO}$ .

**(iii) Saturation region.** The region to the left of ordinate  $V_{CB} = 0$  is called saturation region.

1. Note the exponential increase in  $I_c$  as the voltage  $V_{CB}$  increases towards zero volt.

2. Here, both the  $B-E$  and  $C-B$  junctions are forward-biased.

**Dynamic output resistance ( $R_0$ ) :** It is also called the *a.c. output resistance*.

It is experimentally found from the definition

$$R_0 = \left( \frac{\Delta V_{CB}}{\Delta I_C} \right)_{I_E=\text{const.}}$$

Since the characteristic curves are almost horizontal,  $R_0$  is very high. Its typical value is  $500 \text{ k}\Omega$ .

**Current gain.** DC current gain and *a.c. current gain* are respectively defined by

$$\alpha_{dc} = \left( \frac{I_C}{I_E} \right)_{V_{CB}=\text{const.}} \quad \text{and} \quad \alpha_{ac} = \left( \frac{\Delta I_C}{\Delta I_E} \right)_{V_{CB}=\text{const.}}$$

Both  $\alpha$ 's have a value slightly less than unity.

### 35.4 C – C CONFIGURATION

The *common-collector configuration* is shown in Fig. 35.10 with the proper current directions and voltage notation.

The common-collector configuration is used primarily for impedance-matching purposes since it has a high input impedance and low output impedance.

#### Common Collector Characteristics

Fig. 35.11 shows the circuit arrangement to draw the characteristics of a PNP transistor connected in C – C configuration.

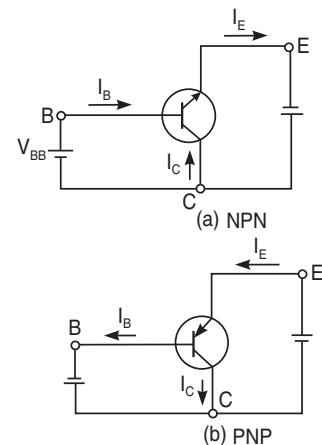


Fig. 35.10

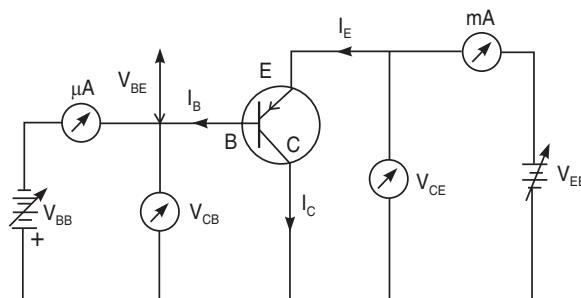


Fig. 35.11

**1. Input characteristics :** The curve showing the variation of base current  $I_B$  with collector-base voltage  $V_{CB}$  at constant collector-emitter voltage  $V_{CE}$  is called the input characteristic.

Fig. 35.12 gives a plot of  $I_B$  versus  $V_{CB}$  for different values of  $V_{CE}$ . These characteristics are quite different from those for CB or CE circuit because the input voltage  $V_{CB}$  is largely determined by the value of  $CE$  voltage. It may be seen that with an increase in  $V_{CB}$ ,  $I_B$  gets decreased.

**2. Output Characteristics :** The curves showing the variation of the emitter current  $I_E$  with collector-emitter voltage  $V_{CE}$  at constant base current  $I_B$  are called the output characteristics of the transistor.

Fig. 35.13 gives the plot of  $I_E$  versus  $V_{CE}$  for several fixed values of  $I_B$ .

For all practical purposes, the output characteristics of the common-collector configuration are the same as for the common-emitter configuration because  $I_C \approx I_E$ .

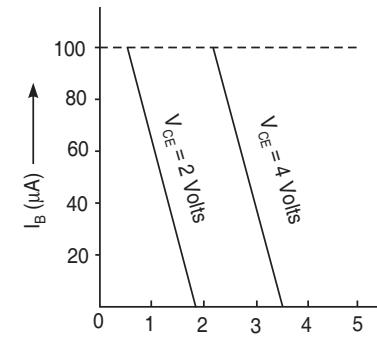


Fig. 35.12

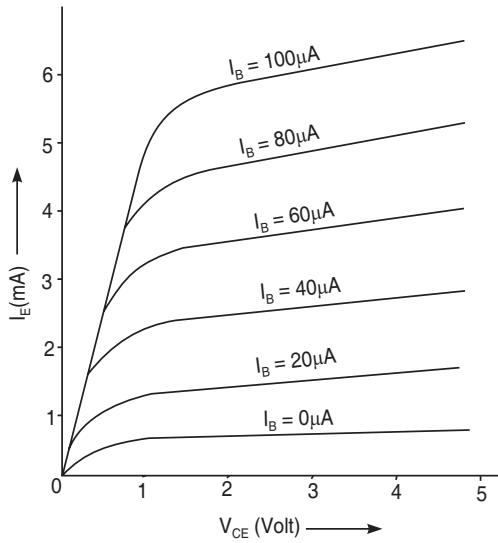


Fig. 35.13

### 35.5 TRANSISTOR CIRCUIT CONFIGURATIONS

- A transistor can be connected in the following three different configurations:

(1) **Common-base configuration.** Fig. 35.14 shows the *common-base configuration*:

(i) **Definition of static current gain,  $\alpha_{dc}$ .** It is the ratio of static (dc) collector current  $I_C$  to the static emitter current  $I_E$  at a constant collector voltage with respect to the base,

$$\alpha_{dc} = \left( \frac{I_C}{I_E} \right)_{V_{CB} = \text{constant}} \quad \dots (1)$$

(ii) **Definition of small signal current gain ( $\alpha_{ac}$ ).** It is the ratio of small change in collector current to the small change in the emitter current at a constant collector voltage with respect to the base.

$$\alpha_{ac} = \left( \frac{\partial I_C}{\partial I_E} \right)_{V_{CB} = \text{constant}} \quad \dots (2)$$

- Since  $\alpha_{dc}$  and  $\alpha_{ac}$  do not differ from each other much, it is simply denoted as  $\alpha$ .

#### (2) Common-Emitter Configuration

Fig. 35.15 shows the *common-emitter configuration*.

- Static current gain ( $\beta_{dc}$ )** is the ratio of the static (dc) collector current  $I_C$  to the static base current  $I_B$  at a constant collector to emitter voltage.

$$\beta_{dc} = \frac{I_C}{I_B}$$

- Small signal current gain ( $\beta_{ac}$ )** is the ratio of small change in the collector current to corresponding small change in the base current at constant collector voltage with respect to the emitter.

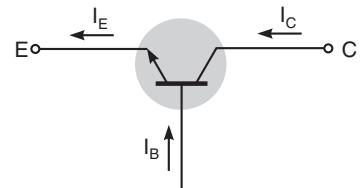
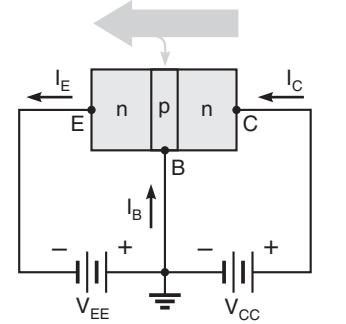


Fig. 35.14

$$\beta_{ac} = \left( \frac{\partial I_C}{\partial I_B} \right)_{V_{CE} = \text{Constant}}$$

**Relation between  $\alpha$  and  $\beta$ .** By definition,

$$\alpha = \left( \frac{\Delta I_C}{\Delta I_E} \right)_{V_{CB}} \text{ and } \beta = \left( \frac{\Delta I_C}{\Delta I_B} \right)_{V_{CE}}$$

For both the types,

$$\begin{aligned} I_E &= I_B + I_C \\ \therefore \Delta I_E &= \Delta I_B + \Delta I_C \end{aligned}$$

$$\text{Dividing by } \Delta I_C, \frac{\Delta I_E}{\Delta I_C} = \frac{\Delta I_B}{\Delta I_C} + 1$$

$$\begin{aligned} \frac{1}{\alpha} &= \frac{1}{\beta} + 1 \\ \therefore \beta &= \frac{\alpha}{1 - \alpha} \end{aligned}$$

(3) **Common collector configuration.** Fig. 35.16 shows the *common-collector configuration*.

*dc current gain*,

$$\gamma_{dc} = \left( \frac{I_E}{I_B} \right)_{V_{EC}}$$

- The input current is the base current  $I_B$ .
- The output current is the emitter current  $I_E$ .
- The ratio of change in emitter current ( $\Delta I_E$ ) to the change in base current ( $\Delta I_B$ ) is called **ac current gain**.

$$\gamma_{ac} = \left( \frac{\Delta I_E}{\Delta I_B} \right)_{V_{EC}}$$

(1) **Relation between  $\gamma$  and  $\alpha$**

$$(i) \quad \gamma_{dc} = \left( \frac{I_E}{I_B} \right) \quad \dots (1)$$

$$\text{and} \quad I_E = (I_B + I_C) \quad \dots (2)$$

$$\text{Dividing by } I_E, \text{ we get} \quad I = \frac{I_B}{I_E} + \frac{I_C}{I_E}$$

or

$$I = \frac{1}{\gamma_{dc}} + \alpha_{dc} \quad (\because \alpha_{dc} = \frac{I_C}{I_E})$$

$$\gamma_{dc} = \frac{1}{1 - \alpha_{dc}} \quad \dots (3)$$

$$(ii) \quad \gamma_{ac} = \left( \frac{\Delta I_E}{\Delta I_B} \right) \quad \dots (4)$$

$$\text{and} \quad I_E = (I_B + I_C) \quad \dots (5)$$

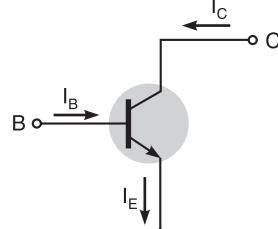
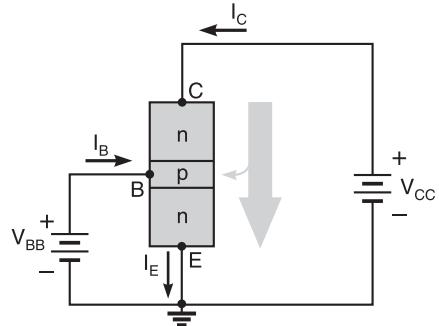


Fig. 35.15

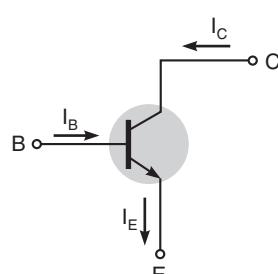
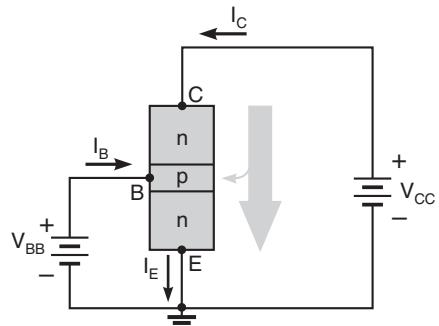


Fig. 35.16

$$\Delta I_E = \Delta I_B + \Delta I_C$$

Dividing by  $\Delta I_E$ ,

$$1 = \frac{\Delta I_B}{\Delta I_E} + \frac{\Delta I_C}{\Delta I_E}$$

or

$$1 = \frac{1}{\gamma_{ac}} + \alpha_{ac}$$

$$\therefore \gamma_{ac} = \frac{1}{1 - \alpha_{ac}} \quad \dots (6)$$

**(2) Relation between  $\gamma$  and  $\beta$** 

$$(i) \quad \gamma_{dc} = \frac{I_E}{I_B} = \frac{I_B + I_C}{I_B} = 1 + \frac{I_C}{I_B} = 1 + \beta_{dc}$$

$$(ii) \quad \gamma_{ac} = \frac{\Delta I_E}{\Delta I_B} = \frac{\Delta I_B + \Delta I_C}{\Delta I_B} = 1 + \beta_{ac}$$

$\gamma = \frac{1}{1 - \alpha}$
$\gamma = 1 + \beta$

**Table 35.1: Comparison of CB, CE, CC configurations**

No.	Characteristic	Common base CB	Common emitter CE	Common collector CC
1.	Input resistance	Low (about 100 $\Omega$ )	Low (about 750 $\Omega$ )	very high (about 750 k $\Omega$ )
2.	Output resistance	Very high (about 450 k $\Omega$ )	High (about 45 k $\Omega$ )	Low (about 50 $\Omega$ )
3.	Current gain	Slightly less than one denoted as $\alpha$	High denoted as $\beta$ $\beta = \frac{\alpha}{1 - \alpha}$	High denoted as $\gamma$ $\gamma = \frac{1}{1 - \alpha} = 1 + \beta$
4.	Voltage gain	about 150	about 500	Less than 1
5.	Applications	For high frequency applications	For audio-frequency applications	For impedance matching.

**TRANSISTOR BIASING****35.6 TRANSISTOR BIASING**

A transistor is used as an amplifier. For faithful amplification, the following three conditions must be satisfied :

- (i) The emitter-base junction should be forward-biased,
- (ii) The collector-base junction should be reverse-biased, and
- (iii) There should be proper zero signal collector current.

The proper flow of zero signal collector current (proper operating point of a transistor) and the maintenance of proper collector-emitter voltage during the passage of signal is known as transistor biasing.

When a transistor is not properly biased, it works inefficiently and produces distortion in the output signal.

### Requirements of biasing circuit

Following are the requirements of a proper biasing circuit :

(i) Proper selection of operating point ( $Q$ -point) such that on applying the input signal, the instantaneous operating point does not move either to the saturation region or to the cut-off region.

(ii) To stabilize the collector current against temperature variations.

(iii) To make the operating point independent of transistor parameters (*i.e.*, if the transistor is replaced by another transistor of the same type, the operating point should not shift).

**Stability Factor.** The transistor parameters  $\beta$ ,  $I_{C0}$  and  $V_{BE}$  are functions of temperature. Among these, the change in  $I_{C0}$  with temperature is more significant than other changes.

The stability factor  $S$  is defined as the rate of change of collector current  $I_C$  with respect to the reverse saturation current  $I_{C0}$  keeping  $\beta$  and  $V_{BE}$  constant, *i.e.*,

$$S = \frac{\partial I_C}{\partial I_{C0}} = \frac{\Delta I_C}{\Delta I_{C0}}$$

This expression shows that smaller is the value of  $S$ , higher is the stability. So the stability factor  $S$  should be kept as small as possible. The lowest value of  $S$ , that can be obtained, is unity since  $I_C$  must include  $I_{C0}$ . Closer is value of  $S$  to unity, lesser will be the variation of operating point with temperature.

There are three types of biasing techniques.

#### 35.6.1. Fixed Bias or Base Bias

In this method, a high resistance  $R_B$  is connected between the base and the positive end of the power supply  $V_{CC}$ , for *n-p-n* transistor (Fig. 35.17). Let us determine the value of  $R_B$ , so that required collector current  $I_C$  ( $= \beta I_B$ ) flows in the zero signal conditions.

Applying Kirchhoff's voltage law to the base-emitter loop  $ABENA$ , we get

$$V_{CC} = I_B R_B + V_{BE}$$

or

$$I_B R_B = V_{CC} - V_{BE}$$

∴

$$R_B = \frac{V_{CC} - V_{BE}}{I_B}$$

The value of  $V_{BE}$  is negligibly small, when compared with  $V_{CC}$ .

$$\therefore R_B = \frac{V_{CC}}{I_B}$$

The supply voltage  $V_{CC}$  is fixed. Once the value of  $R_B$  is selected, the base current  $I_B$  is also fixed. Hence the name fixed bias circuit.

#### Advantages

1. The circuit is very simple since only one resistor is required to bias the base.
2. The biasing conditions can be easily set up.
3. There is no loading of the source by the biasing circuit.

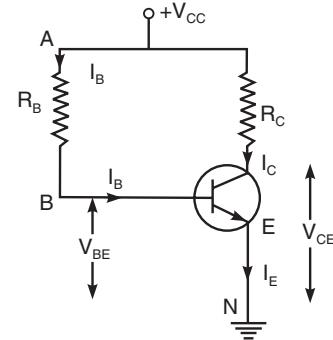


Fig. 35.17

### Disadvantages

1. Stabilisation is poor. The collector current  $I_C = \beta I_B$ . Variation of  $\beta$  will cause corresponding variation of collector current. Such collector current variation cannot be checked by the fixed bias circuit. The operating point is highly unstable.
2. The possibility for thermal runaway is large.  
This type is used rarely.

#### 35.6.2. Collector to Base Bias Circuit

A resistance  $R_B$  is connected between the collector and base (Fig. 35.18). The bias voltage is obtained from the collector of the transistor. Here the required zero signal base current is determined by the collector-base voltage  $V_{CB}$  and not by  $V_{CC}$ . The required value of  $R_B$  needed to give the zero signal collector current  $I_C$  can be determined by applying Kirchhoff's voltage law to Fig. 35.18.

$$\begin{aligned} V_{CC} &= (I_C + I_B) R_C + I_B R_B + V_{BE} \\ V_{CC} &= (I_C R_C + I_B R_B + V_{BE}) \\ &\quad (\text{since } I_B \ll I_C) \end{aligned}$$

$$\text{or } V_{CC} = \beta I_B R_C + I_B R_B + V_{BE} \quad (\because I_C = \beta I_B)$$

$$\text{or } R_B = \frac{V_{CC} - V_{BE} - \beta I_B R_C}{I_B}$$

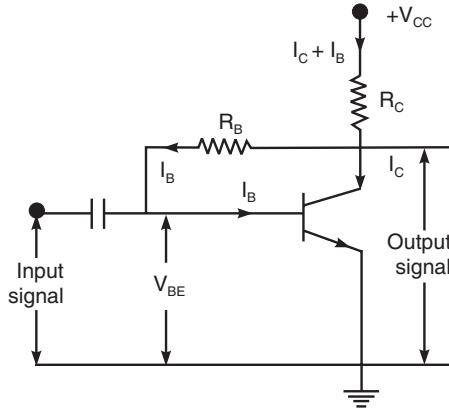


Fig. 35.18

**Stabilisation.** This circuit provides stabilisation of the operating point as discussed below.

When  $I_C$  tends to increase either because of rise in temperature or because of the transistor being replaced by another of larger  $\beta$ , then voltage  $V_{CE}$  decreases due to greater drop across  $R_C$  (since  $V_{CE} = V_{CC} - I_C R_C$ ). Consequently, lesser voltage is available across  $R_B$ . Hence base current  $I_B$  decreases which in turn tends to decrease the collector current to its original value. Thus a mechanism exists in the circuit which does not allow the collector current to increase and hence the operating point is stabilised.

**Disadvantages.**  $R_B$  not only provides a dc feedback for the stabilisation of operating point but it also causes a feedback of a.c. signal from output (collector) terminal to input (base) terminal. This reduces the voltage gain of the amplifier.

#### 35.6.3. Self Bias or Emitter Bias (Voltage-Divider Bias)

The circuit arrangement is shown in Fig. 35.19. Two resistors  $R_1$  and  $R_2$  are connected across supply voltage  $V_{CC}$  and provide biasing. The emitter resistor  $R_E$  provides stabilisation.  $R_1$  and  $R_2$  form a potential divider across  $V_{CC}$ . The resistor  $R_E$  causes a voltage drop in a direction so as to reverse-bias the emitter junction. Since the junction must be forward biased, the base voltage is obtained from the supply through  $R_1 - R_2$  network. The net forward bias across the emitter junction is equal to  $V_B$  minus the d.c. voltage drop across  $R_E$ .

The improvement in the operating point stability may be explained as follows : Let there be a rise in temperature. This causes a rise in  $I_{C0}$ , a rise in  $I_C$ , and in  $I_E$ . Now the current in  $R_E$  increases. As a result, the voltage drop across  $R_E$  increases and consequently the base current decreases. This

decreases the collector current. Thus the presence of  $R_E$  reduces the increase in  $I_C$  and improves the operating point stability. Figure 35.20 shows how the rising tendency of  $I_C$  is checked by the circuit.

**Circuit Analysis.** As  $I_B$  is very small, it can be taken that same current  $I_1$  is flowing through both the resistors  $R_1$  and  $R_2$ . Let us calculate collector current  $I_C$ .

$$I_1 = \frac{V_{CC}}{(R_1 + R_2)} \quad \dots(1)$$

The voltage  $V_2$  developed across  $R_2$  is given by

$$V_2 = \left( \frac{V_{CC}}{(R_1 + R_2)} \right) R_2 \quad \dots(2)$$

Applying KVL to the base circuit, we have

$$V_2 = V_{BE} + V_E = V_{BE} + I_E R_E$$

or

$$V_2 = V_{BE} + I_C R_E \quad (\because I_E \approx I_C)$$

∴

$$I_C = \frac{V_2 - V_{BE}}{R_E} \quad \dots(3)$$

Here  $I_C$  is almost independent of transistor parameters and hence good stabilisation is ensured.

The collector emitter voltage  $V_{CE}$  can be calculated as follows :

Applying KVL to the output circuit, we have

$$\begin{aligned} V_{CC} &= I_C R_L + V_{CE} + I_E R_E \\ &= I_C R_L + V_{CE} + I_C R_E \end{aligned}$$

$$\therefore V_{CE} = V_{CC} - I_C (R_L + R_E) \quad \dots(4)$$

**Advantages.** This biasing method is almost universally used on account of the following advantages :

- (i) Several combinations of potential divider resistors  $R_1$  and  $R_2$  are possible.
- (ii) A common power supply  $V_{CC}$  is sufficient for providing the bias related to emitter-base and collector-base junctions.
- (iii) The gain of the transistor is not influenced by negative feedback.
- (iv) The operating point is stable irrespective of change in temperature.

### 35.7 THERMAL RUNAWAY

The maximum average power  $P_{D(max)}$  which a transistor can dissipate depends upon the transistor construction. It may lie in the range from a few milliwatts to 200 W. This maximum power is limited by the temperature that the collector-to-base junction can withstand. The range of temperature lies between 60°C and 100°C for Ge transistor and 150°C and 225°C for Si transistor. If the temperature increases beyond this range, then the transistor burns out. The increase in the collector junction temperature is due to *thermal runaway*.

The flow of collector current  $I_c$  produces heat within the transistor. This increases the transistor temperature. If the heat is not quickly drawn off, the temperature of the collector-base junction will

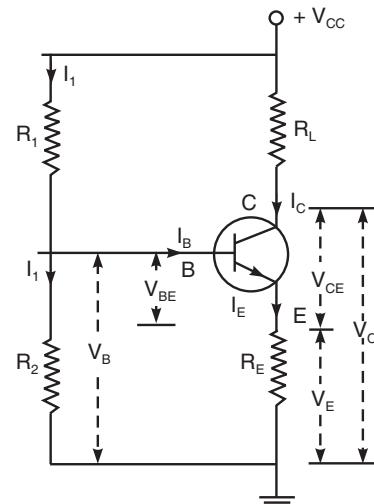


Fig. 35.19

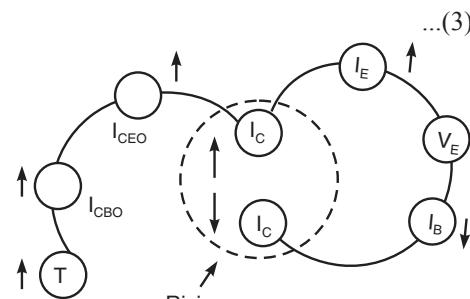


Fig. 35.20

increase. This increases the minority carrier density. Therefore, the collector leakage current  $I_{co}$  increases. This leakage current is highly influenced by temperature changes. It almost doubles for every  $6^{\circ}\text{C}$  rise in temperature in *Ge* and for every  $10^{\circ}\text{C}$  rise in *Si*. Any increase in  $I_{co}$  is magnified  $(1 + \beta)$  times and increases  $I_c$  considerably as seen from equation

$$I_c = \beta I_B + (1 + \beta) I_{co}.$$

This increased collector current further raises the transistor temperature that leads to further increase in  $I_{co}$ . If this succession of increases is allowed to continue, collector current may become very large in a matter of seconds, thereby damaging the transistor. **The self-destruction of an unstabilised transistor is known as thermal runaway.**

In order to avoid thermal runaway, some form of stabilisation is necessary. It is important that the operating point is stabilised.

The following two techniques are used for stabilization:

(i) *Stabilization techniques*. The technique consists in the use of resistive biasing circuit which permits such a variation of base current  $I_B$  so as to maintain  $I_c$  almost constant in spite of variation of  $I_{co}$ ,  $\beta$  and  $V_{BE}$ .

(ii) *Compensation techniques*. In this technique, temperature sensitive devices such as diodes, transistors, thermistors etc., are used. Such devices produce compensating voltages and currents in such a way that the operating point is maintained stable.

In power transistors, provisions are made to remove heat developed at the junction during working. This is achieved by using *heat sink*, which is a metal sheet connected to the collector and whose surface radiates heat quickly.

### Heat Sinks

As power transistors handle large currents, they get heated up during operation. Since transistor is a temperature dependent device, the heat generated must be taken out in order to keep the temperature within permissible limits. Generally, the transistor is fixed on a metal sheet (usually aluminium) so that additional heat is transferred. The metal sheet that serves to dissipate the additional heat from the power transistor is known as heat sink. Most of the heat within the transistor is produced at the collector junction. The heat sink increases the surface area and allows heat to escape from the collector junction easily. The result is that temperature of the transistor is sufficiently lowered. Thus heat sink is a direct practical means of combating the undesirable thermal effects e.g., thermal runaway. It may be noted that the ability of any heat sink to transfer heat to the surrounding depends upon the material, volume, area, shape and contact between case and sink and movement of air around the sink. Finned aluminium heat sinks yield the best heat transfer per unit cost. Metal case transistors are more readily cooled by conduction than plastic ones.

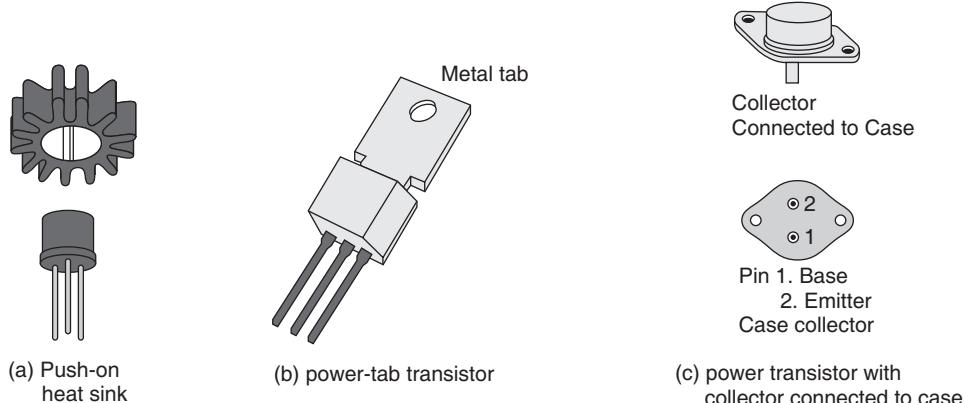


Fig. 35.21

In a power transistor, the collector is directly connected to transistor case. The case is connected to the chassis which acts as a heat sink. To prevent shorting, a thin mica washer is provided between case and chassis. Special silicon grease (heat sink compound) is used to get thermal contact between heat sink and casting.

Fig. 35.21 (a) shows one type of heat sink. When this is pushed onto the transistor case, heat radiates more quickly because of the increased surface area of the fins.

Figure 35.21 (b) shows the power-tab transistor. The metal tab provides a path out of the transistor for heat. This metal tab can be fastened to the chassis of electronics equipment. Because the chassis is a massive heat sink, heat can easily escape from the transistor to the chassis.

Large power transistors like Fig. 35.21(c) have the collector connected directly to the case to let heat escape as easily as possible. The transistor case is then fastened to the chassis. To prevent the collector from shorting to the chassis ground, a thin insulating washer and a thermal conductive paste are used between the transistor case and the chassis.

### LOAD-LINE ANALYSIS

## 35.8 THE D.C. LOAD LINE

When no signal is applied to a transistor circuit it is said to be in the *quiescent condition*. Fig. 35.22 shows a CE amplifier in the quiescent condition.

The collector circuit d.c. supply battery  $V_{CC}$  sends the collector current  $I_C$  through the load resistance  $R_L$  and the transistor. The voltage drop across the load is  $I_C R_L$ . Then the value of collector-emitter voltage  $V_{CE}$  at any instant is given by

$$V_{CE} = V_{CC} - I_C R_L \quad \dots(1)$$

$$\text{or} \quad I_C = \left( -\frac{1}{R_L} \right) V_{CE} + \frac{V_{CC}}{R_L} \quad \dots(2)$$

Here  $(-1/R_L)$  and  $(V_{CC}/R_L)$  are constants. Therefore, if  $I_C$  is plotted against  $V_{CE}$  (on the transistor's output characteristics), we get a straight line. This line is called the *dc load line* because it depends upon the value of load  $R_L$  used in the circuit.

The slope of the line is  $(-1/R_L)$ . The intercept on the Y-axis is  $(V_{CC}/R_L)$ .

**Plotting of DC load line.** To draw load line, we need two points of the straight line. They are located as follows :

When collector current  $I_C = 0$ , then collector-emitter voltage  $V_{CE}$  is given by Eq. (1).

$$V_{CE} = V_{CC}. \quad \dots(3)$$

Thus, we get point A at which  $I_C = 0$  and  $V_{CE} = V_{CC}$ .

When collector-emitter voltage  $V_{CE} = 0$ , then collector current is given by Eq. (2).

$$I_C = V_{CC}/R_L. \quad \dots(4)$$

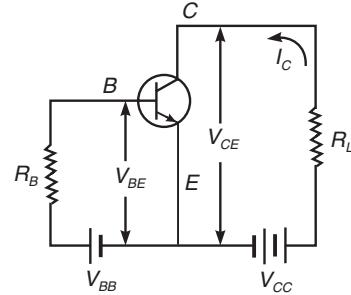


Fig. 35.22

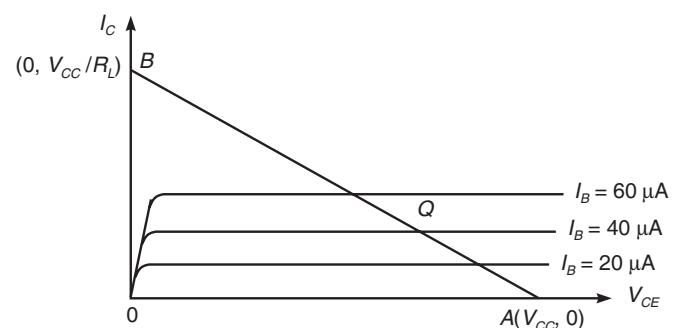


Fig. 35.23

Thus, we get point  $B$  at which  $I_C = V_{CC}/R_L$  and  $V_{CE} = 0$ .

Joining points  $A$  and  $B$ , we get the d.c. load line (Fig. 35.23).

### Location of the Operating Point

The operating point must naturally lie on the load line. To find exactly where on this line it lies we have to consider the base current  $I_B$ . The base current is decided by the values of  $R_B$ ,  $V_{BB}$  and  $V_{BE}$ . Application of Kirchhoff's voltage law to the base circuit leads us to the equation

$$V_{BB} = I_B R_B + V_{BE} \quad \dots(5)$$

The base current is, therefore, given by

$$I_B = \frac{V_{BB} - V_{BE}}{R_B}.$$

Once the base current is known the point where the characteristic for this base current intersects the load line gives the operating point. This point is called the *quiescent operating point* or simply the  $Q$  point.

### 35.9 A. C. LOAD LINE

AC Load line is the locus of  $Q$  point (operating point) when a.c. signal is applied at the input of the amplifier. It enables us to determine the values of collector current and collector-emitter voltage when signal is applied at the input.

The d.c. load line is  $AB$ . The a.c. load line is  $CD$  (Fig. 35.24). The intersection of a.c. load line and d.c. load line represents the  $Q$ -point.

In a.c. load line, a.c. load resistance ( $R_{ac}$ ) is taken into account, while in d.c. load line, d.c. load resistance ( $R_L$ ) is considered. DC load line can be drawn as described earlier.

In order to draw a.c. load line, two points on a plot of the output characteristics are required.

Let  $(V_{CE})_0$  and  $(I_C)_0$  be the zero signal voltage and zero signal collector current corresponding to the operating point, located on the d.c. line.

Let  $R'_L$  be the a.c. load resistance (effective load resistance found from a.c. equivalent circuit).

Let  $(I_C)_0 \times R'_L = x$ . It has the dimensions of voltage.

Let  $\frac{(V_{CE})_0}{R_L} = y$ . It has the dimensions of current.

Then, for the a.c. load line, the cut-off point on the  $V_{CE}$  axis is

$$V'_{CE} = (V_{CE})_0 + x.$$

The saturation current on the  $I_C$  axis is

$$I'_C = (I_C)_0 + y.$$

A line drawn with slope  $= -1/R'_L$  passing through the  $Q$ -point and having  $(V'_{CE}, 0)$  and  $(0, I'_C)$  as the extreme coordinates gives the a.c. load line for the load resistance  $R'_L$ .

Slope of a.c. load line is  $\tan \theta_2 = -1/R'_L$ .

Slope of d.c. load line is  $\tan \theta_1 = -1/R_L$ .

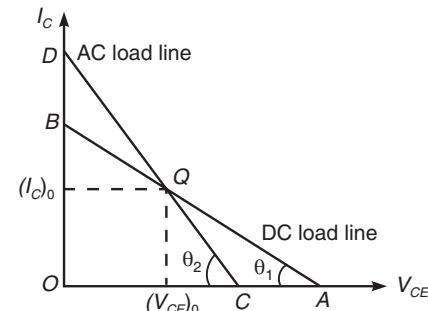


Fig. 35.24

### HYBRID PARAMETERS

## 35.10 TWO-PORT REPRESENTATION OF A TRANSISTOR

A transistor having three terminals is an active device and can be used in any of the three configurations, *CB*, *CE* and *CC*. In these configurations one of the terminals is common to the input and output circuits. Hence a transistor acts as a two port network, input port and the output port (Fig. 35.25).

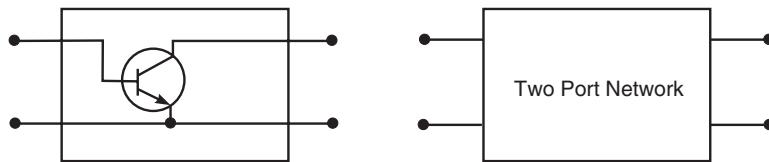


Fig. 35.25

For amplification purposes, the alternating currents, voltages to be amplified (called signals) are applied to the input port (pair of terminals) and the amplified signal is taken from output port (pair of output terminals). Then the performance of the amplifier is obtained by considering the transistor as a 'black box'.

### 35.10.1. The Hybrid Equivalent Circuit

To analyse the performance of transistors, they are conveniently represented by an equivalent circuit. The hybrid equivalent or the *h-parameter equivalent circuit* is widely used for small signal low frequency applications because of the following reasons.

- (i) The *h*-parameters can be measured easily.
- (ii) They are more independent of each other and other variables like frequency and operating point etc.
- (iii) The value of *h*-parameters nearly corresponds to actual values of input and output impedances and current gain for many applications.
- (iv) The *h*-parameters are real numbers at audio frequencies.
- (v) They are particularly suitable for circuit analysis and design and are specified by the transistor manufacturers.

**Hybrid Parameters.** A four terminal network or a two port network can be treated as a Black Box with two input terminals and two output terminals (Fig. 35.26).

For each pair of terminals, there are two variables, the current  $I$  and voltage  $V$ .

These four variables can be related by the following equations :

$$V_1 = h_{11} I_1 + h_{12} V_2 \quad \dots(1)$$

$$I_2 = h_{21} I_1 + h_{22} V_2 \quad \dots(2)$$

The parameters  $h_{11}$ ,  $h_{12}$ ,  $h_{21}$  and  $h_{22}$ , which relate the four variables of the two port system, are called *h-parameters*. They may be defined by first putting  $V_2 = 0$  (output terminals short circuited) and then  $I_1 = 0$  (Input terminals open circuited) in the above equations. Thus

$$h_{11} = \left| \frac{V_1}{I_1} \right|_{V_2=0} = \text{Input impedance (with output shorted)} = h_i$$

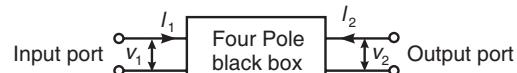


Fig. 35.26

$$h_{21} = \left| \frac{I_2}{I_1} \right|_{V_2=0} = \text{Forward current ratio (with output shorted)} = h_f$$

$$h_{12} = \left| \frac{V_1}{V_2} \right|_{I_1=0} = \text{Reverse voltage ratio (with input open)} = h_r$$

$$h_{22} = \left| \frac{I_2}{V_2} \right|_{I_1=0} = \text{Output admittance (with input open)} = h_0$$

$h_{11}$  and  $h_{22}$  (i.e.,  $h_i$  and  $h_0$ ) have dimensions of impedance and admittance respectively.  $h_{12}$  and  $h_{21}$  (i.e.,  $h_r$  and  $h_f$ ) have no units. Hence the name hybrid parameters. Another subscript ( $b$  for Base,  $e$  for Emitter and  $c$  for Collector) is added to designate the configuration of the transistor.

Thus,  $h_{ib}$ ,  $h_{rb}$ ,  $h_{fb}$ , and  $h_{ob}$  describe the  $CB$  configuration,  $h_{ie}$ ,  $h_{re}$ ,  $h_{fe}$ ,  $h_{oe}$  and  $h_{ic}$ ,  $h_{rc}$ ,  $h_{fc}$  and  $h_{oc}$  describe respectively the  $CE$  and  $CC$  configurations.

Typical values of  $h$ -parameters for a transistor for different configurations are listed in the following table.

Parameter	$CB$	$CE$	$CC$
$h_i (\Omega)$	30 – 80	600 – 2000	600 – 2000
$h_r$ (ratio)	$5 - 200 \times 10^{-6}$	$3 \times 10^{-4}$	1
$h_f$ (ratio)	– 0.98	10 – 100	– 51 – 100
$h_0$ (mhos)	$1.5 \mu \text{ mhos}$	$20 \mu \text{ mhos}$	$20 \mu \text{ mhos}$

### 35.10.2. Determination of the $h$ Parameters

$$\text{We have, } V_1 = h_{11} I_1 + h_{12} V_2 \quad \dots(1)$$

$$I_2 = h_{21} I_1 + h_{22} V_2 \quad \dots(2)$$

#### Determination of $h_{ie}$

It can be obtained from Eq. (1) by making  $V_2 = 0$ .

$$V_1 = h_{11} I_1$$

$$h_{11} = \frac{V_1}{I_1} = \text{Input resistance.}$$

$h_{11}$  gives the input resistance of transistor when  $V_2 = 0$ .  $V_2$  will be zero when we short circuit the output terminals.  $h_{11}$  is also called  $h_{ie}$  for common emitter mode.

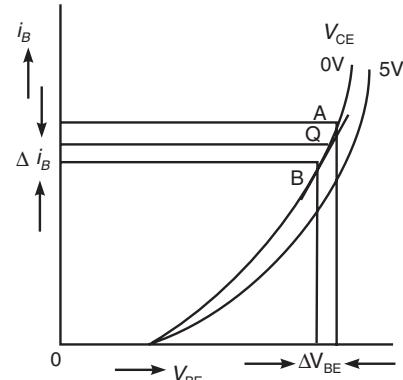


Fig. 35.27

Consider the input characteristics shown in Fig. 35.27. On the characteristics through the operating point  $Q$  choose two points  $A$  and  $B$  on either side of  $Q$  and very close to it. On the two axes read the intervals  $\Delta i_B$  and  $\Delta V_{BE}$  corresponding to  $AB$ . Then we have

$$h_{ie} = \frac{\Delta V_{BE}}{\Delta i_B}$$

as  $V_{CE}$  is the same at  $A$  and  $B$ .

**Determination of  $h_{re}$** 

Putting  $I_1 = 0$  in Eq. (1), we get

$$h_{12} = \frac{V_1}{V_2}$$

$h_{12}$  is also called  $h_{re}$  for CE mode.

Through the operating point  $Q$  draw a line parallel to the  $V_{BE}$  axis to cut an adjacent characteristic at  $A$  (Fig. 35.28).

$$h_{re} = \frac{\Delta V_{BE}}{V_{CE_2} - V_{CE_1}}$$

**Determination of  $h_{fe}$** 

Putting  $V_2 = 0$  in Eq. (2), we get

$$h_{21} = \frac{I_2}{I_1}.$$

$h_{21}$  is also called  $h_{fe}$  for CE mode.

Consider the output characteristics of a transistor (Fig. 35.29).

Through the operating point  $Q$  draw a line parallel to the  $i_c$  axis. Let this line cut the two adjacent characteristics on either side at  $A$  and  $B$ . Through  $A$  and  $B$  draw lines parallel to the  $V_{CE}$  axis to cut the  $i_c$  axis at  $i_{c1}$  and  $i_{c2}$ . Let the base current corresponding to  $A$  be  $i_{B1}$  and that corresponding to  $B$  be  $i_{B2}$ . Then,

$$h_{fe} = \frac{i_{c1} - i_{c2}}{i_{B1} - i_{B2}}$$

**Determination of  $h_{0e}$** 

Putting  $I_1 = 0$  in Eq. (2), we get

$$h_{22} = \frac{I_2}{V_2}.$$

$h_{22}$  is also called  $h_{0e}$  for CE mode.

In relation to the output characteristic curves  $h_{22}$  is defined as

$$h_{22} \approx \frac{\Delta i_c}{\Delta V_{CE}} \quad (i_B = \text{constant})$$

Through the operating point  $Q$  draw a line parallel to the characteristics (Fig. 35.30). On this line choose two close points  $A$  and  $B$  one on either side of the operating point. The intervals  $\Delta V_{CE}$  and  $\Delta i_c$  corresponding to  $AB$  are read on the two axes. Then

$$h_{0e} = \frac{\Delta i_c}{\Delta V_{CE}}$$

as  $i_B$  is the same at  $A$  and  $B$ .

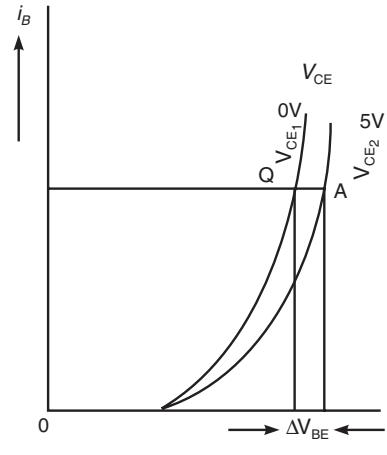


Fig. 35.28

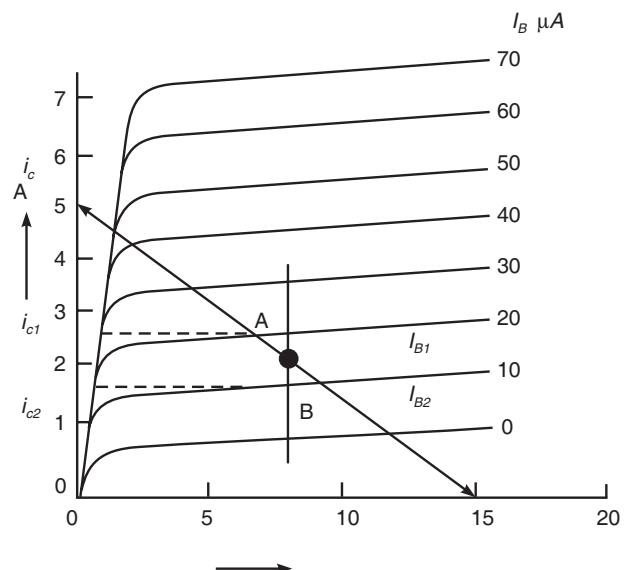


Fig. 35.29

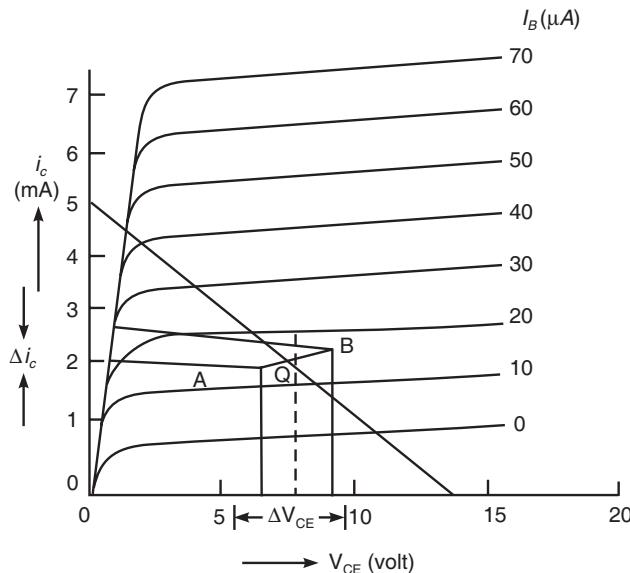


Fig. 35.30

### Advantages of $h$ -parameters

- (1) The  $h$ -parameters representation leads to simple algebraic expressions for the performance quantities, e.g., input impedance, output impedance, current gain, voltage gain and power gain of a transistor amplifier.
- (v) A set of  $h$ -parameters for a transistor is generally specified by the transistor manufacturers.

### Limitations of $h$ parameters

- (1) The  $h$  parameter model gives correct answers for small ac signals only. It is because a transistor behaves as a linear device for small signals only.
- (2) It is very difficult to get the exact values of  $h$  parameters for a particular transistor. It is because these parameters are subject to considerable variation—unit to unit variation, variation due to change in temperature and variation due to change in operating point. In predicting an amplifier performance, care must be taken to use  $h$  parameter values that are correct for the operating point being considered.
- (3) The parameters of the hybrid equivalent circuit are defined in general terms for any operating conditions. In other words, the hybrid parameters may not reflect the actual operating conditions but simply provide an indication of the level of each parameter to expect no matter what conditions actually exist.

## 35.11 PHOTO TRANSISTOR

Photo transistor is an  $n-p-n$  or  $p-n-p$  transistor in which the forward current is controlled by the incident light at the  $pn$  junction. Due to transistor action, the current is amplified. For the normal operation of the photo-transistor ( $n-p-n$ ), the base is kept open [Fig. 35.31]. The emitter region is given a negative potential while collector region is given a positive potential. Radiation is concentrated on the region near the collector junction  $J_c$ .

The photo-transistor is usually connected in a common-emitter configuration with the base open. The junction  $J_E$  is slightly forward-biased. The junction  $J_c$  is reverse-biased. Assume, first, that there is no radiant excitation. Under these circumstances minority carriers are gener-

ated thermally. The electrons crossing from the base to the collector, as well as the holes crossing from the collector to the base, constitute the reverse saturation collector current  $I_{CO}$ . The collector current is given by

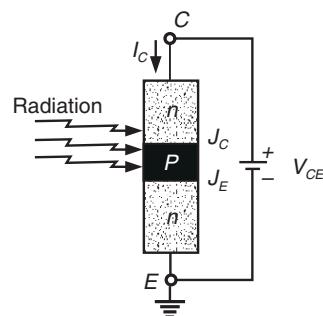
$$I_C = (\beta + 1) I_{CO} \quad \dots(1)$$

If the light is now turned on, additional minority carriers are photogenerated, and these contribute to the reverse saturation current in exactly the same manner as do the thermally generated minority charges. Let  $I_L$  be the component of the reverse saturation current due to the light. Then, the total collector current is

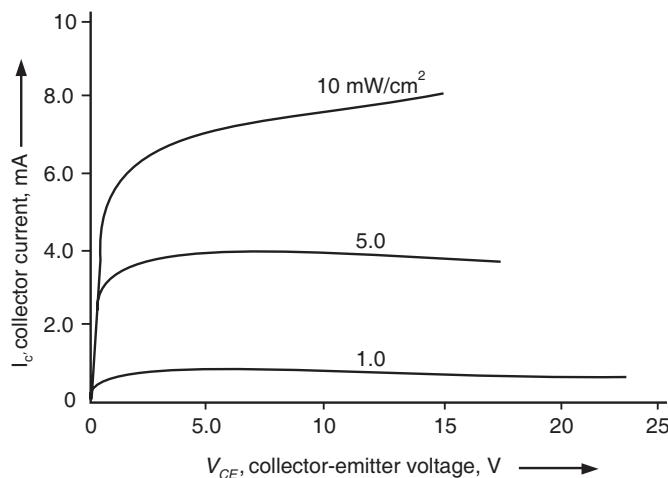
$$I_C = (\beta + 1) (I_{CO} + I_L) \quad \dots(2)$$

Due to transistor action, the current caused by the radiation is multiplied by the large factor  $(\beta + 1)$ .

The output volt-ampere characteristics of an *n-p-n* silicon phototransistor are shown in Fig. 35.32 for three different values of illumination intensities.



**Fig. 35.31**



**Fig. 35.32**

Photo transistors are used in light operated relays. They are also used to compare the intensities of different light sources and for infrared detection.

### EXERCISE

- Explain the working of a *n-p-n* transistor. *(Madras, 1995)*
- Define  $\alpha$  and  $\beta$  parameters of the transistors. Establish the relation  $\beta = \frac{\alpha}{1-\alpha}$ . *(Rohilkhand 1990)*
- Describe an experiment to draw the characteristic curves for a transistor in C-E mode. *(Bangalore 2000)*
- Explain the input and output characteristics of a transistor connected in common base mode. Hence explain the ac current gain.
- Explain the need for biasing a transistor. What are the different methods of biasing? *(K.U. April 1991)*
- Why is voltage divider bias commonly used in transistor? Discuss its advantages.
- What is meant by *Q*-point and *dc* load line? How will you draw *d.c.* load line on the output characteristics of a transistor? What is its importance?
- Describe the two port model of a transistor. Define the *h*-parameters of a transistor. Describe how the *h*-parameters of a transistor can be found from its characteristics.
- Justify the need for heat sink in power transistors.

## BJT AMPLIFIERS

**After reading this chapter, you should be able to**

- |                          |  |
|--------------------------|--|
| <input type="checkbox"/> | Discuss the important characteristics of the CE amplifier.   |
| <input type="checkbox"/> | Explain and analyze the operation of class A power amplifiers  |
| <input type="checkbox"/> | Discuss multistage amplifiers and analyze their operation.   |
| <input type="checkbox"/> | Analyze two-stage BJT CE capacitor-coupled amplifier circuit to determine its performance.             |
| <input type="checkbox"/> | Explain why voltage gain of an amplifier with negative feedback is smaller than that with no feedback. |

### 36.1 COMMON-EMITTER TRANSISTOR AMPLIFIER

Fig. 36.1 shows the common-emitter amplifier circuit using *n-p-n* transistor. The emitter is common to both the input and output circuits. The emitter is forward biased by using base-emitter battery  $V_{BB}$ . Due to the forward bias, the resistance of input circuit is low. The collector is reverse biased by using collector-emitter battery  $V_{CC}$ . The low input voltage signal is applied in base-emitter circuit (input circuit). The amplified output is obtained across the collector and emitter.

In a common-emitter circuit, the collector-current is controlled by the base-current rather than the emitter-current. We know that in a transistor a large collector-current corresponds to a very small base-current. Therefore, with input signal applied to the base, a very small change in base-current produces a much more larger change in the collector-current. Thus, a very substantial *current-gain* is obtained.

**(a) D.C. current gain :** It is the ratio of the collector current to the base current. It is denoted by  $\beta$ .

$$\therefore \text{d.c. current gain, } \beta = \frac{I_C}{I_B}.$$

**(b) A.C. current gain :** It is defined as the ratio of the change in collector current to the change in base current at constant collector voltage. It is denoted by  $\beta_{ac}$ .

$$\therefore \beta_{ac} = \left( \frac{\Delta I_C}{\Delta I_B} \right)_{V_{CC}}$$

Its value is quite large as compared to 1 and lies between 150 and 50.

**(c) A.C. voltage gain :** It is the ratio of the change in output voltage to the change in input voltage. It is denoted by  $A_V$ .

$$A_V = \beta_{ac} \times \text{resistance gain} .$$

$$(d) \text{A.C. power gain} = \frac{\text{Change in output power}}{\text{Change in input power}}$$

$$\text{a.c. power gain} = \beta_{ac}^2 \times \text{resistance gain} .$$

**(e) Phase relationship.** In the common-emitter amplifier, the output signal is in antiphase with the input signal.

#### Practical Circuit of Common-Emitter Amplifier

A transistor gives a faithful amplification only if proper associated circuit elements are used. Fig. 36.2 shows a practical single stage transistor amplifier.

Its various circuit elements and their functions are explained below :

**(i) Biasing circuit.** The resistors  $R_1$ ,  $R_2$ , and  $R_E$  form the biasing and stabilisation circuit. The biasing circuit must establish a proper operating point. Otherwise a part of the negative half cycle of signal may be cut off in the output.

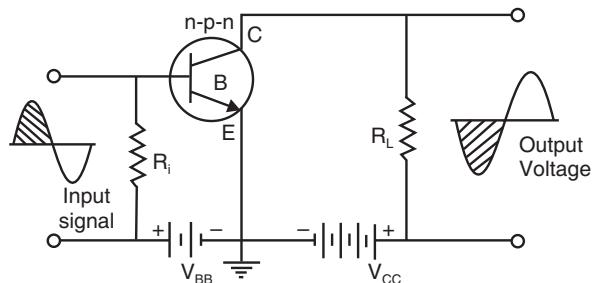
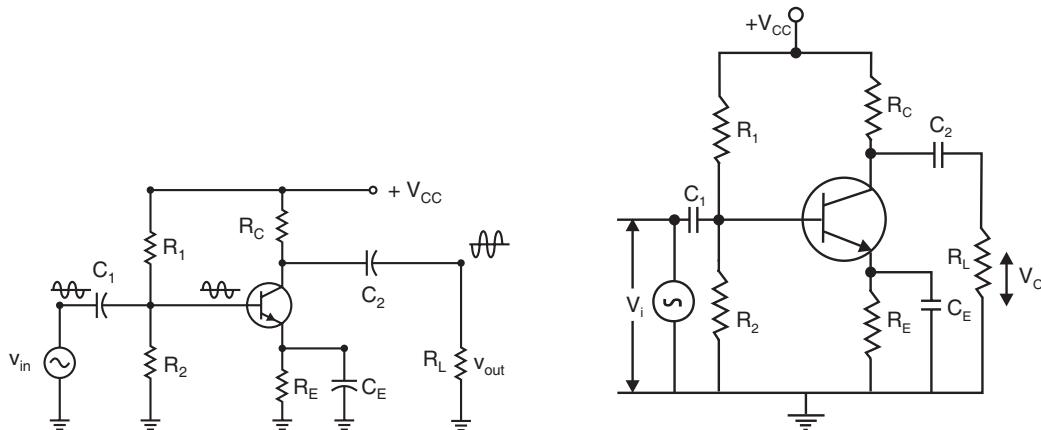


Fig. 36.1



**Fig. 36.2.** VDB (Voltage-divider biased) amplifier with waveforms.

(The student has to draw only one of the two diagrams)

**(ii) Input capacitor  $C_1$ .** An electrolytic capacitor ( $C_1$ ) is used to couple the a.c. signal voltage to the base of the transistor. If it is not used, the resistance of the signal source will be in parallel to  $R_2$  and the bias voltage of the base will change. Thus  $C_1$  allows only a.c. from the signal source to flow into the input circuit and isolates the signal source from  $R_2$ .

**(iii) Emitter bypass capacitor  $C_E$ .** An emitter bypass electrolytic capacitor is used in parallel with  $R_E$  to provide a low resistance path to the amplified A.C. signal. If it is not used, then amplified A.C. signal flowing through  $R_E$  will cause a voltage drop across it, thereby reducing the output voltage.

**(iv) Coupling capacitor  $C_2$ .** The coupling electrolytic capacitor  $C_2$  couples one stage of amplification to the next stage. It prevents the d.c. component of the output of the first stage from reaching the input of the second stage. But it allows the passage of A.C. signal. If it is not used,  $R_C$  will come in parallel with the resistance  $R_1$  of the biasing network of the next stage and the bias conditions of the next stage will be drastically changed.

**Working.** When a weak A.C. signal is applied to the base of transistor, a small base current starts flowing. Due to the transistor action a much larger A.C. current flows through the collector load  $R_C$ . As the value of  $R_C$  is quite high, a large voltage drop appears across  $R_C$ . Thus a weak signal appears in amplified form in the collector-circuit. In this way, transistor acts as an amplifier.

### Voltage Gain

The voltage gain of an amplifier is defined as the ac output voltage divided by the ac input voltage.

$$A_V = \frac{v_{out}}{v_{in}} \quad \dots(1)$$

$$v_{out} = A_V v_{in} \quad \dots(2)$$

Eq. (2) is useful when you want to calculate the value of  $v_{out}$ , given the values of  $A_V$  and  $v_{in}$ .

**Example.** If we measure an ac load voltage of 50 mV with an ac input voltage of 100  $\mu$ V, the voltage gain is:

$$A_V = \frac{50 \text{ mV}}{100 \mu\text{V}} = 500$$

This says that the ac output voltage is 500 times larger than the ac input voltage.

### A.C. Equivalent Circuit

Both d.c. and a.c. conditions prevail in a transistor amplifier. When no signal is applied, d.c. base current  $I_B$ , collector current  $I_C$  and emitter current  $I_E$  flow due to biasing circuit. With the application of signal, a.c. base current  $i_b$ , a.c. collector current  $i_c$  and a.c. emitter current  $i_e$  also flow. However, the total current in any branch is the sum of d.c. and a.c. currents through the branch.

For drawing a.c. equivalent circuit of transistor amplifier, we consider only the a.c. conditions. Thus,

(i) All d.c. sources are reduced to zero (i.e.,  $V_{CC} = 0$ ).

(ii) All the capacitors are short circuited.

The a.c. equivalent circuit is shown in Fig. 36.3. It is used to calculate the voltage gain.

Collector load  $R_C$  appears in parallel with load  $R_L$ . Therefore, effective load  $R_{AC}$  for a.c. is given by

$$R_{AC} = \frac{R_C R_L}{R_C + R_L}$$

Output voltage,  $V_0 = i_c R_{AC}$

Input voltage,  $V_i = i_b R_i$

where  $R_i$  is the input impedance.

Voltage gain ( $A_V$ ) is the ratio of a.c. output voltage to the a.c. input signal voltage.

$$\therefore A_V = \frac{V_0}{V_i} = \frac{i_c R_{AC}}{i_b R_i} = \beta \frac{R_{AC}}{R_i}$$

where  $\beta = \frac{i_c}{i_b}$  is the alternating current amplification factor.

**Power gain.** Power gain is given by

$$\text{Power gain} = \text{current gain} \times \text{voltage gain}$$

$$A_p = \frac{i_c}{i_b} \times \frac{i_c R_{AC}}{i_b R_i} = \beta^2 \times \frac{R_{AC}}{R_i}.$$

#### Definitions.

1. **Gain.** The amount by which an electrical signal is increased or amplified is called **gain**.

*The ratio of the output electrical quantity to the input one of the amplifier is called its gain.*

It can be current gain or voltage gain or power gain.

2. **Frequency response.** The voltage gain of an amplifier varies with signal frequency.

Fig. 36.4 shows the variation of gain with frequency.

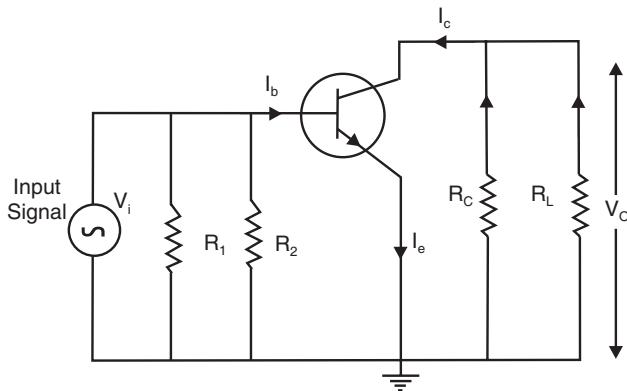


Fig. 36.3

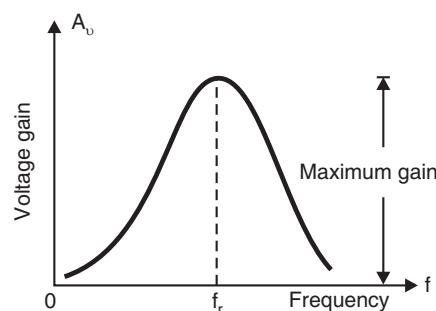


Fig. 36.4

It is called **frequency response** graph.

- At low frequencies, the gain increases with the frequency.
  - The gain becomes maximum at  $f_r$ , called *resonant frequency*.
  - If the frequency of signal increases beyond  $f_r$ , the gain decreases. At high frequencies, the gain is found to decrease with increase of frequency.
- 3. Bandwidth.** The range of frequency over which the voltage gain is equal to or greater than 70.7% of the maximum gain is called bandwidth.

The **bandwidth (BW)** of a resonant circuit is defined as:

$$BW = f_2 - f_1.$$

Here,  $f_1$  = lower half-power frequency,

$f_2$  = upper half-power frequency.

The half-power frequencies are identical to the frequencies at which the voltage gain equals 0.707 times the maximum gain, as shown in Fig. 36.5. The smaller  $BW$  is, the narrower the bandwidth of the amplifier.

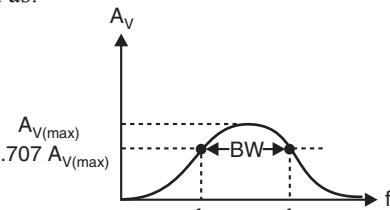


Fig. 36.5

## 36.2 COMMON-BASE TRANSISTOR AMPLIFIER

Fig. 36.6 shows the common-base amplifier circuit using *n-p-n* transistor. The weak signal to be amplified is applied between the emitter and the base. The output is taken across the load  $R_L$  connected between the collector and the base. The d.c. voltage  $V_{EE}$  is applied in the input circuit in order to keep the input section always forward biased irrespective of polarity of the input signal. The d.c. voltage  $V_{CC}$  is chosen such that it keeps the output section always reverse biased.

The input circuit has a low resistance. Hence a small change in the signal voltage will cause an appreciable change in the emitter current. This causes almost the same change in the collector current due to transistor action. Thus, collector current flowing through the load resistor causes a large voltage drop. Thus a weak signal applied in the input circuit gets amplified in the collector circuit. *The output voltage is in phase with the input signal.*

**Current amplification factor :** The ratio of the change in the collector current to the change in the emitter current at constant collector-base voltage ( $V_{CB}$ ) is called the current amplification factor  $\alpha$ .

$$\text{Thus } \alpha = \left( \frac{\Delta I_C}{\Delta I_E} \right)_{V_{CB}}.$$

The current amplification factor is always less than 1.

**Example 1.** Explain why *CE* configuration is widely used in amplifier circuits.

**Solution.** A good amplifier stage is one which has a high input resistance and a low output resistance. A transistor in the *CB* configuration has a low input resistance (about  $20\ \Omega$ ) and a high output resistance (about  $1\ M\Omega$ ). This is just the opposite to what is desirable in a good amplifier. Hence the *CB* configuration is not useful for amplification. In a *CE* configuration, the input resistance is about  $1\ k\Omega$  and the output resistance is about  $10\ k\Omega$ . Therefore a transistor in the *CE* configuration makes a good amplifier. Further the current gain, voltage gain and power gain are all much greater in the *CE* configuration than in the *CB* configuration.

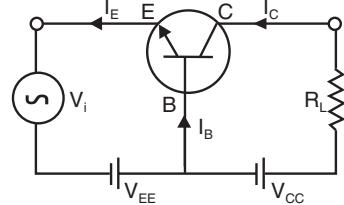


Fig. 36.6

### 36.3 HYBRID EQUIVALENT CIRCUIT OF COMMON Emitter AMPLIFIER

Fig. 36.7 shows the common-emitter *NPN* transistor amplifier circuit.  $R_g$  is the output resistance of input signal and  $R_L$  is the load resistance.

The general *h*-parameter expressions become,

$$V_i = V_{be} = h_{ie} I_b + h_{re} V_o \quad \dots(1)$$

$$I_c = h_{fe} I_b + h_{0e} V_0 \quad \dots(2)$$

where

$$V_0 = V_{ce}$$

From Eq. (1), we get

$$I_b = \frac{V_i - h_{re} V_o}{h_{ie}} \quad \dots(3)$$

The d.c. voltage of the collector with respect to the emitter is given by

$$V_{ce} = V_{cc} - I_c R_L$$

$\therefore$

$$dV_{ce} = -R_L dI_c$$

or

$$V_{ce} = -R_L i_c$$

In terms of usual notations, we can write

$$V_{ce} = -R_L I_c$$

or

$$V_0 = -R_L I_c$$

Substituting the value of  $V_0$  in Eq. (2), we have

$$I_c = h_{fe} I_b - h_{0e} R_L I_c$$

or

$$h_{fe} I_b = h_{0e} R_L I_c + I_c$$

$$= \frac{I_c R_L}{1/h_{0e}} + \frac{I_c R_L}{R_L} \quad \dots(4)$$

Equation (3) indicates that the base-emitter circuit is equivalent to a.c. voltage source of  $h_{re} V_0$  which opposes the a.c. input voltage  $V_i$  and is connected in series with the input resistance  $h_{ie}$ .

Equation (4) indicates that the collector-emitter circuit is equivalent to current source which supplies a current  $h_{fe} I_b$  and in parallel of which are connected the output resistance  $1/h_{0e}$  and load resistance  $R_L$ .

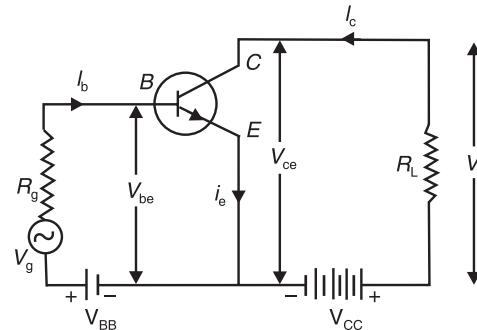


Fig. 36.7

( $\because V_{cc}$  is constant)

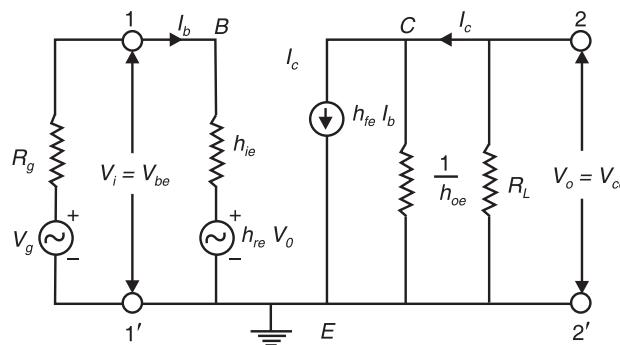


Fig. 36.8

Accordingly, the equivalent circuit is drawn in Fig. 36.8. Here the a.c. voltage source  $h_{re} V_0$ , which acts in opposition to the input signal  $V_i$ , represents the 'feedback' of the output voltage to

the input circuit. The current source of magnitude  $h_{fe} I_b$  may be looked as if the input current  $I_b$  is amplified and appears as  $h_{fe} I_b$  in the output circuit. Thus  $h_{fe} = \beta$ , the current amplification factor.

### 36.3.1. Analysis of a Transistor CE Amplifier using h-parameters

Figure 36.8 shows the  $h$ -parameters equivalent circuit of a common emitter transistor amplifier. Here,

- $h_{ie}$  = input impedance,
- $h_{0e}$  = output admittance,
- $h_{fe}$  = forward current gain,
- $h_{re}$  = reverse voltage transfer ratio of the transistor.

The signal source  $V_g$  is across the input port along with its source impedance  $R_g$ . The load resistance  $R_L$  appears across the output port.  $V_i$  and  $V_0$  are the input and output signals respectively. The input and output currents are taken to be positive, while flowing inward. This circuit is an *a.c.* equivalent circuit and *d.c.* values do not appear in the circuit.  $I_b$  and  $I_c$  are the input and output currents, with the presence of the source and load.

We will now derive expressions for current gain, voltage gain, input impedance, output impedance and power gain.

(i) **Current Gain.** Let  $Z$  be the equivalent impedance of  $1/h_{0e}$  and  $R_L$  in parallel. Then,

$$\frac{1}{Z} = 1/\frac{1}{h_{0e}} + \frac{1}{R_L} = h_{0e} + \frac{1}{R_L}$$

or  $Z = \frac{R_L}{1 + h_{0e} R_L}$

Voltage across  $R_L$  = voltage across  $Z$

or  $I_c R_L = h_{fe} I_b (Z) = h_{fe} I_b \left( \frac{R_L}{1 + h_{0e} R_L} \right)$

or  $\frac{I_c}{I_b} = \frac{h_{fe}}{1 + h_{0e} R_L}$

Current Gain  $A_{ie} = \frac{\text{Output Current}}{\text{Input Current}}$

$$\therefore A_{ie} = \frac{I_c}{I_b} = \frac{h_{fe}}{1 + h_{0e} R_L}. \quad \dots(1)$$

(ii) **Input impedance.** The input impedance  $Z_{ie}$  of the transistor is the impedance at the input terminals 1 and 1'.

$$\text{Input impedance } Z_{ie} = \frac{\text{Input Voltage}}{\text{Input Current}} = \frac{V_i}{I_b}$$

But

$$\begin{aligned} V_i &= h_{ie} I_b + h_{re} V_0 \\ &= h_{ie} I_b + h_{re} (-I_c R_L) \quad (\because V_0 = -I_c R_L) \\ Z_{ie} &= \frac{V_i}{I_b} = h_{ie} - h_{re} R_L \left( \frac{I_c}{I_b} \right) \end{aligned}$$

$$\therefore Z_{ie} = h_{ie} - h_{re} R_L A_{ie} = h_{ie} - \frac{h_{re} \cdot h_{fe} \cdot R_L}{(1 + h_{0e} \cdot R_L)} \quad \dots(2)$$

## (iii) Voltage gain.

$$\text{Voltage gain } A_{ve} = \frac{\text{Output Voltage}(V_0)}{\text{Input Voltage}(V_i)}$$

But

$$\begin{aligned} V_0 &= -I_c R_L \\ A_{ve} &= -\frac{I_c R_L}{V_i} = -\left(\frac{I_c}{I_b}\right)\left(\frac{I_b}{V_i}\right)R_L \\ &= -A_{ie}\left(\frac{1}{Z_{ie}}\right)R_L = -\frac{A_{ie} R_L}{Z_i} \end{aligned} \quad \dots(3)$$

Substituting the value of  $Z_i = Z_{ie} = h_{ie} - h_{re} R_L A_{ie}$  from Eq. (2),

$$A_{ve} = -\frac{A_{ie} R_L}{h_{ie} - h_{re} R_L A_{ie}} = \frac{R_L}{\frac{h_{ie}}{A_{ie}} - h_{re} R_L}$$

Substituting

$$\begin{aligned} A_{ie} &= \frac{h_{fe}}{1 + h_{0e} R_L} \text{ from Eq. (1), we get} \\ A_{ve} &= -\frac{R_L}{\frac{h_{ie}(1 + h_{0e} R_L)}{h_{fe}} - h_{re} R_L} \\ &= -\frac{h_{fe} R_L}{h_{ie}(1 + h_{0e} R_L) - h_{fe} h_{re} R_L} \\ &= -\frac{h_{fe} R_L}{h_{ie} + (h_{ie} h_{0e} - h_{fe} h_{re}) R_L} \\ &= -\frac{h_{fe} R_L}{h_{ie} + R_L \Delta h} \end{aligned} \quad \dots(4)$$

where  $\Delta h = h_{ie} h_{0e} - h_{fe} h_{re}$ .

The negative sign shows that the input and the output are  $180^\circ$  out of phase.

(iv) Output impedance. The output impedance  $Z_0$  of an amplifier is defined as the ratio of the output voltage to the output current with the input signal generator  $V_g$  reduced to zero and replaced by its internal resistance  $R_g$  and an a.c. voltage source  $V_0$  (rms) applied to the output terminals as shown in Fig. 36.9. Thus

$$Z_{0e} = \frac{V_0}{I_c}$$

where  $I_c$  is the current sent by the applied source.

Since the current through the output resistance  $1/h_{0e}$  is  $I_c - h_{fe} I_b$ , the output voltage  $V_0$  is given by

$$V_0 = (I_c - h_{fe} I_b) \frac{1}{h_{0e}}$$

or

$$h_{0e} V_0 = I_c - h_{fe} I_b \quad \dots(5)$$

But the base current  $I_b$  is given by

$$I_b = -\frac{h_{re} V_0}{h_{ie} + R_g}$$

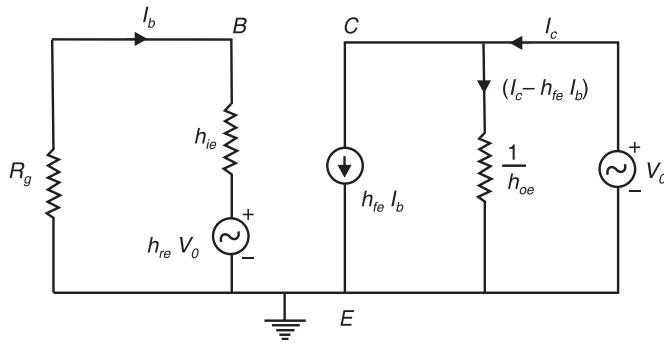


Fig. 36.9

Substituting the value of  $I_b$  in Eq. (5), we get

$$h_{0e} V_0 = I_c + \frac{h_{fe} h_{re}}{h_{ie} + R_g} V_0$$

or  $V_0 \left( h_{0e} - \frac{h_{fe} h_{re}}{h_{ie} + R_g} \right) = I_c$

or  $Z_{0e} = \frac{V_0}{I_c} = \frac{1}{h_{0e} - \frac{h_{fe} h_{re}}{h_{ie} + R_g}}$

$\therefore Z_{0e} = \frac{h_{ie} + R_g}{h_{0e} (h_{ie} + R_g) - h_{fe} h_{re}} \quad \dots(6)$

(v) **Power gain.** Power gain of the amplifier is the product of current gain and voltage gain. Thus power gain

$$A_{pe} = |A_{ve}| \times |A_{ie}|$$

Substituting the values of  $A_{ve}$  and  $A_{ie}$  from Eqs. (4) and (1), we get

$$\begin{aligned} A_{pe} &= \left( \frac{h_{fe} R_L}{h_{ie} + R_L \Delta h} \right) \left( \frac{h_{fe}}{1 + h_{0e} R_L} \right) \\ &= \frac{h_{fe}^2 R_L}{(1 + h_{0e} R_L)(h_{ie} + R_L \Delta h)} \quad \dots(7) \end{aligned}$$

where  $\Delta h = h_{ie} h_{0e} - h_{fe} h_{re}$ .

In actual practice,  $h_{0e}$ ,  $h_{re}$  are very small quantities.  $h_{0e} < 1$  and  $R_L \Delta h < h_{ie}$ .

$\therefore A_{pe} = \frac{h_{fe}^2 R_L}{h_{ie}}$ .

## POWER AMPLIFIERS

### 36.4 POWER AMPLIFIER

A transistor amplifier which raises the power level of the signals is known as power amplifier. Here a concentrated effort is made to obtain maximum output power. A power amplifier actually takes power from the *d.c.* supply and delivers it as useful signal power to the load. The ratio of signal power delivered as output to the *d.c.* power input is known as the efficiency of the power amplifier.

In the case of power amplifier, the power drawn by the load impedance is high. This is because of high collector current in this power amplifier. High collector current is the result of low load resistances driven by these amplifiers. These amplifiers (power) are used in the last stage of cascaded amplifier. The last stage drives low resistances like that of a loudspeaker. Power amplifiers use *power transistor* as amplifying device. Power transistor is different from ordinary bipolar transistor in following aspects :

- (i) Doping level of emitter and base is high.
- (ii) Ohmic resistance between emitter and base is decreased by increasing the contact area between the base layer and base lead giving it ring like form.
- (iii) The area of collector is more and collector is often attached to a metallic heat sink in order to dissipate the generated heat.

Power amplifiers are used mainly in four modes — *class A*, *class B*, *class AB* and *class C*.

**Class A Power Amplifier.** It operates in the linear region at all times (Fig. 36.10).

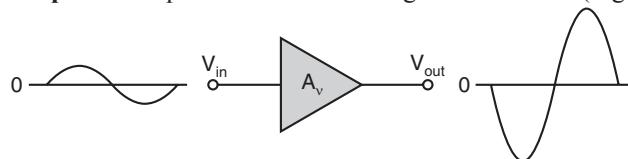


Fig. 36.10

- Output is shown  $180^\circ$  out of phase with the input (inverted).
- Class A power amplifiers are large-signal amplifiers with the objective of providing power (rather than voltage) to a load.
- In a power amplifier, it is necessary to consider the problem of heat dissipation in components.

**Power Gain.** The *power gain* of an amplifier is the ratio of the power delivered to the load to the input power.

$$A_p = \frac{P_L}{P_{in}}$$

Here,

$A_p$  = power gain

$P_L$  = signal power delivered to the load

$P_{in}$  = signal power delivered to the amplifier.

**Efficiency.** The efficiency of any amplifier is the ratio of the signal power supplied to a load to the power from the dc supply.

#### 36.4.1. Class A Power Amplifier

Figure 36.11 shows a typical *class A transistor power amplifier* operating in the common emitter mode. Here resistors  $R_1$  and  $R_2$  form voltage divider circuit to provide forward biasing of the emitter-base junction.  $R_E$  is the emitter resistor for bias stabilisation and  $C_E$  is the emitter bypass capacitor which prevents the a.c. voltages from appearing across  $R_E$ . The capacitor  $C_B$  is called '*Blocking capacitor*'. It is used to block d.c. component of the input signal so that only a.c. signal reaches the base. The amplifier is directly coupled to the load resistance  $R_L$ .

##### Operation of the Amplifier

The collector characteristics of the transistor used in Fig. 36.11 is shown in Fig. 36.12. The load line corresponding to  $R_L$  is drawn on the characteristics. The operating point  $Q$  is selected in almost the middle of a.c. load line.

When no signal is applied, the  $Q$ -point is operating point. When an a.c. signal is applied, the operating point shifts. It keeps on oscillating along the load line with  $Q$  as the mean position. The maximum signal condition occurs when the operation swings between saturation and cut-off.

**Efficiency.** The curves, in Fig. 36.12, show the voltage and current values at saturation and cut-off. In saturation stage, collector current is maximum and collector voltage is minimum. In cut-off stage, voltage is maximum and current is minimum. We denote voltage and current at saturation by  $V_{C\min}$  and  $I_{C\max}$  while these quantities at cut-off are denoted by  $V_{C\max}$  and  $I_{C\min}$  and at  $Q$ -point are  $V_{CQ}$  and  $I_{CQ}$ .

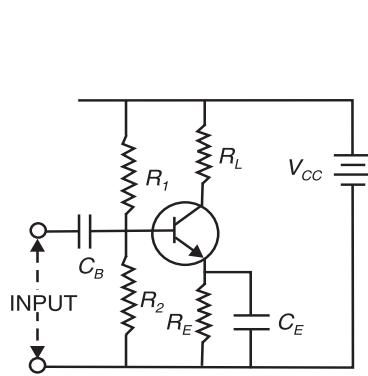


Fig. 36.11

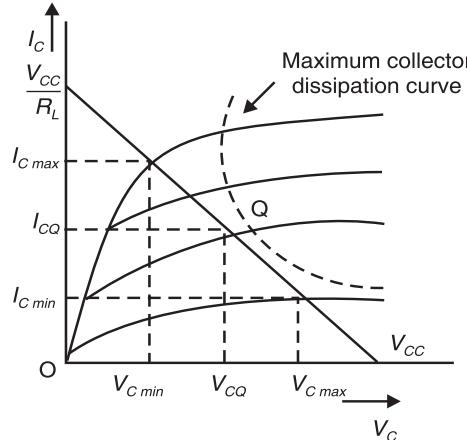


Fig. 36.12

Efficiency  $\eta$  is given by

$$\eta = \frac{\text{output a.c. power}}{\text{input d.c. power}} \quad \dots(1)$$

Input d.c. power is given by

$$P_{in} = V_{cc} I_{CQ} \quad \dots(2)$$

Output a.c. power is given by

$$P_{out} = V_{Crms} I_{Crms} \quad \dots(3)$$

Here  $V_{Crms}$  and  $I_{Crms}$  are the r.m.s. values of voltage and current. The signal makes excursions between  $V_{C\max}$  and  $V_{C\min}$  (or  $I_{C\max}$  and  $I_{C\min}$ ). So, peak values  $V_{Cm}$  and  $I_{Cm}$  of voltage and current are given by

$$V_{Cm} = \frac{V_{C\max} - V_{C\min}}{2}$$

$$I_{Cm} = \frac{I_{C\max} - I_{C\min}}{2}$$

Therefore, the a.c. output power

$$\begin{aligned} P_{out} &= V_{Crms} I_{Crms} \\ &= \frac{V_{Cm}}{\sqrt{2}} \times \frac{I_{Cm}}{\sqrt{2}} \\ &= \left( \frac{V_{C\max} - V_{C\min}}{2\sqrt{2}} \right) \times \left( \frac{I_{C\max} - I_{C\min}}{2\sqrt{2}} \right) \end{aligned}$$

$$= \frac{(V_{Cmax} - V_{Cmin})(I_{Cmax} - I_{Cmin})}{8} \quad \dots(4)$$

Therefore, the efficiency of power conversion

$$\eta = \frac{P_{a.c.}}{P_{d.c.}} = \frac{(V_{Cmax} - V_{Cmin})(I_{Cmax} - I_{Cmin})}{8V_{CC} I_{CQ}} \quad \dots(5)$$

For ideal collector characteristic curves

$$\begin{aligned} V_{Cmin} &= 0, \quad V_{Cmax} = V_{CC} \\ I_{Cmin} &= 0, \quad I_{Cmax} = 2I_{CQ} \\ \therefore \eta &= \frac{(V_{CC})(2I_{CQ})}{8V_{CC} I_{CQ}} = 0.25. \end{aligned}$$

Thus the maximum efficiency obtainable from a class A amplifier when coupled directly to the load resistance is only 25%.

### 36.4.2. Transformer Coupled Class A Power Amplifier

Figure 36.13 shows the circuit diagram of a transformer coupled class A power amplifier.  $R_1$  and  $R_2$  form a potential divider arrangement to give base bias.  $V_{CC}$  provides reverse bias to the collector-base junction. The transformer is used to couple the transistor to the load. The transformer matches the output impedance of the transistor to the load resistance so that power transfer is maximum.

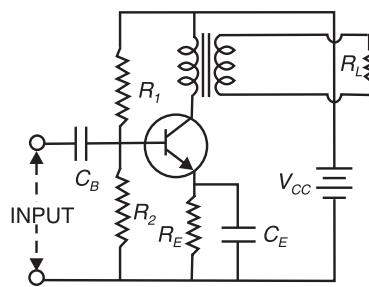


Fig. 36.13

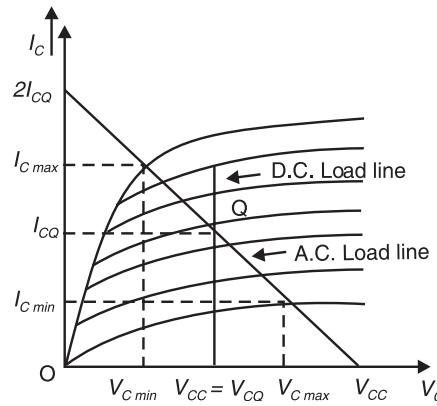


Fig. 36.14

In the circuit shown, load resistance  $R_L$  is connected across the secondary coil of the transformer. Its value as seen from the primary coil terminal is given by

$$R_L' = n^2 R_L.$$

Here,  $n$  is the turns ratio of the transformer given by

$$n = \frac{N_p}{N_s}.$$

Here  $N_p$  and  $N_s$  are the number of turns in primary and secondary coils, respectively, in practical circuits. Secondary coil also has d.c. resistance, say  $R_s$ . The total resistance as seen at the primary coil terminals is given by

$$R_L' = n^2 (R_s + R_L) \quad \dots(1)$$

Efficiency,

$$\eta = \frac{\text{output a.c. power}}{\text{input d.c. power}}.$$

d.c. input power is given by

$$P_{in} = V_{CC} I_{CQ} \quad \dots(2)$$

where  $I_{CQ}$  is the collector current at the operating point.

Output a.c. power is given by

$$\begin{aligned} P_{out} &= V_{Crms} I_{Crms} \\ &= \frac{V_{Cm}}{\sqrt{2}} \cdot \frac{I_{Cm}}{\sqrt{2}} \\ &= \frac{(V_{Cmax} - V_{Cmin})(I_{Cmax} - I_{Cmin})}{8} \end{aligned} \quad \dots(3)$$

From Fig. 36.14, it can be seen that

$$\begin{aligned} V_{Cmax} &= 2V_{CC}, \quad V_{Cmin} = 0 \\ I_{Cmax} &= 2I_{CQ}, \quad I_{Cmin} = 0 \\ P_{out} &= \frac{4V_{CC} I_{CQ}}{8} = \frac{V_{CC} I_{CQ}}{2}. \end{aligned}$$

$\therefore$  The efficiency of power amplifier is

$$\eta = \frac{P_{a.c.}}{P_{d.c.}} = \frac{\frac{V_{CC} I_{CQ}}{2}}{V_{CC} I_{CQ}} = \frac{1}{2}$$

$$\therefore \eta = 0.5$$

$\therefore$  The efficiency of transformer coupled class A amplifier is 50% in ideal case.

## COMPARISON OF AMPLIFIER CONFIGURATIONS

The choice of configuration and the type of transistor used depends on the desired input and output impedances, current and voltage gains and the frequency response. The following table gives a comparison of small signal amplifiers in the three possible configurations.

Parameter	Common base	Common emitter	Common collector
Voltage gain	High	Very-high	nearly 1
Current gain	nearly 1	high	highest
Input impedance	lowest	moderate	highest
Output impedance	highest	moderate	lowest
Phase reversal	no	yes	no

**Class B Amplifier.** Class B amplifier is biased at cutoff so that it operates in the linear region for  $180^\circ$  of the input cycle and is in cutoff for  $180^\circ$ .

Fig. 36.15 shows class B amplifier operation (noninverting).

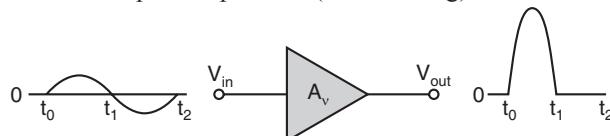


Fig. 36.15

The output waveform is shown relative to the input in terms of time ( $t$ ).

The circuit only conducts for the positive half of the cycle.

To amplify the entire cycle, it is necessary to add a second class B amplifier that operates on the negative half of the cycle. The combination of two class B amplifiers working together is called **push-pull** operation.

We shall discuss a class B push-pull amplifier. It uses two *complementary symmetry* BJTs to reproduce the entire waveform.

### 36.5 CLASS B PUSH-PULL AMPLIFIER

Fig. 36.16 shows the circuit arrangement of a class *B* push-pull amplifier.

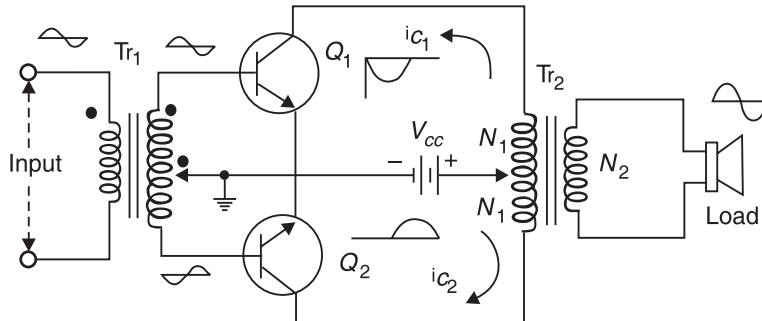


Fig. 36.16

It consists of two identical NPN transistors  $Q_1$  and  $Q_2$  placed in parallel (back to back) and operated as class *B* amplifiers. The input transformer  $Tr_1$  has a centre tapped secondary winding which provides opposite polarity signals to the two transistors. So the input signal drives the two transistors on alternate half cycles. The collector terminals of the transistors are connected across the primary of the output transformer  $Tr_2$ . The *dc* supply voltage  $V_{cc}$  is applied at the centre tap of this primary. The load resistance (loudspeaker) is connected across the secondary of the output transformer. When no signal is applied, both the transistors are cut-off.

#### Working :

(i) When the input signal is applied, the phase splitter transformer  $Tr_1$  produces two signals which are  $180^\circ$  out of phase with each other. The transistors  $Q_1$  and  $Q_2$  are driven by these two signals.

(ii) During positive half cycle of the signal, transistor  $Q_1$  conducts because its base is driven positive. Now collector current  $i_{c1}$  flows as shown in Fig. 36.16. Transistor  $Q_2$  does not conduct because its base has negative voltage. Thus  $i_{c2}$  is zero. In this way one positive half cycle of output signal appears across the primary of  $Tr_2$ .

(iii) During the negative half cycle of the input signal,  $Q_2$  is forward biased and allows a current  $i_{c2}$  to flow. Transistor  $Q_1$  becomes nonconducting.

Thus only one transistor conducts at a time.

The output transformer serves to join the two currents producing an almost undistorted output waveform as shown in Fig. 36.17.

(i) Fig. 36.17 (a) shows the input signal  $v_i$ .

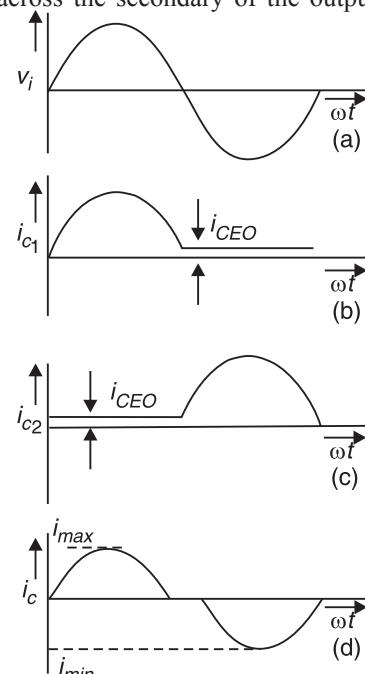


Fig. 36.17

- (ii) The collector currents in  $Q_1$  and  $Q_2$  transistors are represented by Figs. (b) and (c).  
 (iii) The current responsible for output voltage across the load is represented by Fig. (d).

#### Cross-Over Distortion :

When the d.c. base voltage is zero, both transistors are off. In order that a transistor begins to conduct, the input signal voltage must have exceeded  $V_{BE}$  (Fig. 36.18). The transistor remains cut off until the base voltage exceeds 0.3 V for Ge and 0.7 V for Si.

Because of this there will be a time interval between the positive and negative alternations of the input, when neither transistor is conducting. The output does not follow the input around the zero input voltage condition. Thus the output collector current is not a uniformly enlarged sine wave at low values of current. The resulting distortion in the output waveform is called crossover distortion.

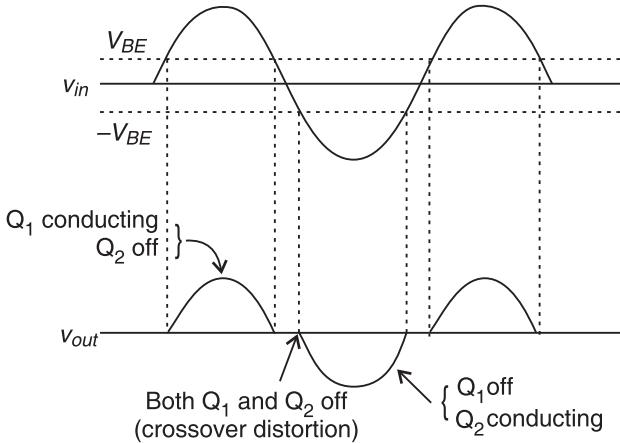


Fig. 36.18

#### Power Efficiency of Class B Push-Pull Amplifier

The current in each transistor is the average value of half sine loop.  
 For half sine loop,  $I_{dc}$  is given by

$$\begin{aligned} I_{dc} &= \frac{(I_c)_{\max}}{\pi} \\ \therefore (P_{in})_{dc} &= 2 \times \left[ \frac{(I_c)_{\max}}{\pi} \times V_{CC} \right] \end{aligned} \quad \dots(1)$$

Here factor 2 is introduced as there are two transistors in push-pull amplifier.

R.M.S. value of collector current =  $(I_c)_{\max} / \sqrt{2}$ .

R.M.S. value of output voltage =  $V_{CC} / \sqrt{2}$ .

(Under ideal conditions of maximum power)

$$\therefore (P_0)_{ac} = \frac{(I_c)_{\max}}{\sqrt{2}} \times \frac{V_{CC}}{\sqrt{2}} = \frac{(I_c)_{\max} \times V_{CC}}{2} \quad \dots(2)$$

Now overall maximum efficiency

$$\begin{aligned} \eta_{\text{overall}} &= \frac{(P_0)_{\text{a.c.}}}{(P_{\text{in}})_{\text{d.c.}}} \\ &= \frac{(I_c)_{\max} \times V_{CC}}{2} \times \frac{\pi}{2(I_c)_{\max} \times V_{CC}} = \frac{\pi}{4} = 0.785 = 78.5\% \end{aligned}$$

**Advantages.** (i) The output has much less distortion due to cancellation of even harmonics.

(ii) The maximum efficiency of class B push pull amplifier is quite high ( $\approx 79\%$ ).

**EXAMPLE 1.** Prove that a push-pull circuit balances all the even harmonics in the output and leaves the third harmonic term as the principal source of distortion.

(Rohilkhand, 1999, Kumaun, 2000)

**SOLUTION.** A power amplifier has to handle large signal inputs. Hence there will be harmonic distortions (frequency components in the signal getting modified) due to non-linear characteristics of the transistors used in the amplifier.

Let the input signal be sinusoidal. The base currents in the two transistors will be opposite in phase. Hence the collector currents will also be opposite in phase.

The collector currents in transistors  $Q_1$  and  $Q_2$  are given by

$$i_{c_1} = A_0 + A_1 \sin \omega t + A_2 \sin 2\omega t + A_3 \sin 3\omega t + \dots, \quad \dots(1)$$

and

$$i_{c_2} = A_0 + A_1 \sin (\omega t + \pi) + A_2 \sin 2(\omega t + \pi) + A_3 \sin 3(\omega t + \pi) \quad \dots(2)$$

or

$$i_{c_2} = A_0 - A_1 \sin \omega t + A_2 \sin 2\omega t - A_3 \sin 3\omega t + \dots \quad \dots(2)$$

The total current through the primary of output transformer is

$$i = i_{c_1} - i_{c_2}$$

$$\therefore i = 2[A_1 \sin \omega t + A_3 \sin 3\omega t + \dots] \quad \dots(3)$$

The voltage induced in the secondary of the output transformer is proportional to  $i$ .

$$\therefore v_0 \propto i$$

$v_0 = ki$  where  $k$  is the constant of proportionality.

$$\therefore v_0 = 2k[A_1 \sin \omega t + A_3 \sin 3\omega t + \dots] \quad \dots(4)$$

Thus, the output is free from even harmonics.

The principal source of distortion is the third harmonic  $3\omega$ .

## MULTISTAGE AMPLIFIERS

**Multistage Amplifiers.** Single-stage amplifiers are connected in sequence with various coupling methods to form multistage amplifiers.

- The output of one amplifier drives the input of the next.
- The output of the first stage is used as the input to a second stage.
- The output of the second stage is used as the input to the third stage, and so on.
- The basic purpose of a multistage arrangement is to *increase* the overall *voltage gain*.

The *total voltage gain* of a multistage amplifier is the *product of the individual gains (sum of dB gains)*.

### 36.6 MULTISTAGE VOLTAGE GAIN

Fig. 36.19 shows a multistage amplifier.

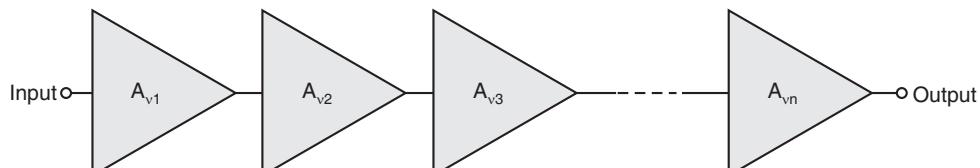


Fig. 36.19

Each triangular symbol represents a separate amplifier. The overall voltage gain,  $A_v$ , of a multistage amplifier is the product of the individual voltage gains.

$$A_v = A_{v1}A_{v2}A_{v3} \dots A_{vn}$$

Here,  $n$  is the number of **stages**.

### Voltage Gain Expressed in Decibels

Voltage gain is the output voltage divided by the input voltage:

$$A_v = \frac{V_{\text{out}}}{V_{\text{in}}}.$$

**Decibel voltage gain** is defined as:

$$A_{v(\text{dB})} = 20 \log A_v \quad \dots(1)$$

If an amplifier has a voltage gain of 100,000, it has a decibel voltage gain of:

$$A_{v(\text{dB})} = 20 \log 100,000 = 100 \text{ dB.}$$

Consider a two-stage amplifier (Fig. 36.20).

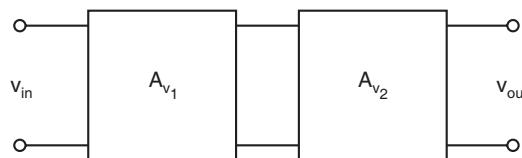


Fig. 36.20

The total voltage gain is the product of the individual voltage gains:

$$A_v = (A_{v1})(A_{v2}) \quad \dots(2)$$

Total decibel voltage gain is

$$A_{v(\text{dB})} = 20 \log A_v = 20 \log (A_{v1})(A_{v2}) = 20 \log A_{v1} + 20 \log A_{v2}$$

This can be written as:

$$A_{v(\text{dB})} = A_{v1(\text{dB})} + A_{v2(\text{dB})} \quad \dots(3)$$

This equation says that the total decibel voltage gain of two cascaded stages equals the *sum* of the individual decibel voltage gains. The same idea applies to any number of stages.

**EXAMPLE 1.** A certain cascaded amplifier arrangement has the following voltage gains:  $A_{v1} = 10$ ,  $A_{v2} = 15$ , and  $A_{v3} = 20$ . What is the overall voltage gain? Also express each gain in decibels (dB) and determine the total voltage gain in dB.

**SOLUTION.**

$$A'_v = A_{v1}A_{v2}A_{v3} = (10)(15)(20) = 3000$$

$$A_{v1(\text{dB})} = 20 \log 10 = 20.0 \text{ dB}$$

$$A_{v2(\text{dB})} = 20 \log 15 = 23.5 \text{ dB}$$

$$A_{v3(\text{dB})} = 20 \log 20 = 26.0 \text{ dB}$$

$$A'_{v(\text{dB})} = 20.0 \text{ dB} + 23.5 \text{ dB} + 26.0 \text{ dB} = 69.5 \text{ dB}$$

## 36.7 DIRECT-COUPLED MULTISTAGE AMPLIFIER

**Circuit Diagram.** Fig. 36.21 shows a two-stage direct-coupled amplifier.

There are no coupling or bypass capacitors.

**Working.** The dc collector voltage of the first stage provides the base-bias voltage for the second stage.

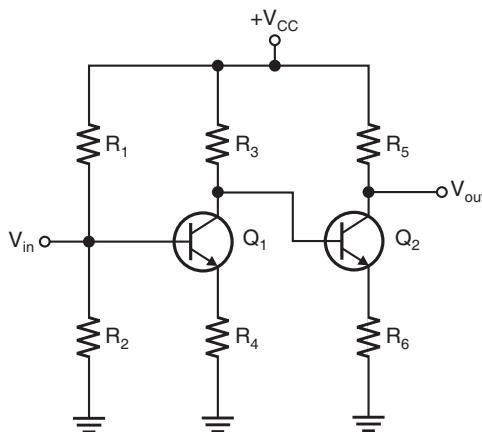


Fig. 36.21

When a weak signal is applied to the input of the first transistor  $Q_1$ , the amplified output appears at the collector. This amplified output is given as the input at the base of second transistor  $Q_2$ . This is further amplified by  $Q_2$ .

- Fig. 36.22 shows the frequency response of dc amplifier.

**Advantage.** It can be used to amplify low frequencies all the way down to dc (0 Hz) without loss of voltage gain because there are no capacitive reactances in the circuit.

#### Disadvantages

1. It cannot be used for amplifying high frequencies.
2. Due to temperature variation, the operating point is shifted.
- Normally in integrated circuits, most of the operational amplifiers are direct coupled.

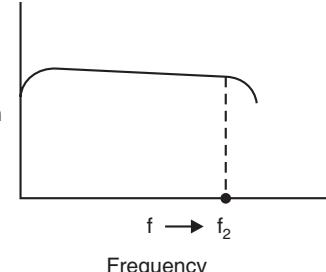


Fig. 36.22

## 36.8 RESISTANCE-CAPACITANCE (R-C) COUPLED AMPLIFIER

Fig. 36.23 shows the circuit arrangement of a two stage  $RC$  coupled  $CE$  mode transistor amplifier.

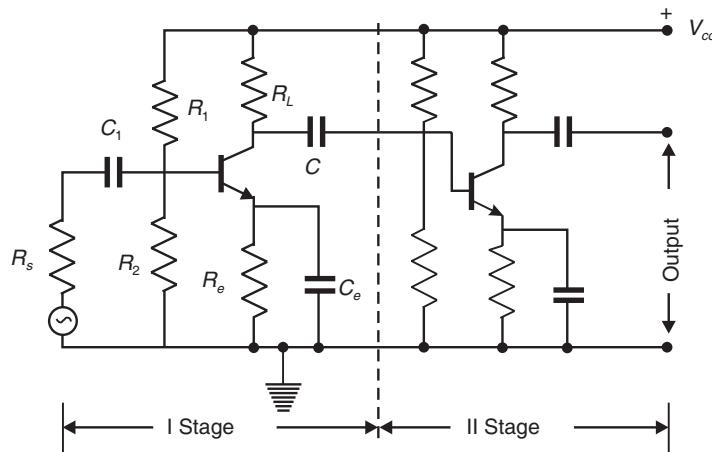


Fig. 36.23

### Function of Circuit Components

(i) **Biassing and stabilization network.** The resistances  $R_1$ ,  $R_2$  and  $R_e$  form the biassing and stabilization network. The biassing circuit establishes a proper operating point. Otherwise, a part of the negative half cycle of the signal  $e_s$  may be cut off in the output.  $R_e$  also reduces the variation of collector current with temperature.

(ii) **Input capacitor  $C_1$ .** Electrolytic capacitor  $C_1$  is used to couple the a.c. signal voltage to the base of the transistor. If it is not used, the resistance of the signal source will be in parallel to  $R_2$  and the bias voltage of the base will change.

(iii) **Emitter bypass capacitor  $C_e$ .** The capacitor  $C_e$  connected across the emitter resistor  $R_e$  is of large value. Its reactance  $1/\omega C_e$  is small. It, therefore, offers a low reactance path to the alternating component of emitter current and thus bypasses resistor  $R_e$  at audio frequencies. Consequently, the potential difference across  $R_e$  is due to the d.c. component of the current only. If it is not used, the amplified a.c. signal flowing through  $R_e$  will cause a voltage drop across it, thereby reducing the output voltage.

(iv) **Coupling capacitor  $C$ .** The coupling capacitor  $C$  transmits a.c. signals but blocks d.c. It couples the output of the first stage amplifier to the input of the second stage. It blocks the d.c. voltage of the first stage from reaching the base of the second stage. Thus the d.c. biasing of the next stage is not disturbed. That is why, the capacitor  $C$  is also called a *blocking capacitor*.

**Operation.** When an a.c. signal is applied to the base-emitter circuit of first transistor, it appears in the amplified form across the load resistor  $R_L$ . The amplified signal is applied between base and emitter circuit of second stage through coupling capacitor  $C$ . The function of coupling capacitor is to allow the passage of amplified a.c. signal and block the passage of d.c. currents and voltages from I stage to the input of the II stage. The second stage does further amplification of the signal. Thus the amplifiers connected in cascade arrangement amplify the signal and the overall gain is considerably increased.

### Frequency Response Curve

The voltage gain of the amplifier depends on the frequency of a.c. signal applied. The curve representing the variation of gain of an amplifier with frequency is called frequency response curve. The frequency response curve of RC coupled amplifier is shown in Fig. 36.24.

It is divided into three regions.

The voltage gain is uniform in the midfrequency range.

The voltage gain falls off in both the low frequency and high frequency ranges.

This behaviour is explained below :

(i) **Low Frequency Range.** At low frequencies ( $< 50$  Hz), the *gain decreases with decreasing frequency due to coupling capacitor  $C$* . As the frequency gets lower and lower, the reactance ( $1/\omega C$ ) of the coupling capacitor gets higher and higher. Consequently, an increasing portion of the voltage drop takes place across  $C$ . Thus the voltage reaching from one stage to the next stage decreases. Hence the voltage gain drops.

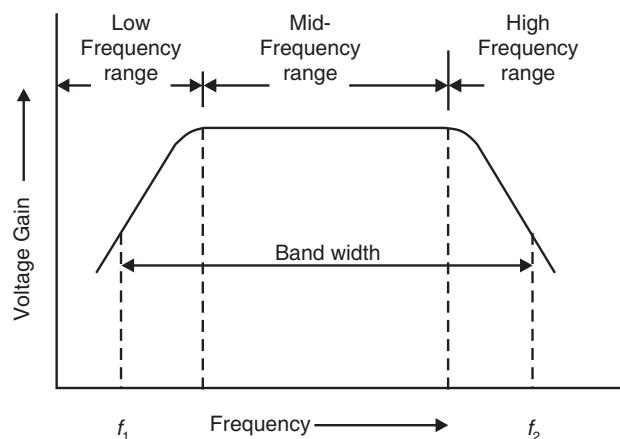


Fig. 36.24

The other component due to which the gain decreases at low frequencies is the bypass capacitor  $C_e$ . As frequency decreases, the reactance of capacitor  $C_e$  increases and becomes comparable to emitter resistance  $R_e$ . Therefore, it cannot shunt the resistor  $R_e$  effectively.

Because of these two reasons, the voltage gain decreases in the low frequency region.

**Lower cut off frequency  $f_1$ .** It is defined as the frequency at which the magnitude of the voltage gain in the low frequency range falls to  $1/\sqrt{2}$  of the mid-frequency gain.

**(ii) Mid Frequency Range.** In the middle frequency range (50 Hz to 20 kHz), the voltage gain of the amplifier is uniform because resistor values are independent of frequency changes. As the frequency increases in this range, reactance of coupling capacitor decreases which tends to increase the gain. But it simultaneously increases the loading effect of the first stage which decreases the gain. These two factors cancel each other. Hence the gain at mid frequencies remains constant.

**(iii) High Frequency Range.** In the high frequency range ( $> 20$  kHz), the voltage gain falls due to the capacitances that appear in shunt across the output. These capacitances consist of the collector capacitance of the transistor and the stray wiring capacitances. With increase in frequency, the reactance of these capacitances decreases and an increased part of the output is bypassed to earth. Consequently the output voltage of the amplifier decreases. Hence the voltage gain falls at high frequencies.

Moreover, at high frequencies, capacitive reactance of base-emitter junction is reduced. As a result, the base current increases. Hence the current amplification factor  $\beta$  decreases. This has the effect of further reducing the voltage gain.

**Upper cut-off frequency  $f_2$ .** It is defined as the frequency at which the magnitude of the high frequency gain is  $1/\sqrt{2}$  times the magnitude of the mid frequency gain.

**Bandwidth.** The frequency difference  $f_2 - f_1$  gives a measure of the bandwidth of the amplifier.

#### Expression for the voltage gain in mid-frequency range

We make the following simplifying assumptions :

1. The transistors used in the two stages are identical. The input resistance  $h_{ie}$  and forward current transfer ratio  $h_{fe}$  of each is the same.
2. The output admittance  $h_{oe}$  and the reverse voltage gain  $h_{re}$  are very small. Therefore they are neglected.
3. The bias resistors  $R_1, R_2$  are of sufficiently large value so that they do not affect the a.c. operation of the circuit.
4. The reactance of the emitter bypass capacitor is negligibly small.

In middle frequency range, the reactance of the coupling capacitor  $C$  is negligibly small. Fig. 36.25 shows the simplified hybrid equivalent circuit of the first stage of  $RC$  coupled amplifier.

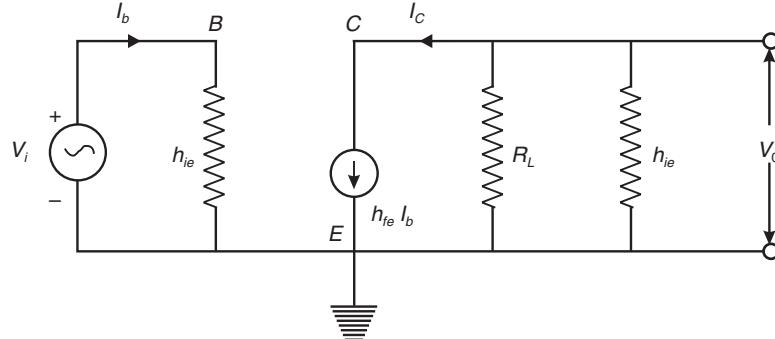


Fig. 36.25

The effective load resistance  $Z_L$  is the parallel combination of load resistance  $R_L$  and the input resistance  $h_{ie}$  of the second stage. Thus

$$\frac{1}{Z_L} = \frac{1}{R_L} + \frac{1}{h_{ie}}$$

or

$$Z_L = \frac{h_{ie}R_L}{h_{ie} + R_L} \quad \dots(1)$$

Therefore, the output voltage  $V_0$  across  $Z_L$  is

$$V_0 = -h_{fe}I_b Z_L \quad \dots(2)$$

The input voltage  $V_i$  is

$$V_i = h_{ie}I_b \quad \dots(3)$$

Therefore, the voltage gain is given by

$$A_{vm} = \frac{V_0}{V_i} = -\frac{h_{fe}I_b Z_L}{h_{ie}I_b} = -\frac{h_{fe}Z_L}{h_{ie}}.$$

Substituting the value of  $Z_L$  from Eq. (1), we get,

$$A_{vm} = -\frac{h_{fe}}{h_{ie}} \left( \frac{h_{ie}R_L}{h_{ie} + R_L} \right) = -\frac{h_{fe}R_L}{h_{ie} + R_L} \quad \dots(4)$$

The negative sign in Eq. (4) shows that there is a phase shift of  $180^\circ$  between the output and input signal voltages.

It is evident from Eq. (4) that the voltage gain of  $RC$  coupled common-emitter amplifier in mid-frequency range is independent of frequency.

- The **total voltage gain** of the amplifier,  $A_V = (A_{V_1})(A_{V_2})$ .

#### Advantages

1. It has very flat *frequency versus gain curve*. So it has excellent frequency response. The gain is constant over the audio frequency range which is the region of most importance for speech, music etc. It provides excellent audio fidelity over a wide range of frequency.
2. It is small, light and cheap.
3. The overall amplification of *R.C.* coupling is higher than that of other forms of coupling.
4. There are no coils or transformers which can pick up undesirable signals. Therefore, it has minimum possible non-linear distortion. There are no magnetic fields to interfere with its working and signal.

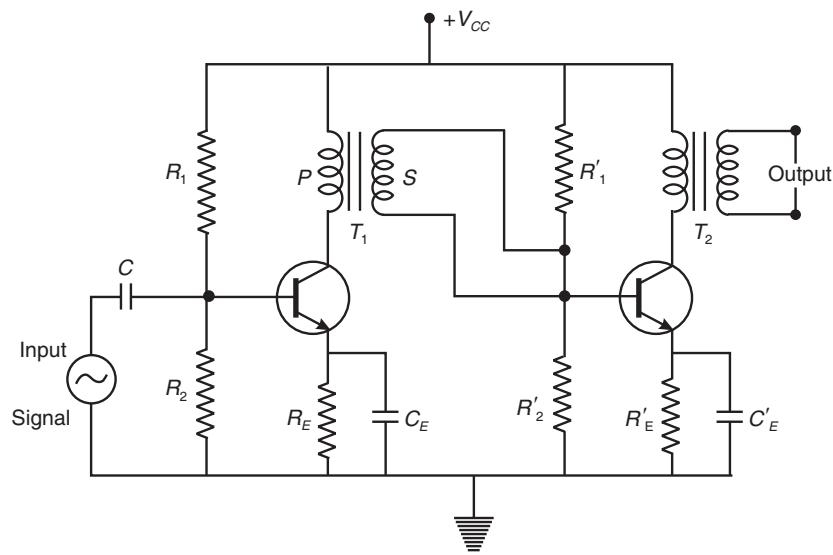
#### Disadvantages

1. It has low voltage and power gain. It is because the low resistance presented by the input of each stage to the preceding stage decreases the effective *load resistance* ( $R_{AC}$ ) and hence the gain.
  2. It becomes noisy with age, particularly in moist climates.
  3. Impedance matching is poor.
- The output impedance of  $RC$  coupled amplifier is high (several hundred ohms). But the input impedance of a loud speaker is low (a few ohms). Hence, little power is transferred to the speaker.

**Applications:** It has excellent audio fidelity over a wide range of frequency. Therefore, they are widely used as voltage amplifiers *e.g.*, in the initial stages of public address system. If other type of coupling (*e.g.*, transformer coupling) is employed in the initial stages, this results in frequency distortion which may be amplified in next stages. However, because of poor impedance matching,  $RC$  coupling is rarely used in the final stages.

### 36.9 TRANSFORMER COUPLED (TC) AMPLIFIER

Fig. 36.26 shows the circuit of a two stage transformer coupled amplifier.  $T_1$  is the coupling transformer. It is used to couple the two stages.  $T_2$  is the output transformer. It is used to transfer the output of the second stage to the load. The primary  $P$  of the coupling transformer  $T_1$  serves as the collector load. Its secondary  $S$  gives output to the next stage.



**Fig. 36.26**

The resistors  $R_1$ ,  $R_2$  as well as  $R'_1$  and  $R'_2$  form voltage-divider circuits. The resistors  $R_E$  and  $R'_E$  are the emitter stabilisers. The capacitors  $C_E$  and  $C'_E$  are the bypass capacitors.  $C$  is the input coupling capacitor.

#### Operation

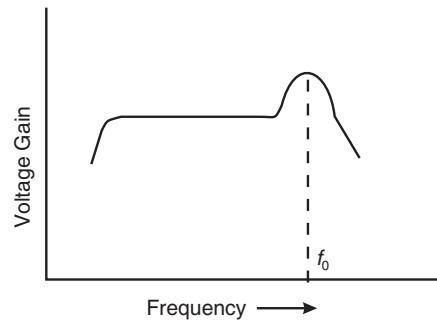
The signal is applied to the base of the first transistor. It appears in an amplified form across the primary  $P$  of the coupling transformer  $T_1$ . This amplified voltage is transferred to the secondary of the transformer by magnetic induction. From there, it is applied to the base of second transistor. Thus the amplified output of first stage becomes the input of the second stage.

The second transistor further amplifies the input. The amplified signal appears across the primary of the output transformer  $T_2$ . The secondary of this transformer transfers this output to the load.

#### Frequency Response

Fig. 36.27 shows the frequency response curve of a transformer coupled amplifier.

- (i) The frequency response is poor *i.e.*, the gain is constant only over a small range of frequency.
- (ii) There is a decrease in gain at low frequencies.
- (iii) At high frequencies, the distributed capacitance between different turns of windings acts as a bypass capacitor. The output voltage decreases. Hence the gain decreases.



**Fig. 36.27**

There is a peak or exaggerated gain at resonant frequency  $f_0$  of the tuned circuit which is formed by inductance and distributed winding capacitance in the circuit.

Thus all frequencies in a composite signal, such as music, speech etc., are not equally amplified. Hence transformer coupled amplifier shows *frequency distortion*.

### Advantages

1. The operation of a transformer coupled amplifier is more efficient than an RC coupled system because of the low d.c. resistance of the primary which is connected to the collector circuit.
2. By suitably adjusting the turns ratio of the transformer, very good impedance matching can be obtained. Hence very good overall gain can be obtained.
3. No signal power is lost in the collector or base resistors.

### Disadvantages

1. The frequency response is poor.
2. The coupling and output transformers are bulky and costly.
3. At radio frequencies, the inductance and interwinding capacitance in the transformer create problems.
4. The transformer windings tend to produce a *hum* in the output.

### Applications

A transformer coupled amplifier is mostly employed for *impedance matching* and *power amplification*.

## 36.10 DISTORTION IN AMPLIFIERS

An ideal amplifier must produce an amplified output voltage which must be an enlarged true copy of the input voltage, without any change in the wave-shape. Practical amplifiers, however, create some change in the wave-shape of the output. This is known as 'distortion'.

There are three types of distortion.

(i) **Amplitude (or Non-linear) Distortion.** The difference in the magnitudes of positive peak and negative peak of the amplified voltage is called amplitude distortion. Amplitude distorted waveform is shown in Fig. 36.28.

In this type of distortion, the output of the amplifier is not linearly related to the input. So the distortion is also called non-linear distortion. In large signal amplifiers, new frequencies are present in the output which are *not* present in the input signal. So the distortion is called harmonic distortion. Harmonic distortion can be reduced by using negative feedback circuits.

(ii) **Frequency Distortion.** We know that the frequency response of an amplifier is not uniform. Whereas the voltage-gain is sufficiently uniform over the middle-frequency range, it sharply falls at low as well as at high frequencies. Therefore, if the input signal is a complex wave, certain frequency components are amplified more than others. The output wave is not a faithful reproduction of the signal. The distortion so produced is called 'frequency distortion'.

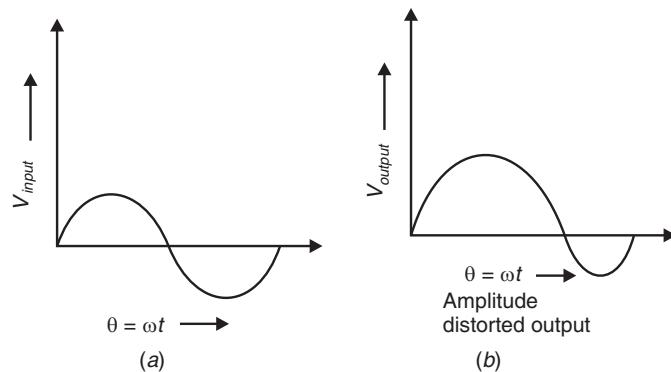


Fig. 36.28

**(iii) Phase Distortion.** In a practical amplifier, different frequencies may suffer different phase-changes. Therefore, relative phase of some components of a complex input wave may be shifted. This makes the shape of the output wave different from that of the signal. Special phase equalising circuits are employed to avoid phase distortion.

## FEEDBACK AMPLIFIERS

### 36.11 FEEDBACK PRINCIPLE

In an amplifier, a small part of the output signal may be fed to the input circuit by suitable means. This is known as feedback. Feedback is a process in which there is a transfer of energy from output of an electronic system to its input. The amount of feedback is determined by the feedback network.

**(a) Positive feedback amplifier :** When the feedback energy (Voltage or current) is in phase with the input signal and thus aids it, it is called *positive feedback*. When the input and feedback signals are in same phase, the magnitude of the input signal enhances as an effect of the feedback signal. The amplifier based on positive feedback is called *positive feedback amplifier*. They are generally used as oscillators.

**(b) Negative feedback amplifier :** When the feedback energy (voltage or current) is out of phase with the input signal and thus opposes it, it is called *negative feedback*. When the input and feedback signals are completely out of phase, the magnitude of the input signal diminishes by the superimposition of feedback signal. The amplifier based on negative feedback is called negative feedback amplifier.

The feedback amplifiers are classified into two general classes : voltage feedback and current feedback. In voltage feedback, the voltage fed back to the input terminals is proportional to the output voltage. In current feedback, the voltage fed back to the input terminals is proportional to the current through the load.

The amount of amplifier gain with feedback is called the *closed loop gain*. Without feedback, the gain is called *open loop gain*.

#### 36.11.1. The Effect of Feedback on the Gain of an Amplifier

Fig. 36.29 shows the block diagram of an amplifier with feedback loop. The feedback circuit usually consists of resistors. A fraction  $\beta$  of the output voltage is fed back into the input circuit through the feedback circuit. Let  $A$  be the gain of the amplifier *i.e.*, the ratio of output voltage  $V_0$  to the input voltage  $V_i$

$$A = \frac{V_0}{V_i}$$

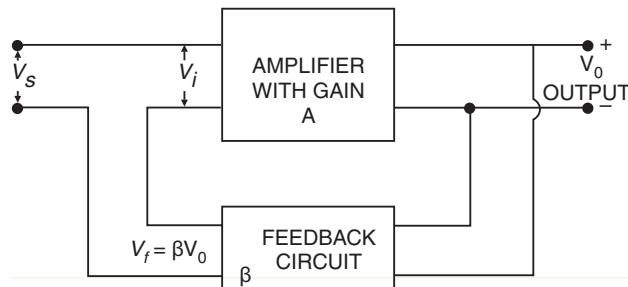


Fig. 36.29

This is called the open-loop gain.

This is the gain of the amplifier without feedback.

The feedback network extracts a voltage  $V_f = \beta V_0$  from output  $V_0$  of the amplifier. This voltage is added (positive feedback) or subtracted (negative feedback) from the signal voltage  $V_S$ . Now

$$V_i = V_S + V_f = V_S + \beta V_0 \quad (\text{Positive feedback})$$

$$V_i = V_S - V_f = V_S - \beta V_0 \text{ (Negative feedback)}$$

The quantity  $\beta = V_f/V_0$  is called feedback ratio or feedback fraction.

Gain of the amplifier without feedback,  $A = V_0/V_i$ .

or

$$V_0 = AV_i \quad \dots(1)$$

Let us assume that

$$V_i = V_S + \beta V_0 \text{ (positive feedback)} \quad \dots(2)$$

From Eqs. (1) and (2), we get

$$V_0 = A[V_S + \beta V_0] = AV_S + A\beta V_0$$

or

$$V_0[1 - A\beta] = AV_S$$

∴

$$\frac{V_0}{V_S} = \frac{A}{1 - A\beta} \quad \dots(3)$$

Let  $A_f$  be the overall gain (gain with feedback) of the amplifier. This is defined as the ratio of output voltage  $V_0$  to the applied signal voltage  $V_S$ .

The resultant gain of the amplifier with feedback

$$A_f = \frac{\text{Output voltage}}{\text{Input signal voltage}} = \frac{V_0}{V_S} \quad \dots(4)$$

$A_f$  is called the *closed-loop gain*.

From Eqs. (3) and (4)

$$A_f = \frac{A}{1 - A\beta} \quad \text{positive feedback} \quad \dots(5)$$

$$A_f = \frac{A}{1 + A\beta} \quad \text{negative feedback} \quad \dots(6)$$

(i) For positive feedback,  $A_f > A$ .

The positive feedback thus increases the gain of the amplifier.

(ii) For negative feedback,  $A_f < A$ .

The negative feedback thus decreases the gain of the amplifier.

(iii) In the case of positive feedback, if  $\beta A = 1$ , then  $A_f \rightarrow \infty$ .

This means that the amplifier becomes an oscillator.

Hence, the conditions for the amplifier to be an oscillator are

(i) *the feedback must be positive* and

(ii) *the feedback factor must be unity i.e.,  $\beta A = 1$* .

### 36.11.2. Advantages of Negative Feedback

(i) **Gain Stability :** The gain of the amplifier employing negative feedback is given by

$$A_f = \frac{A}{1 + \beta A} \quad \dots(1)$$

If  $\beta A \gg 1$ , then above equation approximates to,

$$A_f = \frac{A}{1 + \beta A} \approx \frac{1}{\beta} \quad \dots(2)$$

The term  $\beta$  is the feedback ratio and it is a function of the feedback network only. It does not depend upon the amplifier parameter changes or supply voltage variations. Thus the negative feedback improves stability of gain of the amplifier.

**(ii) Reduction of distortion :** The distortion in the output is reduced, because there is partial cancellation of out-of-phase signals. Both amplitude distortion and frequency distortion are reduced.

**(iii) Reduction of output noise :** The output noises and hum in an amplifier are reduced by negative feedback.

**(iv) Increase in input impedance :** Negative feedback increases the input impedance of an amplifier, which will not then draw much current from the source.

**(v) Decrease in output impedance :** Application of negative series voltage feedback in an amplifier *reduces the output impedance* which is a desirable characteristic for an amplifier because an amplifier with low output impedance can deliver power (or voltage) to the load without much loss.

**(vi) Increase in band width :** Negative feedback decreases the lower cut-off frequency but increases the upper cut-off frequency. Thus negative feedback increases the band width. But this increase in band width is achieved at the cost of reduction in the gain of the amplifier as shown in Fig. 36.30.

If  $A$  is the gain of amplifier without feedback and  $A_f$  with negative feedback, then since gain-band width product remains the same

$$A \times BW = A_f \times BW'$$

or  $A \times (f_2 - f_1) = A_f \times (f'_2 - f'_1)$

### 36.11.3. Practical Feedback Circuits

#### 1. Current Negative Feedback Transistor Amplifier

In these circuits, the voltage feedback to the input of the amplifier is proportional to the output current.

##### Circuit Details

Consider the circuit shown in Fig. 36.31. It is an  $R - C$  coupled amplifier without emitter bypass capacitor across emitter resistance  $R_E$ .  $R_1$  and  $R_2$  provide the base bias.

##### Circuit Operation

When a signal is applied at the input, there will be change in emitter current. The a.c. emitter current  $i_e$  produces a voltage across the emitter resistance  $R_E$  and this voltage  $= i_e R_E$  with emitter becoming more positive during the increase of the signal voltage. Effectively, the value of  $v_{be}$  is decreased *i.e.*, the voltage developed across  $R_E$  is in series with and opposing the applied signal,  $v_s$ . This is negative feedback and it is proportional to the current through the emitter resistance. Thus the circuit works as a *negative current feedback* amplifier.

##### Expression for closed loop gain

The voltage  $i_e R_E$  across emitter resistance opposes the signal voltage  $v_s$  as it is in opposite direction to the signal.

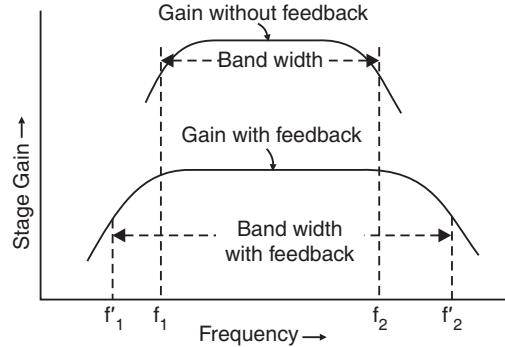


Fig. 36.30

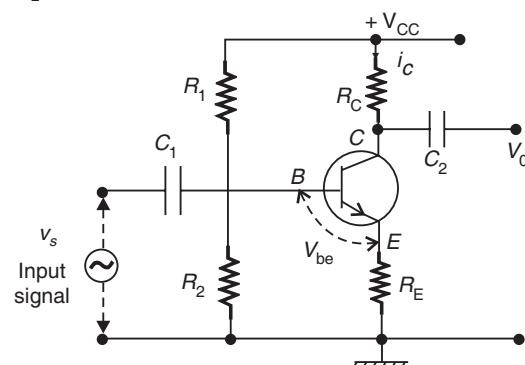


Fig. 36.31

The net input to the amplifier after negative feedback is

$$\begin{aligned} v_{be} &= v_s - i_e R_E \\ v_{be} &= v_s - i_c R_E \text{ since } i_e = i_c \end{aligned} \quad \dots(1)$$

Output voltage  $v_0$  is given by

$$\begin{aligned} v_0 &= -i_c R_c \\ \therefore i_c &= -v_0 / R_c \end{aligned} \quad \dots(2)$$

Putting this value in Eq. (1), we get

$$v_{be} = v_s + v_0 \cdot \frac{R_E}{R_C} \quad \text{where } \beta = R_E/R_C \text{ is the feedback ratio.}$$

$$\frac{v_{be}}{v_0} = \frac{v_s}{v_0} + \beta$$

$$\frac{1}{A_f} = \frac{1}{A} + \beta$$

since the closed loop gain  $A_f = v_0/v_{be}$  and the open loop gain  $A = v_0/v_s$ ,

$$\therefore A_f = \frac{A}{1 + A\beta} \quad \dots(3)$$

But since  $\beta A \gg 1$ , the above expression approximates to

$$A_f \approx \frac{A}{\beta A} = \frac{1}{\beta} = \frac{R_C}{R_E} \quad \dots(4)$$

$\therefore$  The closed loop gain of the negative feedback amplifier is

$$A_f = \left( \frac{R_C}{R_E} \right)$$

### 36.12 Emitter Follower

Fig. 36.32 (a) shows the emitter follower circuit. It is an excellent example of an amplifier with negative feedback.

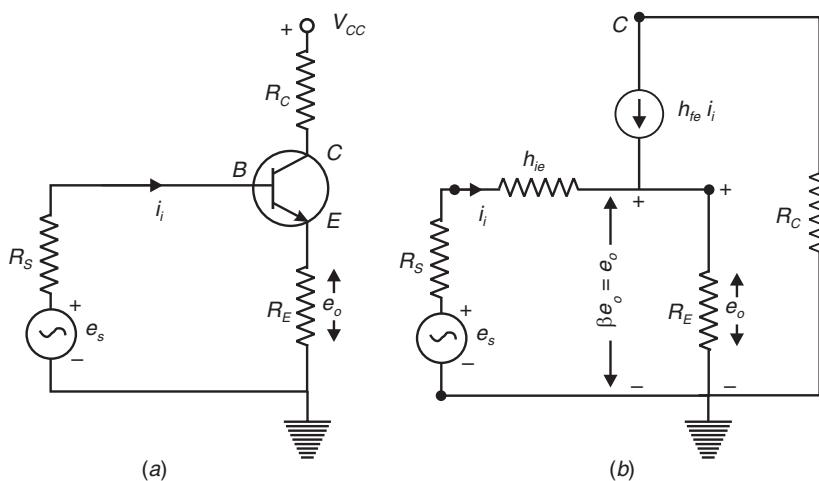


Fig. 36.32

The output voltage  $e_0$  is developed across emitter resistance  $R_E$ . When a signal  $e_s$  is applied at the input, the resulting a.c. emitter current  $i_e$  produces an output voltage  $i_e R_E$  across  $R_E$ . This voltage opposes the signal voltage as it is in opposite direction to the signal. This voltage is feedback at the input of the amplifier. The feedback voltage is proportional to the emitter current i.e., output current. So the circuit provides negative current feedback.

The output is taken from the emitter resistance  $R_E$  and the voltage variations across base-emitter junction (input) follow the emitter. So the circuit is named as an *emitter follower*. It is also called *common collector amplifier* because the transistor collector is at a.c. ground and common between input and output.

The equivalent circuit, using *CE* hybrid model, of the emitter follower is given in Fig. 36.32 (b).

**Voltage gain.** The voltage gain  $A$  of the amplifier, without feedback, may be obtained by connecting the grounded end of  $e_s$  to  $E$ . Then,

$$e_s = (R_S + h_{ie})i_i \quad \dots(1)$$

$$\text{Output voltage } e_0 = h_{fe}i_i R_E \quad \dots(2)$$

Voltage gain without feedback

$$A = \frac{e_0}{e_s} = \frac{h_{fe}R_E}{R_S + h_{ie}} \quad \dots(3)$$

Whole of the output voltage is fed to the input. So, feedback fraction  $\beta = 1$ .

$$\text{i.e., } \beta e_0 = e_0.$$

The voltage gain of the amplifier with negative feedback

$$A_f = \frac{A}{1 + \beta A} = \frac{A}{1 + A} \quad (\because \beta = 1) \dots(4)$$

Thus the voltage gain of the emitter follower is always less than unity.

Substituting the value of  $A$  from Eq. (3) in Eq. (4), we get

$$A_f = \frac{\frac{h_{fe}R_E}{R_S + h_{ie}}}{1 + \frac{\frac{h_{fe}R_E}{R_S + h_{ie}}}{\frac{h_{fe}R_E}{R_S + h_{ie}}}} = \frac{h_{fe}R_E}{R_S + h_{ie} + h_{fe}R_E} \quad \dots(5)$$

### Input Impedance

Put  $R_S = 0$  in Eq. (3). Then we get,

voltage gain without feedback

$$A = \frac{h_{fe}R_E}{h_{ie}} \quad \dots(6)$$

From Fig. 36.32 (b), input impedance of the amplifier without feedback,

$$Z_i = h_{ie} \quad \dots(7)$$

We know that input impedance of the amplifier with negative feedback,

$$Z_{if} = (1 + \beta A)Z_i$$

Substituting the value of  $A$  from Eq. (6) and of  $Z_i$  from Eq. (7) we get

$$Z_{if} = h_{ie} \left[ 1 + 1 \cdot \frac{h_{fe}R_E}{h_{ie}} \right] \quad (\because \beta = 1)$$

or

$$Z_{if} = h_{ie} + h_{fe}R_E \quad \dots(8)$$

Thus the input impedance of emitter follower is considerably increased by negative feedback.

### Output Impedance

The output impedance without feedback is the ratio of output voltage to output current.

Let

$e$  = open circuit output voltage

and  $i$  = short circuit output current i.e., by connecting the grounded end of  $e_s$  to  $E$  in Fig. 36.32 (b).

$$e = \lim_{R_E \rightarrow \infty} e_0 = \lim_{R_E \rightarrow \infty} \frac{h_{fe} R_E e_S}{R_S + h_{ie}} \quad [\text{using Eq. (3)}]$$

$$i = \frac{h_{fe} e_S}{R_S + h_{ie}}$$

$$\therefore Z_0 = \frac{e}{i} = \lim_{R_E \rightarrow \infty} R_E$$

Therefore, the output impedance without feedback is infinite.

Now, we know that output impedance of amplifier with negative feedback

$$\begin{aligned} Z_{of} &= \frac{Z_0}{1+\beta A} = \frac{\lim_{R_E \rightarrow \infty} R_E}{1+1 \times \frac{h_{fe} R_E}{R_S + h_{ie}}} \quad [\text{using Eq. (3)}] \\ &= \lim_{R_E \rightarrow \infty} \frac{(R_S + h_{fe}) R_E}{(R_S + h_{ie}) + h_{fe} R_E} \\ &= \lim_{R_E \rightarrow \infty} \frac{R_S + h_{fe}}{\frac{R_S + h_{ie}}{R_E} + h_{fe}} = \frac{R_S + h_{fe}}{h_{fe}}. \end{aligned}$$

Thus the output impedance of an emitter follower is considerably reduced by negative feedback.

### Emitter Follower (Alternative Analysis)

The **emitter follower** is also called a **common-collector (CC) amplifier**.

**Circuit.** Fig. 36.33 shows an emitter follower circuit with voltage-divider bias.

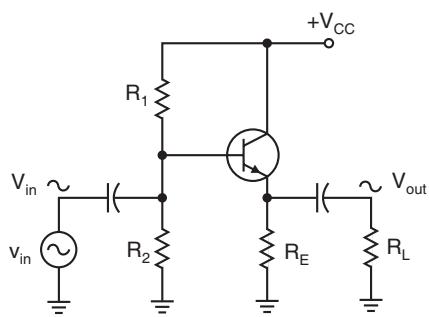


Fig. 36.33

- High current gain and power gain.
- No voltage gain.
- High input impedance and low output impedance.
- Input and output ac voltages are in phase.
- Ideal circuit for impedance matching.

- The input signal is coupled to the base. The output signal is taken from the emitter.
- The collector is at ac ground. So the circuit is a CC amplifier.
- There is no phase inversion. The output voltage is in phase and is approximately equal to the input voltage.
- The voltage gain of a CC amplifier is approximately 1.
- Its main advantages are its high input resistance and current gain.
- The circuit is called an *emitter follower* because the output voltage follows the input voltage.

- Since there is no collector resistor, the total voltage between the collector and ground equals the supply voltage.

**Working.** The input voltage is applied between base and emitter. This sets up an ac emitter current and produces an ac voltage across the emitter resistor. This ac voltage is then coupled to the load resistor. It is a negative current feedback circuit since the voltage feedback is proportional to the emitter current *i.e.*, output current. The negative feedback is massive because the feedback resistance equals all of the emitter resistance. As a result, the voltage gain is ultrastable, the distortion is almost nonexistent, and the input impedance of the base is very high.

The ac emitter resistance,

$$r_e = R_E \parallel R_L \quad \dots(1)$$

This is the external ac emitter resistance. It is different from the internal ac emitter resistance  $r'_e$ .

### Voltage Gain

Fig. 36.34 shows the ac equivalent circuit for emitter follower with the *T* model.

The emitter diode of a transistor acts like an ac resistance  $r'_e$ .

Output voltage,

$$v_{out} = i_e r_e \quad \dots(1)$$

Input voltage,

$$v_{in} = i_e (r_e + r'_e) \quad \dots(2)$$

By definition, the voltage gain is  $A_v = v_{out}/v_{in}$ .

Divide the first equation by the second, and you get the voltage gain of the emitter follower:

$\therefore$  Voltage gain of emitter follower is

$$A_V = \frac{r_e}{r_e + r'_e} \quad \dots(3)$$

Notice that the gain is always less than 1.

Usually, a designer makes  $r_e$  much greater than  $r'_e$ , so that the voltage gain equals 1 (approximately).

The emitter follower is not a voltage amplifier. It has a current gain of  $\beta$ .

The stages near the end of a system need to produce more current because the final load is usually a low impedance. The emitter follower can produce the large output currents needed by low-impedance loads. So emitter follower is a current or power amplifier.

### Input Impedance

Fig. 36.35 shows the ac equivalent circuit with the  $\pi$  model of the transistor.

The current gain transforms the total emitter resistance up by a factor of  $\beta$ .

The input impedance of the base is

$$z_{in(base)} = \beta(r_e + r'_e)$$

If  $r_e \gg r'_e$ , then the input impedance is approximately  $\beta r_e$ .

The input impedance of emitter follower is the combined effect of biasing resistors ( $R_1$  and  $R_2$ ) and the input impedance of transistor base [ $z_{in(base)}$ ]. These resistances are in parallel to the ac signal. So the input impedance  $z_{in}$  of the emitter follower is given by

$$z_{in} = R_1 \parallel R_2 \parallel z_{in(base)}$$

$$z_{in} = R_1 \parallel R_2 \parallel \beta(r_e + r'_e)$$

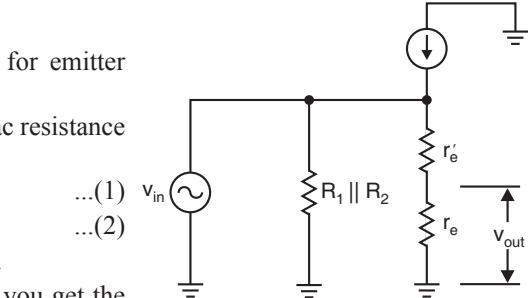


Fig. 36.34

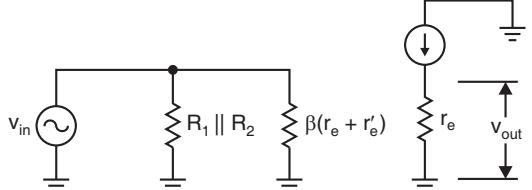


Fig. 36.35

- The emitter follower is characterized by a high input resistance; this is what makes it a useful circuit. Because of the high input resistance, it can be used as a buffer to minimize loading effects when a circuit is driving a low resistance load.

### Output Impedance

The output impedance of an amplifier is the same as its Thevenin impedance.

The output resistance is very low, making the emitter-follower useful for driving low resistance loads.

Maximum power transfer occurs when the load impedance is *matched* (made equal) to the source (Thevenin) impedance.

When maximum load power is wanted, a designer can match the load impedance to the output impedance of an emitter follower. For instance, the low impedance of a speaker can be matched to the output impedance of an emitter follower to deliver maximum power to the speaker.

### Darlington Pair

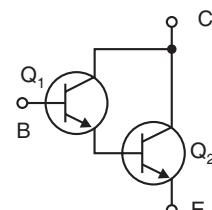
Fig. 36.36 shows a **Darlington pair**.

The Darlington Pair consists of two transistors coupled directly in such a way that emitter of  $Q_1$  is connected to the base of  $Q_2$ . This forms a single package and will act as a single transistor with three terminals. It has an overall current gain of:

$$\beta = \beta_1 \beta_2.$$

The overall current gain equals the product of the individual current gains.

Since its current gain is much higher, a Darlington connection can have a very high input impedance and can produce very large output currents. Darlington connections are often used with voltage regulators, power amplifiers, and high current switching applications.



**Fig. 36.36**

### EXERCISE

- With a neat circuit diagram explain the working of a CE-amplifier. *(Bangalore, 2002)*
- Give the *h*-parameter equivalent circuit of a CE-amplifier. Hence obtain expressions for (i) current gain (ii) voltage gain (iii) input impedance (iv) output impedance (v) power gain.
- Draw the circuit of a direct coupled class A transistor power amplifier and find an expression for the collector efficiency. Hence show that the maximum efficiency obtainable with this amplifier is only 25 percent.
- Draw the circuit of a transformer coupled class A power amplifier. Show that this amplifier is twice as efficient as the direct coupled amplifier.
- Sketch neatly the circuit diagram of a class B push-pull amplifier and calculate its maximum efficiency. *(Osmania University, 2004)*
- Describe the working of an *R-C* coupled amplifier and explain its frequency response curve. *(Rohilkhand, 88; Meerut 84, 86)*
- Give the circuit of *R-C* coupled amplifier. Obtain an expression for the voltage gain in mid-frequency range. *(Rohilkhand, 90; Nagpur, 80, 81, 82, 88)*
- Describe a transformer coupled amplifier. Discuss the frequency response curve.
- Discuss the principles of negative voltage feedback in amplifiers with a neat diagram. Derive an expression for the gain of negative voltage feedback amplifier. *(B.U. 2010)*
- Discuss the advantages of negative feedback. *(B.U. 2015)*
- Describe the action of emitter follower with a neat diagram. *(B.U. 2015)*
- Draw the diagram of emitter follower and its equivalent circuit. Derive expressions for its voltage gain, input impedance and output impedance. *(Nagarjuna, 1993; Andhra, 95)*

## OSCILLATORS

**After reading this chapter, you should be able to**

- ◆ Discuss the basic operating principles of an oscillator
- ◆ Explain the operation of LC sinusoidal oscillators (Hartley oscillator, Colpitt's oscillator)
- ◆ Analyze the basic operation of RC feedback oscillators (Wien bridge oscillator, phase shift oscillator)
- ◆ Explain how crystal-controlled oscillators work.
- ◆ Demonstrate how to design oscillator circuits for a given frequency
- ◆ Analyze square and triangular waveform generators

### 37.1 INTRODUCTION

An oscillator is an electronic device for generating alternating current of a desired frequency. An oscillator converts dc power from the dc source to ac power at the load. Oscillation is achieved through positive feedback in amplifiers.

The essential components of an *LC* oscillator are shown in Fig. 37.1.

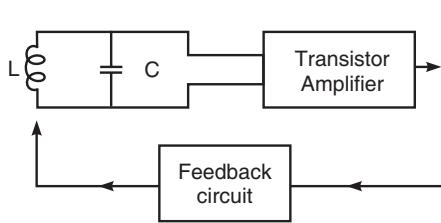


Fig. 37.1



Oscillators

**(i) Tank circuit.** It is a parallel combination of inductor  $L$  and capacitor  $C$ . The frequency of oscillations in the circuit depends upon the values of inductance and capacitance.

$$f = \frac{1}{2\pi} \sqrt{\left(\frac{1}{LC}\right)}.$$

**(ii) Transistor amplifier.** The function of the amplifier is to amplify the oscillations produced by the tank circuit.

**(iii) Feedback circuit.** A positive feedback network to transfer a part of output energy to the tank circuit in proper phase. The amount of energy feedback is enough to compensate for the losses of energy in the *LC* circuit.

#### Condition for Self-Excitation (Barkhausen Criterion for Oscillations)

Consider a feedback amplifier (Fig. 37.2). Let  $A$  = voltage gain of amplifier without feedback,  $e_g$  = input voltage and  $e_o$  = output voltage. Let the feedback network introduce a fraction  $\beta$  of the output into the input. If the feedback is positive, the voltage appearing at the input of the amplifier is  $e_g + \beta e_o$ . It is amplified  $A$  times by the amplifier to give the output  $e_o$ .

$$\therefore (e_g + \beta e_o) A = e_o$$

$$\text{or } e_g = \frac{e_o (1 - \beta A)}{A}.$$

$A$  is called open loop gain and  $A\beta$  is called feedback factor or loop gain. If  $\beta A = 1$ ,  $e_g = 0$ , i.e., an output voltage is obtained without any input signal. Then the amplifier functions as an oscillator. Therefore, the condition for the maintenance of oscillations is

$$\beta A = 1 \quad \dots(1)$$

$$\text{or } \beta = \frac{1}{A} \quad \dots(2)$$

Eq. (1) is called ‘Barkhausen criterion’ for self-sustained oscillations.

In practical oscillators,  $\beta A > 1$ .

In general, the amplifier gain  $A$  and the feedback factor  $\beta$  will be complex. This means that, satisfying Eqn. (1) involves two conditions:

(i) The magnitude of the product  $A\beta$  should be unity.

(ii) The product of the phase angles of the two should be zero or an integral multiple of  $2\pi$ .

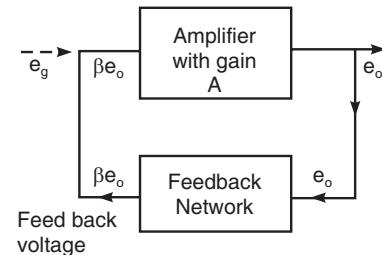


Fig. 37.2

### 37.2 HARTLEY OSCILLATOR

**Circuit diagram.** Figure 37.3 shows the circuit of a shunt-fed Hartley oscillator. The resistors  $R_1$ ,  $R_2$  and  $R_E$  and the supply voltage  $V_{cc}$  establish the d.c. operating point of the transistor. The frequency determining network is made up of the variable capacitor  $C$  and the inductors  $L_1$  and  $L_2$ .

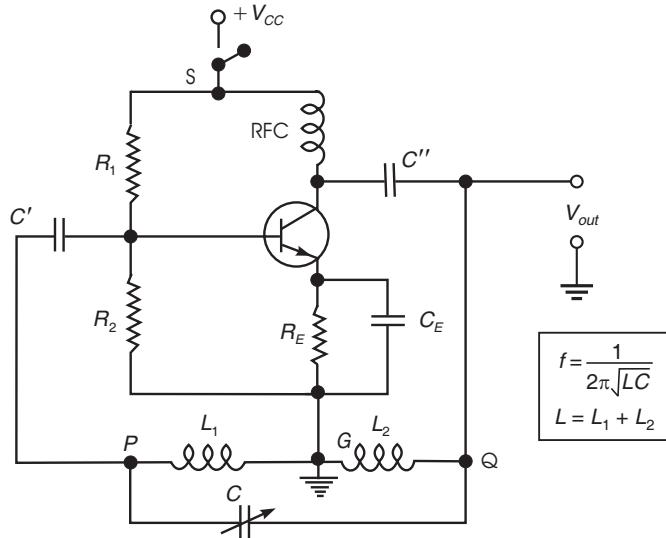


Fig. 37.3

The coil  $L_1$  is inductively coupled to coil  $L_2$  and the combination forms an autotransformer. The capacitor  $C''$  blocks dc and provides an ac path from the collector to the tank circuit. Radio frequency choke RFC provides d.c. load for the collector and also prevents a.c. from reaching the d.c. supply  $V_{cc}$ . Capacitor  $C'$  and resistor  $R_2$  produce self bias voltage. Capacitor  $C_E$  provides a.c. ground and prevents any signal degeneration.  $C_E$  also provides temperature stabilization. The input to the amplifier appears to the base through the capacitor  $C'$ . The output is available at the collector through the capacitor  $C''$  with respect to the ground.

**Working.** When the switch  $S$  is closed, the collector current starts increasing and charges the capacitor  $C$ . When this capacitor is fully charged, it discharges through coils  $L_1$  and  $L_2$ , setting up damped harmonic oscillations in the tank circuit. The oscillatory current in the tank circuit produces an a.c. voltage across  $L_1$  which is applied to the base-emitter junction of the transistor and appears in amplified form in the collector circuit. Feedback of energy from output (collector-emitter circuit) to input (base-emitter circuit) is accomplished through autotransformer action. Thus the energy is feedback into the tank circuit by means of mutual inductance between  $L_1$  and  $L_2$ . This energy supplied to the tank circuit overcomes the losses occurring in it. Consequently the oscillations are sustained in the circuit.

Because of earth connection at  $G$ , the potentials at  $Q$  and  $P$  are  $180^\circ$  out of phase. Thus the phase difference between the voltage across  $L_1$  and that across  $L_2$  is always  $180^\circ$ . A further phase shift of  $180^\circ$  is introduced between the input and output voltage by the transistor itself. Thus the total phase shift becomes  $360^\circ$ , thereby making the feedback positive which is essential for oscillations. Consequently, continuous undamped oscillations are obtained.

#### Expression for frequency

Fig. 37.4 shows the equivalent circuit of the oscillator.  $Z_1$ ,  $Z_2$  and  $Z_3$  represent the inductive or capacitive reactances of oscillator. Here the feedback source of e.m.f.  $h_{re} V_0$  has been omitted because

reverse voltage  $h_{re}$  of a transistor is negligible. The output resistance  $1/h_{0e}$  in parallel with  $Z_2$  is very large. Therefore, it has also been omitted.

The input resistance  $h_{ie}$  of the transistor and impedance  $Z_1$  are in parallel. Their equivalent impedance  $Z'$  is given by

$$\frac{1}{Z'} = \frac{1}{Z_1} + \frac{1}{h_{ie}} = \frac{Z_1 + h_{ie}}{Z_1 h_{ie}}$$

$$\text{or } Z' = \frac{Z_1 h_{ie}}{Z_1 + h_{ie}}$$

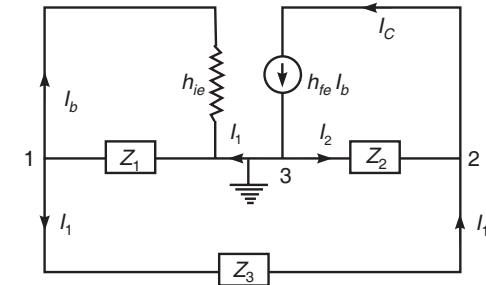


Fig. 37.4

...(1)

This equivalent impedance  $Z'$  is in series with  $Z_3$ . This combination acts in parallel with  $Z_2$ . Thus the *load impedance* ( $Z_L$ ) between the output terminals 2 and 3 is given by

$$\begin{aligned} \frac{1}{Z_L} &= \frac{1}{Z_2} + \frac{1}{Z' + Z_3} \\ \frac{1}{Z_L} &= \frac{1}{Z_2} + \frac{1}{\frac{Z_1 h_{ie}}{Z_1 + h_{ie}} + Z_3} \\ &= \frac{1}{Z_2} + \frac{Z_1 + h_{ie}}{Z_1 h_{ie} + Z_1 Z_3 + Z_3 h_{ie}} \\ &= \frac{1}{Z_2} + \frac{Z_1 + h_{ie}}{Z_1 Z_3 + h_{ie} (Z_1 + Z_3)} \\ &= \frac{Z_1 Z_3 + h_{ie} (Z_1 + Z_3) + Z_2 (Z_1 + h_{ie})}{Z_2 [Z_1 Z_3 + h_{ie} (Z_1 + Z_3)]} \end{aligned}$$

$$\text{or } Z_L = \frac{Z_2 [Z_1 Z_3 + h_{ie} (Z_1 + Z_3)]}{h_{ie} (Z_1 + Z_2 + Z_3) + Z_1 Z_2 + Z_1 Z_3} \quad \dots(2)$$

The *output voltage* between terminals 3 and 2 in terms of current  $I_1$  is

$$\begin{aligned} V_0 &= -I_1 (Z' + Z_3) = -I_1 \left( \frac{Z_1 h_{ie}}{Z_1 + h_{ie}} + Z_3 \right) = -I_1 \left( \frac{Z_1 h_{ie} + Z_1 Z_3 + Z_3 h_{ie}}{Z_1 + h_{ie}} \right) \\ V_0 &= -I_1 \left[ \frac{Z_1 Z_3 + h_{ie} (Z_1 + Z_3)}{Z_1 + h_{ie}} \right] \end{aligned} \quad \dots(3)$$

*Voltage feedback* to the input terminals 3 and 1 is given by

$$V_{fb} = -I_1 Z' = -I_1 \left( \frac{Z_1 h_{ie}}{Z_1 + h_{ie}} \right) \quad \dots(4)$$

Therefore, the feedback fraction  $\beta$  is given by

$$\begin{aligned} \beta &= \frac{V_{fb}}{V_0} = I_1 \left( \frac{Z_1 h_{ie}}{Z_1 + h_{ie}} \right) \left[ \frac{Z_1 + h_{ie}}{Z_1 Z_3 + h_{ie} (Z_1 + Z_3)} \right] \frac{1}{I_1} \\ \beta &= \frac{Z_1 h_{ie}}{Z_1 Z_3 + h_{ie} (Z_1 + Z_3)} \end{aligned} \quad \dots(5)$$

Now from Barkhausen criterion, the condition for the maintenance of oscillation is

$$\beta A_v = 1 \quad \dots(6)$$

Here,  $A_v$  is the voltage gain without feedback. For a *CE* amplifier,

$$A_v = -\frac{h_{fe}}{h_{ie}} Z_L \quad \dots(7)$$

Substituting the value of  $\beta$  from Eq. (5) and  $A_v$  from Eq. (7) in Eq. (6), we get

$$\begin{aligned} & \left( -\frac{h_{fe} Z_L}{h_{ie}} \right) \left[ \frac{Z_1 h_{ie}}{Z_1 Z_3 + h_{ie}(Z_1 + Z_3)} \right] = 1 \\ \text{or } & \left[ \frac{h_{fe} Z_2 \{Z_1 Z_3 + h_{ie}(Z_1 + Z_3)\}}{h_{ie}(Z_1 + Z_2 + Z_3) + Z_1 Z_2 + Z_1 Z_3} \right] \left[ \frac{Z_1}{Z_1 Z_3 + h_{ie}(Z_1 + Z_3)} \right] = -1 \\ \text{or } & \frac{h_{fe} Z_1 Z_2}{h_{ie}(Z_1 + Z_2 + Z_3) + Z_1 Z_2 + Z_1 Z_3} = -1 \\ \text{or } & h_{ie}(Z_1 + Z_2 + Z_3) + Z_1 Z_2 + Z_1 Z_3 = -h_{fe} Z_1 Z_2 \\ \text{or } & h_{ie}(Z_1 + Z_2 + Z_3) + Z_1 Z_2 (1 + h_{fe}) + Z_1 Z_3 = 0 \end{aligned} \quad \dots(8)$$

Now, we apply this general equation to the Hartley oscillator.

Let  $M$  be the mutual inductance between the two coils of self inductance  $L_1$  and  $L_2$ .  $C$  is the capacitance in the tank circuit. Then

$$\begin{aligned} Z_1 &= j\omega L_1 + j\omega M \\ Z_2 &= j\omega L_2 + j\omega M \end{aligned}$$

and

$$Z_3 = \frac{1}{j\omega C} = -\frac{j}{\omega C}$$

Therefore, Eq. (8) gives

$$\begin{aligned} & h_{ie} \left[ (j\omega L_1 + j\omega M) + (j\omega L_2 + j\omega M) - \frac{j}{\omega C} \right] \\ & + (j\omega L_1 + j\omega M)(j\omega L_2 + j\omega M)(1 + h_{fe}) \\ & + (j\omega L_1 + j\omega M) \left( -\frac{j}{\omega C} \right) = 0 \\ \text{or } & j\omega h_{ie} \left[ L_1 + L_2 + 2M - \frac{1}{\omega^2 C} \right] \\ & - \omega^2 (L_1 + M)(L_2 + M)(1 + h_{fe}) + \frac{L_1 + M}{C} = 0 \\ \text{or } & j\omega h_{ie} \left[ L_1 + L_2 + 2M - \frac{1}{\omega^2 C} \right] \\ & - \omega^2 (L_1 + M) \left[ (L_2 + M)(1 + h_{fe}) - \frac{1}{\omega^2 C} \right] = 0 \end{aligned} \quad \dots(9)$$

Equating the imaginary part on both sides of Eq. (9), we get

$$L_1 + L_2 + 2M - \frac{1}{\omega^2 C} = 0$$

$$\text{or } \omega^2 C = \frac{1}{L_1 + L_2 + 2M} \quad \dots(10)$$

$$\text{or } \omega = \frac{1}{\sqrt{(L_1 + L_2 + 2M)C}}$$

*Frequency of oscillation,*

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi\sqrt{(L_1 + L_2 + 2M)C}} \quad \dots(11)$$

The *condition for maintenance of oscillations* is found by equating the real part of Eq. (9) to zero. Thus

$$(L_2 + M)(1 + h_{fe}) - \frac{1}{\omega^2 C} = 0$$

$$\text{or } 1 + h_{fe} = \frac{1}{\omega^2 C(L_2 + M)}$$

Substituting the value of  $\omega^2 C$  from Eq. (10), we get

$$\begin{aligned} 1 + h_{fe} &= \frac{L_1 + L_2 + 2M}{L_2 + M} \\ &= \frac{(L_1 + M) + (L_2 + M)}{L_2 + M} = 1 + \frac{L_1 + M}{L_2 + M} \\ \text{or } h_{fe} &= \frac{L_1 + M}{L_2 + M} \end{aligned} \quad \dots(12)$$

Eq. (12) gives the condition for sustained oscillation.

### Expression for Frequency (II method)

The impedance of tank circuit is

$$\begin{aligned} Z &= (j\omega L_1 + j\omega L_2) - \frac{j}{\omega C} \\ Z &= j\omega \left[ L_1 + L_2 - \frac{1}{\omega^2 C} \right] \end{aligned} \quad \dots(1)$$

At resonance, the net impedance of the oscillatory circuit is zero.

$$\therefore Z = j\omega \left( L_1 + L_2 - \frac{1}{\omega^2 C} \right) = 0 \quad \dots(2)$$

$$\omega^2 = \frac{1}{(L_1 + L_2)C}$$

$$\text{or } \omega = \frac{1}{\sqrt{(L_1 + L_2)C}}$$

$\therefore$  Frequency of oscillation,

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi\sqrt{(L_1 + L_2)C}} \quad \dots(3)$$

The voltage across  $L_1$  is feedback into the circuit. The voltage across  $L_2$  is taken out to the output. The feedback ratio  $\beta$  is given by

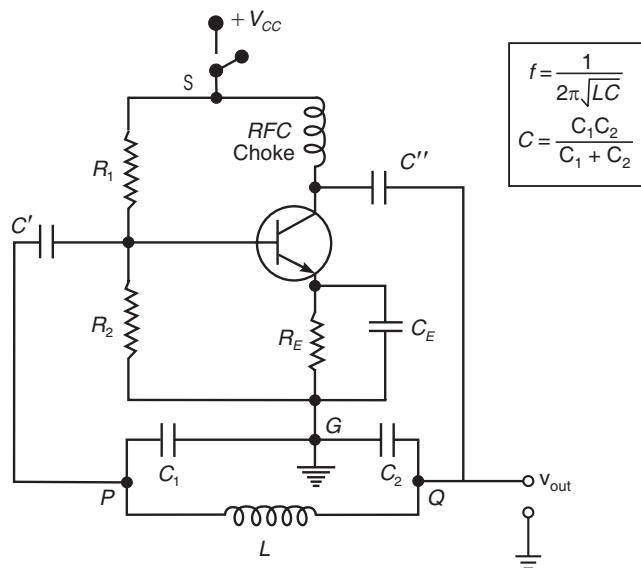
$$\beta = \frac{V_f}{V_0} = \frac{j\omega L_1}{j\omega L_2} = \frac{L_1}{L_2}$$

According to the Barkhausen criterion, the condition for getting sustained oscillations is

$$A = \frac{1}{\beta} \quad \text{or} \quad A = \frac{L_2}{L_1} \quad \dots (4)$$

### 37.3 COLPITT'S OSCILLATOR

**Circuit Diagram.** Fig. 37.5 shows the circuit of a Colpitt's Oscillator. Two capacitors  $C_1$  and  $C_2$  are placed across a common inductance  $L$  and the centre of the two capacitors is tapped. The resistors  $R_1$  and  $R_2$  form a voltage divider across the supply voltage  $V_{CC}$  and provide necessary bias condition for the circuit along with  $R_E$ . The function of capacitor  $C''$  is to block d.c. and provide an a.c. path from collector to the tank circuit.  $C_E$  is a bypass capacitor. Radio frequency choke RFC provides the necessary d.c. load resistance for collector and also prevents a.c. signal from entering the d.c. supply  $V_{CC}$ .  $C''$  conveys feedback from the collector to base circuit.



**Fig. 37.5**

**Working.** When the switch  $S$  is closed, the collector current starts increasing. Capacitors  $C_1$  and  $C_2$  are charged. These capacitors discharge through coil  $L$  setting up damped harmonic oscillations in the tank circuit. The oscillatory current in the tank circuit produces an a.c. voltage across  $C_1$ . The voltage across  $C_1$  is applied to base-emitter junction of the transistor and appears in amplified form in the collector circuit. It overcomes the losses occurring in the tank circuit.

The energy supplied to the tank circuit is in phase with the generated oscillations. The feedback voltage (voltage across the capacitor  $C_1$ ) is  $180^\circ$  out of phase with the output voltage (voltage across the capacitor  $C_2$ ). It is because, the terminal  $G$  is earthed which will be at a zero potential. When  $P$  is at a positive potential with respect to  $G$ ,  $Q$  will be at a negative potential with respect to  $G$  at that instant. Therefore, points  $P$  and  $Q$  are  $180^\circ$  out of phase. A further phase shift of  $180^\circ$  between the output and input voltage is produced by common emitter transistor. Thus the total phase shift is  $360^\circ$ . Thus the phase shift condition for sustained oscillation is satisfied and we get continuous undamped oscillations.

### Expression for Frequency

The frequency and condition of oscillations in Colpitt's oscillator can be obtained by applying the *general equation of oscillator* [Eq. (8) in Hartley Oscillator]. That is,

$$h_{ie}(Z_1 + Z_2 + Z_3) + Z_1 Z_2 (1 + h_{fe}) + Z_1 Z_3 = 0.$$

Here,

$$Z_1 = \frac{1}{j\omega C_1} = -\frac{j}{\omega C_1}$$

$$Z_2 = \frac{1}{j\omega C_2} = -\frac{j}{\omega C_2}$$

and

$$Z_3 = j\omega L.$$

Substituting these values in the general equation, we get

$$\begin{aligned} h_{ie} \left( -\frac{j}{\omega C_1} - \frac{j}{\omega C_2} + j\omega L \right) + \left( -\frac{j}{\omega C_1} \right) \left( -\frac{j}{\omega C_2} \right) (1 + h_{fe}) + \left( -\frac{j}{\omega C_1} \right) j\omega L &= 0 \\ \text{or } -jh_{ie} \left( \frac{1}{\omega C_1} + \frac{1}{\omega C_2} - \omega L \right) - \frac{1 + h_{fe}}{\omega^2 C_1 C_2} + \frac{L}{C_1} &= 0 \\ \text{or } \left( \frac{1 + h_{fe}}{\omega^2 C_1 C_2} - \frac{L}{C_1} \right) + jh_{ie} \left( \frac{1}{\omega C_1} + \frac{1}{\omega C_2} - \omega L \right) &= 0 \end{aligned} \quad \dots(1)$$

Equating the imaginary parts on both sides of Eq. (1), we get

$$\frac{1}{\omega C_1} + \frac{1}{\omega C_2} - \omega L = 0$$

$$\text{or } \frac{C_1 + C_2}{\omega C_1 C_2} - \omega L = 0$$

$$\text{or } \omega^2 L C_1 C_2 = C_1 + C_2 \quad \dots(2)$$

$$\text{or } \omega = \sqrt{\left( \frac{C_1 + C_2}{LC_1 C_2} \right)}$$

*Frequency of oscillation*

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi} \sqrt{\left( \frac{C_1 + C_2}{LC_1 C_2} \right)} \quad \dots(3)$$

Equating the real parts on both sides of Eq. (1), we get

$$\frac{1 + h_{fe}}{\omega^2 C_1 C_2} - \frac{L}{C_1} = 0 \quad \text{or } 1 + h_{fe} = \frac{\omega^2 L C_1 C_2}{C_1}$$

Substituting the value of  $\omega^2 L C_1 C_2$  from Eq. (2), we get

$$1 + h_{fe} = \frac{C_1 + C_2}{C_1} = 1 + \frac{C_2}{C_1}$$

$$\text{or } h_{fe} = \frac{C_2}{C_1} \quad \dots(4)$$

This is the condition for sustained oscillations.

### Expression for frequency of Oscillation (II method)

The impedance of tank circuit is

$$Z = -\frac{j}{\omega C_1} - \frac{j}{\omega C_2} + j\omega L$$

$$Z = -j \left[ \frac{1}{\omega C_1} + \frac{1}{\omega C_2} - \omega L \right]$$

At resonance, the net impedance of the oscillatory circuit is zero.

$$\therefore Z = -j \left( \frac{1}{\omega C_1} + \frac{1}{\omega C_2} - \omega L \right) = 0 \quad \dots (1)$$

$$\frac{1}{\omega C_1} + \frac{1}{\omega C_2} - \omega L = 0$$

$$\text{or } \frac{C_1 + C_2}{\omega C_1 C_2} - \omega L = 0$$

$$\text{or } \omega^2 L C_1 C_2 = C_1 + C_2$$

$$\text{or } \omega = \sqrt{\left( \frac{C_1 + C_2}{L C_1 C_2} \right)} \quad \dots (2)$$

Frequency of oscillation

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi} \sqrt{\left( \frac{C_1 + C_2}{L C_1 C_2} \right)} \quad \dots (3)$$

The voltage across  $C_1$  is feedback into the circuit. The voltage across  $C_2$  is taken out to the output. The feedback ratio  $\beta$  is given by

$$\beta = \frac{V_f}{V_o} = \frac{-\frac{j}{\omega C_1}}{-\frac{j}{\omega C_2}} = \frac{C_2}{C_1}$$

According to the Barkhausen criterion, the condition for getting sustained oscillations is

$$A = \frac{1}{\beta}$$

$$\text{or } A = \frac{C_1}{C_2} \quad \dots (4)$$

**R-C oscillators.** Good frequency stability and waveform can be obtained from oscillators employing resistive and capacitive elements. They provide proper phase for the positive feedback voltage. We shall discuss the principle and working of

- (i) Wien bridge oscillator and
- (ii) Phase shift oscillator.

### 37.4 WIEN-BRIDGE OSCILLATOR

Figure 37.6 shows the Wien bridge network. The frequency of oscillations is determined by the series element  $R_1 C_1$  and parallel element  $R_2 C_2$ . The upper arm consists of resistance  $R$  and capacitance  $C$  in series. The lower arm contains a parallel combination of equal resistance and capacitance.

The ratio of output voltage ( $V_0$ ) of the network to the input voltage ( $V_i$ ) is

$$\begin{aligned} \frac{V_0}{V_i} &= \frac{\text{Impedance of parallel combination}}{\text{Total impedance}} \\ \frac{V_0}{V_i} &= \frac{-\frac{(R \cdot j/\omega C)}{R - j/\omega C}}{R - \frac{j}{\omega C} - \frac{(R \cdot j/\omega C)}{R - j/\omega C}} \\ &= \frac{-jR/\omega C}{\left(R - \frac{j}{\omega C}\right)^2 - \frac{jR}{\omega C}} = \frac{-jR/\omega C}{\left(R^2 - \frac{1}{\omega^2 C^2} - \frac{3jR}{\omega C}\right)} \end{aligned}$$

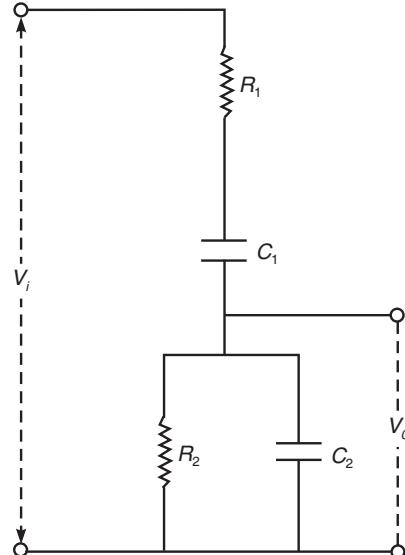


Fig. 37.6

...(1)

The phase shift will be zero if imaginary term vanishes, i.e.,

$$R^2 - \frac{1}{\omega^2 C^2} = 0 \quad \text{or} \quad \omega = \frac{1}{RC}.$$

*Frequency of oscillation*

$$f_0 = \frac{\omega}{2\pi} = \frac{1}{2\pi RC} \quad \dots(2)$$

$$\text{From Eq. (1), } \frac{V_0}{V_i} = \frac{-jR/\omega C}{-3jR/\omega C} = \frac{1}{3} \quad \dots(3)$$

Thus the oscillations will be sustained if the amplifier has a gain just exceeding 3.

**Circuit Diagram.** Figure 37.7 shows the circuit diagram of Wien bridge oscillator.

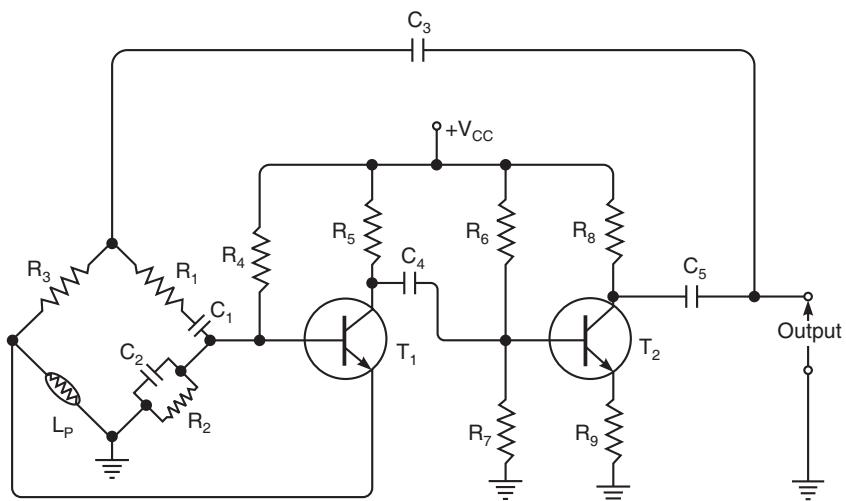


Fig. 37.7

- It is a two stage transistor amplifier with ***R-C bridge circuit***. In the bridge circuit,  $R_1$  in series with  $C_1$ ,  $R_3$ , tungsten lamp  $L_P$  and  $R_2$  in parallel with  $C_2$  form the four arms.
  - Resistances  $R_3$  and  $L_P$  are used to stabilise the amplitude of the output.
  - The transistor  $T_1$  serves as an oscillator and amplifier.
  - The transistor  $T_2$  serves as an inverter to produce a phase shift of  $180^\circ$ .
  - The circuit uses *positive and negative feedbacks*.
- The positive feedback is through  $R_1C_1, R_2C_2$  to the transistor  $T_1$ .
- The negative feedback is through the voltage divider  $R_3 - L_P$  to the emitter of transistor  $T_1$ .
- If  $R_1 = R_2$  and  $C_1 = C_2 = C$ , then

$$f = \frac{1}{2\pi RC} \quad \dots (1)$$

### Working

- When the circuit is started, bridge circuit produces oscillations of frequency determined by Eq. (1).
  - The two transistors produce a total phase shift of  $360^\circ$  so that proper positive feedback is ensured.
  - The negative feedback in the circuit ensures constant output. This is achieved by the temperature sensitive tungsten lamp  $L_P$ . Its resistance increases with current.
- If the amplitude of output tends to increase, more current would provide more negative feedback. The result is that the output would return to original value. A reverse action would take place if the output tends to decrease.
- Wien bridge oscillator finds wide use in variable audio frequency generators. It is a low frequency, low distortion sine wave generator. It gives constant output.

#### 37.4.1 Wien Bridge Oscillator

**Circuit Diagram** A practical oscillator circuit uses an op-amp and *RC* bridge circuit, with the oscillator frequency set by the *R* and *C* components. Fig. 37.8 shows the Wien bridge oscillator circuit using op-amp amplifier.

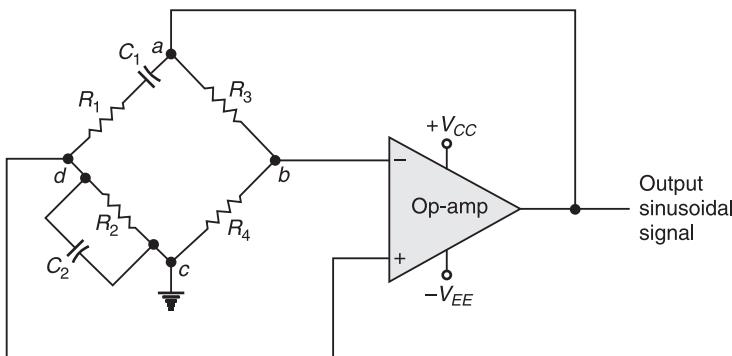


Fig. 37.8

- Resistors  $R_1, R_2$  and capacitors  $C_1, C_2$  form the frequency-adjustment elements.
- Resistors  $R_3$  and  $R_4$  form part of the feedback path.

- (iii) The op-amp output is connected as the bridge input at points *a* and *c*.  
(iv) The bridge circuit output at points *b* and *d* is the input to the op-amp.

### Frequency of Oscillation

Neglecting loading effects of the op-amp input and output impedances, the analysis of the bridge circuit results in

$$\frac{R_3}{R_4} = \frac{R_1}{R_2} + \frac{C_2}{C_1} \quad \dots(1)$$

and

$$f_0 = \frac{1}{2\pi\sqrt{R_1 C_1 R_2 C_2}} \quad \dots(2)$$

If  $R_1 = R_2 = R$  and  $C_1 = C_2 = C$ , the oscillator frequency is

$$f_0 = \frac{1}{2\pi RC} \quad \dots(3)$$

and

$$\frac{R_3}{R_4} = 2 \quad \dots(4)$$

Thus, a ratio of  $R_3$  to  $R_4$  greater than 2 will provide sufficient loop gain for the circuit to oscillate at the frequency calculated using Eq. (3).

**EXAMPLE 1.** Find the frequency of oscillations of a Wien bridge oscillator having  $R_1 = R_2 = 220 k\Omega$  and  $C_1 = C_2 = 250 pF$ .  
(Bharthiar, Nov. 2010)

$$\text{Sol. } f = \frac{1}{2\pi RC} = \frac{1}{2\pi \times (220 \times 10^3) \times (250 \times 10^{-12})} \text{ Hz} = 2892 \text{ Hz}$$

## 37.5 PHASE SHIFT OSCILLATOR

**Principle.** Consider an  $RC$  network, connected in series with a signal generator (Fig. 37.9a). Let a sinusoidal voltage  $V_i = E_0 \sin \omega t$  of frequency  $f = \frac{\omega}{2\pi}$  be applied to the network. The alternating voltage  $V_o$  across  $R$  leads the applied voltage  $V_i$  by an angle  $\phi$ , given by

$$\tan \phi = \frac{1}{\omega CR} = \frac{1}{2\pi f CR}.$$

For a given frequency  $f$ , by adjusting the values of  $C$  and  $R$ ,  $\phi$  can be made equal to  $60^\circ$ .

In a phase shift oscillator, a ladder network consisting of three  $RC$  sections is used (Fig. 37.9 b). Each section will produce a phase shift of  $60^\circ$ .

Consequently a total phase shift of  $180^\circ$  is produced between the input and output voltages i.e.,  $V_o$  leads  $V_i$  by  $180^\circ$ .

**Circuit diagram.** Figure 37.10 shows a transistor phase shift oscillator in  $CE$  mode.  $R_L$  is the collector load resistor.  $C_E$  is the emitter bypass capacitor. The resistors  $R_1$ ,  $R_2$  and  $R_E$  provide the necessary bias conditions for the circuit.

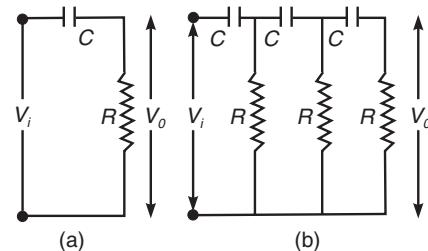


Fig. 37.9

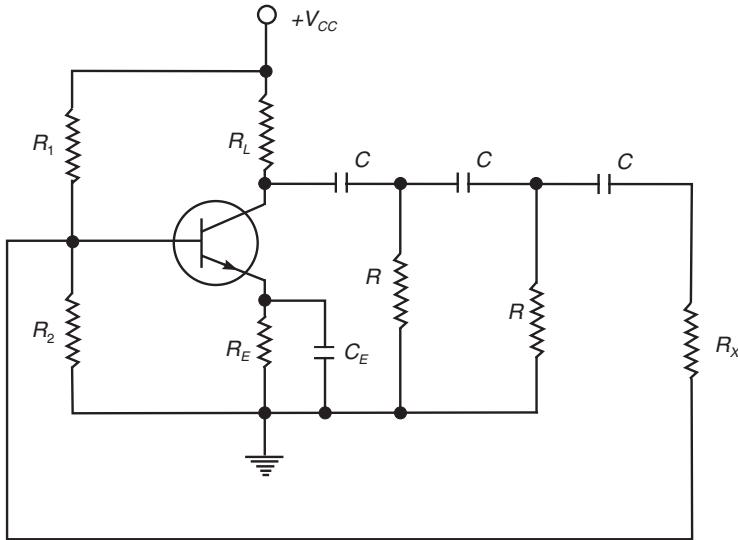


Fig. 37.10

For feedback of energy from collector to base, a ladder network consisting of three identical  $RC$  sections is connected between collector and base. The last section contains a resistance  $R_x$  such that

$$R_x + h_{ie} = R$$

where  $h_{ie}$  is the input resistance of the transistor. Since this resistance is connected with the base of the transistor, the input resistance  $h_{ie}$  of the transistor is added to it to give a total resistance of  $R$ .

**Working.** The circuit is set into oscillations by any random variation caused in the base current by the minor variation in the oscillator power supply  $V_{CC}$  or by the noise inherent in the transistor. This variation in the base current is amplified in the collector circuit and then feedback to the  $RC$  network. The network produces a phase shift of  $180^\circ$  between its input and output voltages. An additional phase shift of  $180^\circ$  is produced by the common emitter transistor. Consequently the total phase shift around the entire loop is  $360^\circ$  or  $0^\circ$  and the circuit acts as an oscillator. The oscillator produces sine wave output voltage continuously.

The frequency of oscillation is given by

$$f = \frac{1}{2\pi C \sqrt{(4RR_L + 6R^2)}}.$$

#### Expression for frequency of oscillation

The equivalent circuit of phase shift oscillator is shown in Fig. 37.11.

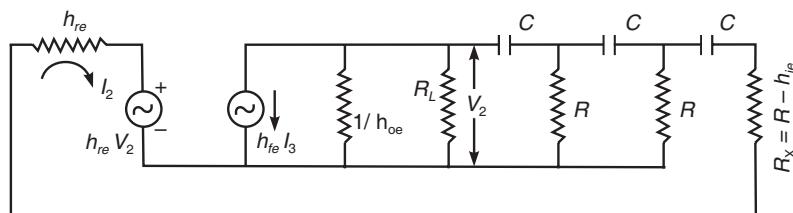


Fig. 37.11

In this circuit  $1/h_{oe}$  can be neglected because it is much larger than  $R_L$ . Also,  $h_{re} V_2$  can be neglected because  $h_{re}$  of a transistor is very small. We can replace the current source  $h_{fe} I_3$  by an

equivalent Thevenin voltage source of generated voltage  $h_{fe} I_3 R_L$ . So the reduced equivalent circuit is shown in Fig. 37.12.

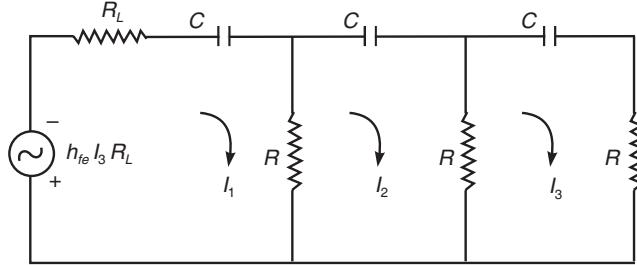


Fig. 37.12

Applying Kirchhoff's voltage laws to the three meshes, we get

$$(R_L + R - jX_c) I_1 - RI_2 + h_{fe} R_L I_3 = 0 \quad \dots(1)$$

$$-RI_1 + (2R - jX_c) I_2 - RI_3 = 0 \quad \dots(2)$$

and

$$0 \times I_1 - RI_2 + (2R - jX_c) I_3 = 0 \quad \dots(3)$$

Here,  $X_c = \frac{1}{\omega C}$  is the capacitive reactance.

The currents  $I_1$ ,  $I_2$  and  $I_3$  are non-zero. So the determinant of the coefficients of  $I_1$ ,  $I_2$  and  $I_3$  must be zero.

$$\begin{vmatrix} R_L + R - jX_c & (-R) & h_{fe} R_L \\ -R & (2R - jX_c) & (-R) \\ 0 & (-R) & (2R - jX_c) \end{vmatrix} = 0 \quad \dots(4)$$

$$(R_L + R - jX_c)(3R^2 - 4RjX_c - X_c^2) - R^2(2R - jX_c) + h_{fe} R_L R^2 = 0 \quad \dots(5)$$

Equating the imaginary parts of Eq. (5), we get

$$-4RX_c(R_L + R) - X_c(3R^2 - X_c^2) + R^2X_c = 0$$

$$\text{or} \quad -4RR_L - 4R^2 - 3R^2 + X_c^2 + R^2 = 0$$

$$\text{or} \quad X_c^2 = 4RR_L + 6R^2 \quad \dots(6)$$

$$\text{or} \quad \frac{1}{\omega^2 C^2} = 4RR_L + 6R^2 \quad \text{or} \quad \omega^2 = \frac{1}{C^2(4RR_L + 6R^2)}$$

$$\therefore f = \frac{\omega}{2\pi} = \frac{1}{2\pi C \sqrt{(4RR_L + 6R^2)}} \quad \dots(7)$$

#### Condition for Maintenance of Oscillations

Equating the real parts of Eq. (5), we get

$$(R_L + R)(3R^2 - X_c^2) - 4RX_c^2 - 2R^3 + h_{fe} R_L R^2 = 0$$

$$\text{or} \quad 3R^2 R_L + 3R^3 - R_L X_c^2 - RX_c^2 - 4RX_c^2 - 2R^3 + h_{fe} R_L R^2 = 0$$

$$\text{or} \quad 3R^2 R_L + R^3 - X_c^2 (5R + R_L) + h_{fe} R_L R^2 = 0$$

$$\text{From Eq. (6), } X_c^2 = 4RR_L + 6R^2$$

$$\therefore 3R^2 R_L + R^3 - (4RR_L + 6R^2)(5R + R_L) + h_{fe} R_L R^2 = 0$$

or  $h_{fe} R_L R^2 = 23R^2 R_L + 29R^3 + 4RR_L^2$

$$\therefore h_{fe} = 23 + 29 \frac{R}{R_L} + 4 \frac{R_L}{R} \quad \dots(9)$$

Usually,  $R = R_L$ . Hence the condition of sustained oscillations is

$$h_{fe} = 56.$$

Thus for sustained oscillations in phase shift oscillator, the minimum value of the forward current gain ratio of the transistor should be 56.

**Variable-frequency Operation.** The phase-shift oscillator is particularly suited to the range of frequencies from several hertz to several hundred kilohertz, and so includes the range of audio frequencies. The frequency of oscillation may be varied by changing any of the impedance elements in the phase-shifting network. For variations of frequency over a large range, the three capacitors are usually varied simultaneously. The phase-shift oscillator is operated in class A in order to keep distortion to a minimum.

## 37.6 CRYSTAL OSCILLATOR

**Principle.** When a suitably cut piezo-electric crystal is subjected to an appropriate alternating voltage, it vibrates mechanically (Fig. 37.13).

The amplitude of the mechanical oscillations becomes maximum when the frequency of the applied a.c. voltage is equal to the natural frequency of the crystal.

Fig. 37.14 shows the electrical equivalent circuit of a crystal. The inductor  $L$ , capacitor  $C$ , and resistor  $R$  are the analogs of the mass, the compliance (the reciprocal of the spring constant), and the viscous-damping factor of the mechanical system.  $C_m$  represents the electrostatic capacitance between electrodes with the crystal as a dielectric.

- (i) **Series resonant frequency  $f_s$ .** We neglect the resistance  $R$ .
- At low frequencies, the impedance of the network is high and capacitive (Fig. 37.15).

- As the frequency is increased,  $R - L - C$  branch approaches its resonant frequency.

At resonant frequency  $f_s$ ,  $X_L = X_C$

$$f_s = \frac{1}{2\pi\sqrt{LC}}$$

It is called the *series resonant frequency* (the zero impedance frequency). At  $f_s$ , the crystal acts as a series resonant circuit.

(ii) **Parallel resonant frequency  $f_p$ .** At  $f_p$ , the crystal acts as a parallel resonant circuit. The crystal offers a very high impedance. The formula for parallel resonant frequency is:

$$f_p = \frac{1}{2\pi\sqrt{LC_T}}$$



Fig. 37.13

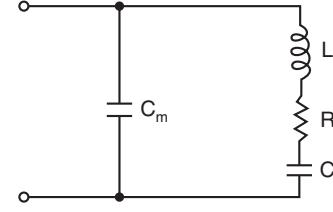


Fig. 37.14

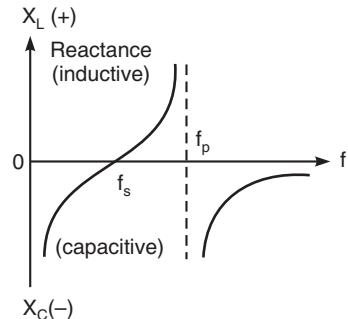


Fig. 37.15

Here,

$$C_T = \frac{C \times C_m}{C + C_m}$$

In any crystal,  $C$  is much smaller than  $C_m$ . Because of this,  $f_p$  is only slightly greater than  $f_s$ .  $f_p \approx f_s$ .

Fig. 37.16 shows the *crystal impedance versus frequency* curve.

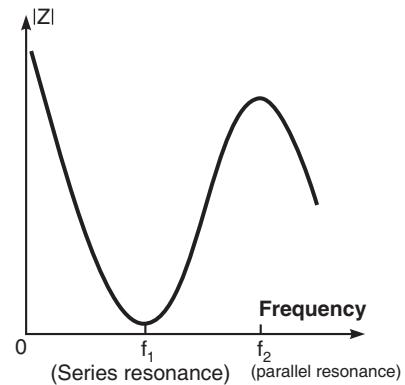


Fig. 37.16

### 37.6.1 Colpitts Crystal Oscillator

Fig. 37.17 shows a Colpitts crystal oscillator.

- The voltage across  $C_2$  is fed back to the input.
- The capacitive voltage divider produces the feedback voltage for the base of the transistor.
- The crystal acts like an inductor that resonates with  $C_1$  and  $C_2$ .
- The oscillation frequency is between the series and parallel resonant frequencies of the crystal.

#### Crystal stability

The frequency of any oscillator tends to change slightly with time. This *drift* is produced by temperature, aging, and other causes. In a crystal oscillator, the frequency drift is very small, typically less than 1 part in  $10^6$  per day. Stability like this is important in electronic wristwatches because they use quartz-crystal oscillators as the basic timing device.

By putting a crystal oscillator in a temperature-controlled oven, we can get a frequency drift of less than 1 part in  $10^{10}$  per day. A clock with this drift will take 300 years to gain or lose 1s. Stability like this is needed in frequency and time standards.

#### Advantages

(1) Crystal oscillator has a high order of frequency stability. The circuit provides a constant frequency because the vibration of crystal is independent of temperature, transistor parameters etc. It mainly depends on the nature in which the crystal is cut and on its thickness.

(2)  $Q$  factor of crystal is given by  $Q = \frac{L\omega}{R}$  and it is of the order of  $10^6$ . For greater frequency stability,  $Q$  factor should be high.

(3) When the frequency of oscillation needs to be accurate and stable, a crystal oscillator is the natural choice. Electronic wristwatches and other critical timing applications use crystal oscillators because they provide an accurate clock frequency.

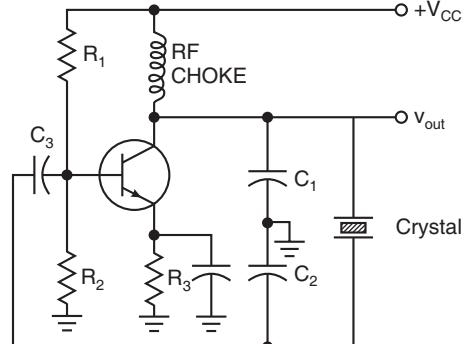


Fig. 37.17

**MULTIVIBRATORS (NON-SINUSOIDAL OSCILLATORS)**

### 37.7 INTRODUCTION

- A circuit which can turn ON or OFF current in an electrical circuit is called a **switching circuit**.

**Electronic switch.** It is a device which can turn ON or OFF current in an electrical circuit with the help of electronic devices *e.g.*, transistors.

#### Ideal switching action of a transistor

- When used as an electronic switch, a transistor is normally operated in cut off and saturation.

(i) **Cutoff—open switch.** When the base input voltage is enough negative, the transistor is **cut off**. No current flows in collector load [Fig. 37.18 (i)]. We assume that collector leakage current  $I_{\text{leakage}} = 0$ .

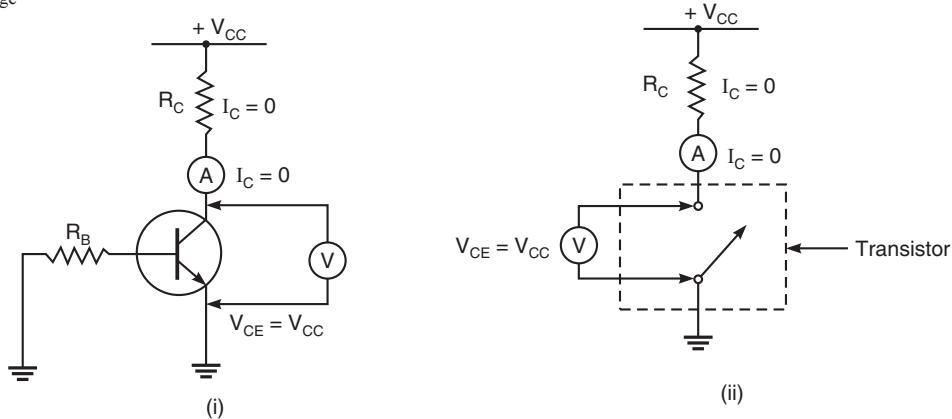


Fig. 37.18

- There is no voltage drop across  $R_C$ .

The output voltage is ideally  $V_{CC}$ , i.e.,

$$I_C = 0 \text{ and } V_{CE} = V_{CC}$$

This condition is similar to that of an **open switch** (*i.e.*, OFF state) [Fig. 37.18 (ii)].

(ii) **Saturation—closed switch.** When the input base voltage is positive enough that transistor saturates, then  $I_{C(\text{sat})}$  will flow through the  $R_C$  [Fig. 37.19 (i)].

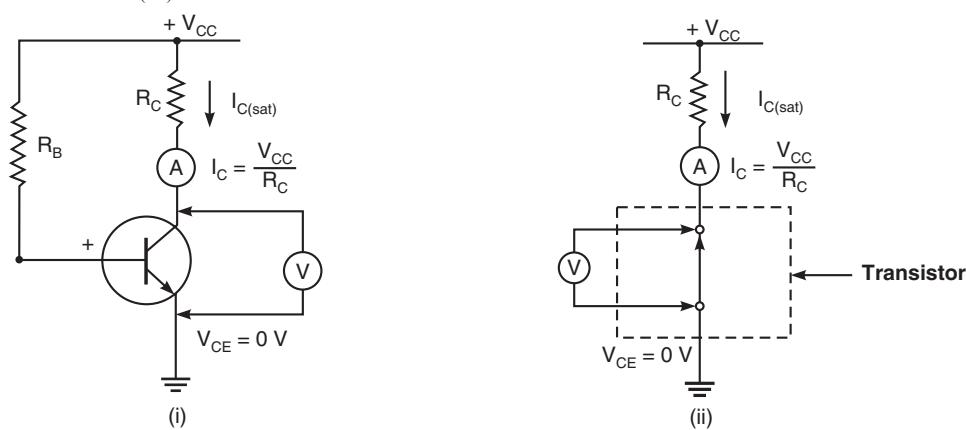


Fig. 37.19

- The entire  $V_{CC}$  will drop across collector load  $R_C$ .

Output voltage is ideally zero i.e.,

$$\frac{V_{CC}}{R_C} \text{ and } V_{CE} = 0 \quad I_C = I_{C(\text{sat})} =$$

This condition is similar to that of a closed switch (*i.e.*, ON state) [Fig. 37.19 (ii)].

- Thus a transistor can act as a switch. Fig. 37.20 shows the switching action of a transistor in terms of *dc* load line.
- The point *A* of the load line represents the ON condition.
- The point *B* of the load line represents the OFF condition.

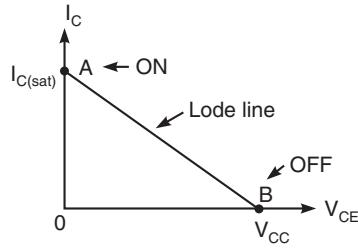


Fig. 37.20

#### Collector leakage current

In an ideal transistor, collector leakage current ( $I_{CEO}$ ) is extremely small and is usually neglected. But in a practical transistor, collector leakage current is not zero.

When the input circuit is reverse biased or input voltage is zero, a small current (a few  $\mu A$ ) flows. This is called collector leakage current and is due to thermally produced minority carriers.

- The value of this leakage current is quite large in *Ge* transistors.
- But in modern silicon transistors, the value of leakage current is low enough to be ignored.

#### Saturation collector current

- It is the maximum collector current for a particular load in a transistor.*

Consider an *npn* transistor having a load  $R_C$  in its collector circuit. (Fig. 37.21).

When the input forward bias is increased, the collector current  $I_C$  also increases because  $I_C = \beta_{DC}I_B$ . However, with the increase in  $I_C$ , the voltage drop across the collector resistor  $R_C$  increases. This results in the decrease of  $V_{CE}$ .

$$V_{CE} = V_{CC} - I_C R_C$$

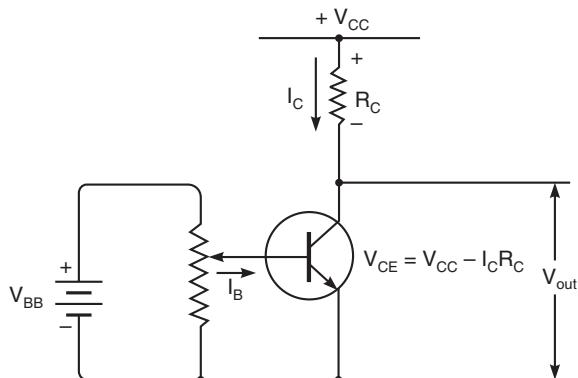


Fig. 37.21

When  $V_{CE}$  drops to knee voltage ( $V_{knee}$ ), any further increase in collector current is not possible since  $\beta_{DC}$  decreases sharply, when  $V_{CE}$  falls below knee voltage. This maximum current is called saturation collector current.

$$\therefore \text{Saturation collector current, } I_{C(\text{sat})} = \frac{V_{CC} - V_{knee}}{R_C}$$

### 37.8 SWITCHING ACTION OF A TRANSISTOR

The switching action of a transistor can be explained with the help of output characteristics. Fig. 37.22 shows the output characteristics of a typical transistor for a CE configuration.

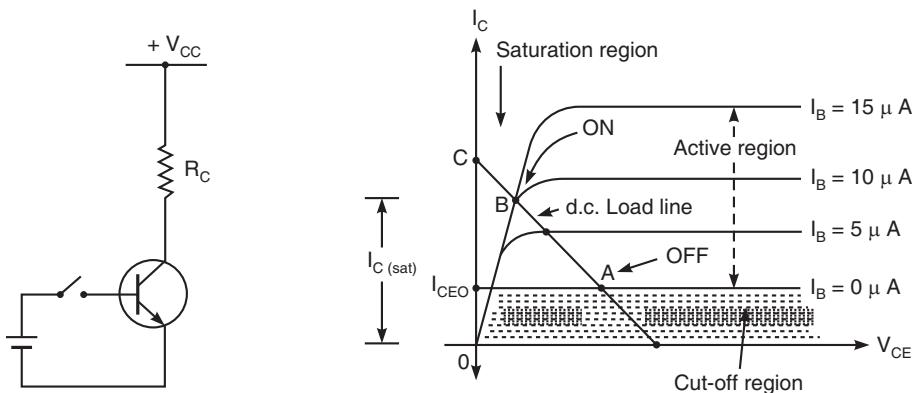


Fig. 37.22

- The load line is drawn for load  $R_C$  and collector supply  $V_{CC}$ .

The characteristic curves are arranged in three regions : OFF, ON (or saturation) and active regions.

(i) **OFF region.** When the input base voltage is zero or negative, the transistor is said to be in the OFF condition. In this condition,  $I_B = 0$ . The collector current is equal to the collector leakage current  $I_{CEO}$ . The value of  $I_{CEO}$  is obtained from the characteristics if we know  $V_{CE}$ .

$$\text{Power loss} = \text{Output voltage} \times \text{Output current}$$

In the OFF condition, the output voltage =  $V_{CC}$  since voltage drop in the load due to  $I_{CEO}$  is negligible.

$$\therefore \quad \text{Power loss} = V_{CC} \times I_{CEO}$$

Since  $I_{CEO}$  is very small as compared to full-load current that flows in the ON condition, power loss in the transistor is quite small in the OFF condition. Thus the efficiency of the transistor is high in the OFF condition.

(ii) **ON or saturation region.** When the input voltage is made so much positive that saturation collector current flows, the transistor is said to be in the ON condition.

The saturation collector current is given by:

$$I_{C(sat)} = \frac{V_{CC} - V_{knee}}{R_C}.$$

$$\text{Power loss} = \text{Output voltage} \times \text{Output current}$$

The output voltage in the ON condition is equal to  $V_{knee}$  and output current is  $I_{C(sat)}$

$$\therefore \quad \text{Power loss} = V_{knee} \times I_{C(sat)}$$

$V_{knee}$  is very small. So power loss is quite low.

Thus, the transistor efficiency is quite high in the ON condition.

- The OFF and ON regions are the stable regions of operation.

The collector current increases from  $I_{CEO}$  to  $I_{C(sat)}$  along the path  $AB$  as the transistor is switched ON. However, when the transistor is switched OFF, the collector current decreases from  $I_{C(sat)}$  to  $I_{CEO}$  along  $BA$ .

*Thus transistor in the saturation and cut off states behaves as a switch.*

### (iii) Active region

- This region exists above the curve for  $I_B$  equal to zero.
- It is the region that lies between OFF and ON conditions.

- It is the unstable (or transient) region through which the operation of the transistor passes while changing from OFF state to the ON state. The path  $AB$  is the active region.

- Electronic oscillators can be classified into two types:

- (1) Sinusoidal oscillators and
- (2) Non-sinusoidal oscillators.

The Hartley oscillator, Colpitt's oscillator, R-C phase shift oscillator and Wein bridge oscillator are examples of sinusoidal oscillators. They generate output voltage varying as a sine wave.

- Multivibrators* generate non-sinusoidal waves at the output such as square wave, rectangular wave or saw-tooth wave.

There are three types of multivibrators.

- (1) Astable multivibrator (free-running multivibrator)
- (2) Monostable multivibrator
- (3) Bistable multivibrator (Flip Flop)

### 37.9 ASTABLE MULTIVIBRATOR

An astable or free running multivibrator generates square waves of its own *i.e.*, without any external excitation. It has no stable state. It has only two quasistable states between which it keeps on oscillating of its own accord.

#### Circuit Details

Fig. 37.23 shows the circuit of an astable multivibrator.  $Q_1$  and  $Q_2$  are two identical transistors.

It is a two-stage  $RC$  coupled-amplifier. The output of each stage is coupled to the input of the other through a capacitor [ $C_1$  or  $C_2$ ]. Each stage will produce a phase shift of  $180^\circ$ . So the total phase shift between the input and output is  $360^\circ$ . Thus a positive feedback takes place and the circuit oscillates. The feedback is so strong that either the transistors are driven to saturation or to cut-off.

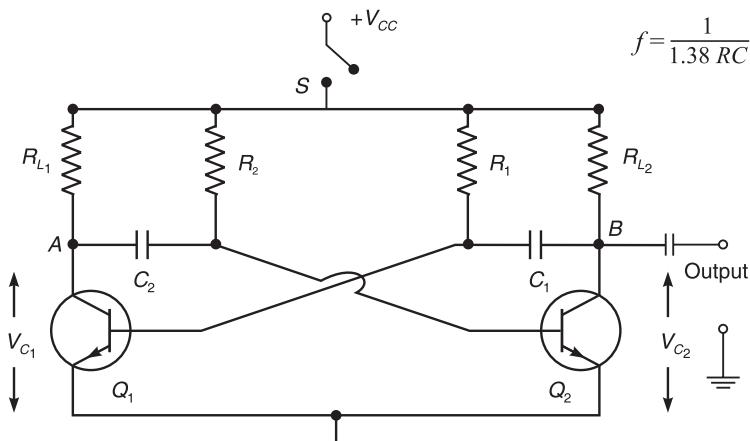


Fig. 37.23

#### Circuit operation

(i) When power  $V_{CC}$  is applied by closing switch  $S$ , collector current starts flowing in  $Q_1$  and  $Q_2$ . The coupling capacitors  $C_1$  and  $C_2$  start charging up. The characteristics of no two similar transistors are identical. Suppose  $Q_1$  conducts more than  $Q_2$ .

Then the collector current of  $Q_1$  will rise rapidly. This will result in a decrease of the collector voltage of  $Q_1$ . The resulting negative signal is applied to the base of  $Q_2$  through  $C_2$ . This drives

$Q_2$  towards cut-off. Consequently, the collector voltage of  $Q_2$  rises towards  $V_{CC}$ . This change in collector voltage of  $Q_2$  (positive going signal) is fed to the base of transistor  $Q_1$  through capacitor  $C_1$ . As a result of this positive going pulse, the collector current of  $Q_1$  is further increased. The process being cumulative, in a short time, transistor  $Q_1$  is saturated while  $Q_2$  is cut-off. These actions are so rapid and instantaneous that  $C_1$  does not get a chance to discharge. Under this situation, whole of  $V_{CC}$  drops across  $R_{L1}$  (since  $Q_1$  is saturated or is on ON state) i.e.,  $V_{C1} = 0$  and point A is at ground (or zero) potential. Also, since  $Q_2$  is cut-off (OFF state), there is no drop across  $R_{L2}$  and point B is at  $V_{CC}$ .

(ii) Capacitor  $C_2$  now begins to discharge which decreases the reverse bias on base of transistor  $Q_2$ . Ultimately a forward bias is re-stabilized at  $Q_2$ . Therefore,  $Q_2$  begins to conduct. Consequently, collector of  $Q_2$  becomes less positive. This negative going voltage signal is applied to the base of transistor  $Q_1$  through the capacitor  $C_1$ . As a result,  $Q_1$  is pulled out of saturation and is soon driven to cut-off. Simultaneously  $Q_2$  is driven to saturation. Now  $V_{C2}$  decreases and becomes almost zero volt when  $Q_2$  gets saturated. Consequently, potential of point B decreases from  $V_{CC}$  to almost zero volt.

The transistor  $Q_1$  remains cut off and  $Q_2$  in conduction until capacitor  $C_1$  discharges through  $R_1$ , enough to decrease the reverse bias of  $Q_1$ .

The whole cycle repeats.

The output of the multivibrator is taken from the collector of either transistor. The output is a *square wave* (Fig. 37.24) with a peak amplitude equal to  $V_{CC}$ .

#### Switching Times

The multivibrator circuit alternates between a state in which  $Q_1$  is ON and  $Q_2$  is OFF and a state in which  $Q_1$  is OFF and  $Q_2$  is ON.

ON time for  $Q_2$  (or OFF time for  $Q_1$ ) is

$$T_1 = 0.69 R_1 C_1 \quad \dots(1)$$

ON time for  $Q_1$  (or OFF time for  $Q_2$ ) is

$$T_2 = 0.69 R_2 C_2 \quad \dots(2)$$

Total period of the wave

$$T = T_1 + T_2 = 0.69 (R_1 C_1 + R_2 C_2) \quad \dots(3)$$

If  $R_1 = R_2 = R$  and  $C_1 = C_2 = C$  i.e., the two stages are symmetrical,

$$T = 0.69 (RC + RC) = 1.38 RC \quad \dots(4)$$

#### Frequency of oscillation

$$\text{Frequency of oscillation, } f = \frac{1}{T} = \frac{1}{1.38RC} = \frac{0.7}{RC}.$$

Here,  $f$  is in hertz,  $R$  in ohms and  $C$  in farads.

### 37.10 MONOSTABLE MULTIVIBRATOR

It has *one stable state* and *one quasi-stable state*. One transistor is always conducting (i.e., in the ON state) and the other is non-conducting (i.e., in the OFF state). A triggering pulse is required to induce a transition from the stable state to the quasi-stable state. After certain time, the multivibrator automatically switches back from quasi-stable state to the stable state and remains in the stable state until another pulse is applied.

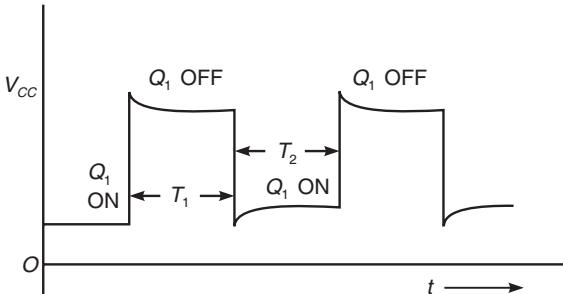
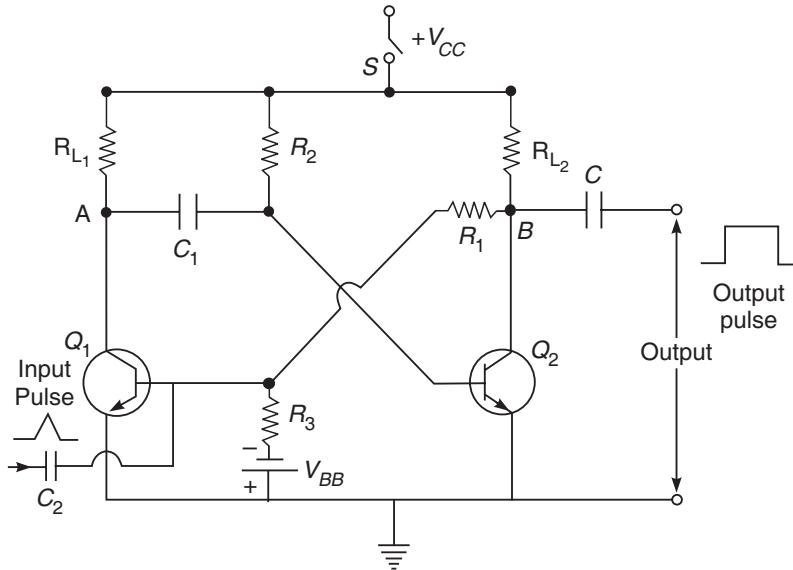


Fig. 37.24

**Circuit Diagram**

Fig. 37.25 shows the circuit diagram of monostable multivibrator.



**Fig. 37.25**

$Q_1$  and  $Q_2$  are two identical NPN transistors with equal collector loads i.e.,  $R_{L1} = R_{L2}$ ,  $V_{BB}$  and  $R_3$  reverse bias the transistor  $Q_1$  and keep it cut-off. The supply voltage  $+V_{cc}$  and resistor  $R_2$  forward bias the transistor  $Q_2$  and keep it at saturation. The input triggering pulse is applied to the base of  $Q_1$  through  $C_2$ . The output is taken from  $Q_2$ .

**Working :**

(i) Initially when switch  $S$  is closed and no triggering pulse is applied at  $C_2$ ,  $Q_1$  is cut-off and  $Q_2$  conducts at saturation. This is the stable state of the multivibrator.

(ii) When a positive trigger pulse of short duration and of sufficient amplitude is applied to the base of  $Q_1$  through  $C_2$ , it forward biases  $Q_1$ . Hence  $Q_1$  starts conducting. Consequently, collector voltage of  $Q_2$  falls due to voltage drop across  $R_{L1}$ . Thus potential of point  $A$  falls which is equivalent to a negative going voltage. This negative potential at the collector of  $Q_1$  is fed to the base of  $Q_2$  through capacitor  $C_1$ . Hence the forward bias of  $Q_2$  decreases. Now the collector current of  $Q_2$  decreases. Consequently, potential of point  $B$  increases due to lesser drop over  $R_{L2}$ .

The potential of  $B$  is applied to the base of  $Q_1$  through  $R_1$ . So the forward bias of  $Q_1$  increases. Now the transistor  $Q_1$  conducts more and more and the potential of point  $A$  approaches zero volt. Quickly  $Q_1$  is driven to saturation and  $Q_2$  to cut-off. This is the quasi-stable state.

(iii) With  $Q_1$  at saturation and  $Q_2$  at cut-off, the circuit will return to initial stable state (i.e.  $Q_2$  at saturation and  $Q_1$  at cut-off) after some time.

During the time when  $Q_1$  changes its state from 'OFF' to 'ON' the capacitor  $C_1$  is charged fully. Since point  $A$  is almost at zero-volt, capacitor  $C_1$  starts to discharge through saturated  $Q_1$  to ground. It decreases the negative potential at the base of  $Q_2$ . As  $C_1$  discharges further, transistor  $Q_2$  is pulled out of cut-off. As  $Q_2$  conducts further, a negative going voltage from point  $B$  via  $R_1$  drives  $Q_1$  into cut-off.

Thus the circuit returns to its initial stable state with  $Q_2$  conducting at saturation and  $Q_1$  cut-off.

To change this state, another pulse is to be applied.

The entire cycle repeats itself.

The width of the output pulse is determined by the time constant of  $C_1R_2$  and given by

$$T = 0.69 C_1 R_2.$$

*MMV* is used to generate new and sharp pulses from distorted and used pulses in computers and tele-communication systems.

### 37.11 BISTABLE MULTIVIBRATOR OR FLIP-FLOP CIRCUIT

It has *two absolutely stable states*. The circuit can stay in any one of the states indefinitely. When an external pulse is applied, the circuit changes from one stable state to another stable state. When another triggering pulse is applied only then it goes back to its original state.

The multivibrator is also known as *flip-flop*. The reason is that one triggering pulse causes the multivibrator to ‘flip’ from one stable state to another stable state while the second pulse causes it to ‘flop’ back to original stable state.

#### Circuit Diagram

Fig. 37.2(b) shows the circuit diagram of bistable multivibrator.

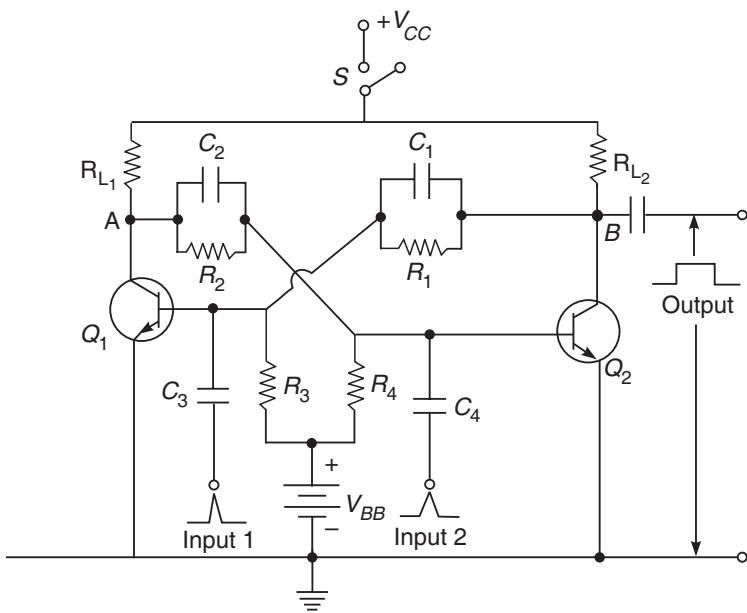


Fig. 37.26

It consists of two identical transistors  $Q_1$  and  $Q_2$  in common-emitter configuration, in such a way that the output of one is given as the input of the other. The feedback is given through the resistors  $R_1$ ,  $R_2$  shunted by capacitors  $C_1$ ,  $C_2$ . The main purpose of the capacitors  $C_1$  and  $C_2$  is to improve the switching speed of the circuit. The bias battery  $V_{BB}$  with resistors  $R_3$ ,  $R_4$  forward biases the bases of  $Q_1$  and  $Q_2$ . Suitable triggering pulses are applied to the bases of  $Q_1$  and  $Q_2$  through  $C_3$  and  $C_4$ . The output can be taken across either  $Q_1$  or  $Q_2$ .

#### Circuit Operation

(i) When the supply voltage  $V_{CC}$  is applied by closing switch  $S$ , collector current starts flowing in transistors  $Q_1$  and  $Q_2$ . Since no two transistors can have the same characteristics, one transistor, say  $Q_1$ , will conduct more rapidly than the other. Then the collector current of  $Q_1$  will rise at a faster rate causing a decrease in its collector voltage. The resulting negative signal is applied to the base of  $Q_2$  through  $C_2$  and drives it towards cut-off. Consequently, the collector voltage of  $Q_2$  rises towards

$V_{CC}$ . This positive going signal is fed to the base of  $Q_1$  through  $C_1$ . As a result, the collector current of  $Q_1$  is further increased. The process being cumulative, in a short time, transistor  $Q_1$  is saturated (ON) while  $Q_2$  is cut-off (OFF). This is the *first stable state* of the multivibrator. When the circuit is left free, it will remain in this stable state indefinitely.

(ii) The multivibrator is driven from the *first stable state* ( $Q_1$  ON and  $Q_2$  OFF) to the *second stable state* ( $Q_1$  OFF and  $Q_2$  ON) by applying either a negative trigger pulse to the base of transistor  $Q_1$  or a positive trigger pulse to the base of transistor  $Q_2$ .

Let a negative pulse be applied to the base  $Q_1$  through  $C_3$ . This reduces the forward bias of transistor  $Q_1$ . As a result, the collector current decreases or the collector voltage increases.

This rising collector voltage is fed to the base of  $Q_2$  where it forward biases the base-emitter junction of  $Q_2$ , increases its collector current and decreases collector voltage. This decreasing collector voltage of  $Q_2$  is fed to the base of  $Q_1$  where it reverse biases base-emitter junction of  $Q_1$  and decreases its collector current.

After few cycles,  $Q_2$  is driven to saturation (ON) and  $Q_1$  to cut off (OFF).

This is the *second stable state* of the multivibrator.

When the circuit is left free, it will remain in this stable state indefinitely, until another pulse is applied.

**Uses** (i) It is used as “Flip-Flop” in digital electronics to store one ‘bit’. The multivibrator is used in many digital operations such as counting and storing the binary information.

(ii) It is used in pulse generation.

(iii) It is used as a frequency divider in timing circuits.

## DIFFERENTIATING AND INTEGRATING CIRCUITS

### 37.12 DIFFERENTIATING CIRCUIT

A circuit in which the output voltage is proportional to the derivative of input voltage is known as differentiating circuit.

A differentiating circuit is a simple  $RC$  series circuit in which the output is taken across the resistor  $R$  as shown in Fig. 37.27.

The output voltage is directly proportional to the rate of change of the signal input, *i.e.*,

$$\text{output voltage } \propto \frac{d}{dt} (\text{input voltage})$$

$$V_{out} \propto \frac{dV_{in}}{dt} \quad \text{or} \quad V_{out} = RC \frac{dV_{in}}{dt}$$

The output will be zero for a constant input signal.

(i) *RC time constant*. The time constant  $RC$  must be many times less than the time period of the input, *i.e.*,  $RC \ll T$ . For good differentiation, the time constant  $RC$  of the circuit should be about one-tenth of the interval of the pulses to be differentiated.

(ii) The reactance offered by the capacitor ( $X_c$ ) must be large.

The AC input voltage can be represented by

$$V_i = V_m \sin \omega t \quad \dots(1)$$

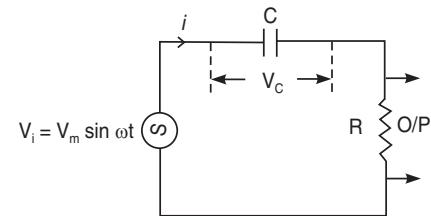


Fig. 37.27

At any instant, let  $i$  be the current in the circuit.

Applying Kirchoff's law to the  $RC$  circuit,

$$V_i = \text{Voltage across } R + \text{voltage across } C.$$

$$V_i = iR + \frac{1}{C} \int_0^t i dt \quad \left( \because V_C = \frac{q}{C} = \frac{1}{C} \int_C^t i dt \right)$$

$$\frac{V_i}{R} = i + \frac{1}{RC} \int_0^t i dt \quad \dots(2)$$

Since  $RC \ll T$ , the second term is much greater than the first term. So first term can be neglected.

$$\therefore \frac{V_i}{R} = \frac{1}{RC} \int_0^t i dt \quad \dots(3)$$

Differentiating Eq. (3) w.r.t. time, we get

$$\begin{aligned} \frac{1}{R} \frac{d}{dt}(V_i) &= \frac{1}{RC} i \\ RC \frac{d}{dt}(V_i) &= iR = V_0 \quad (\because iR = V_0) \end{aligned}$$

$$\text{or} \quad V_0 = RC \frac{d}{dt}(V_i) \quad \dots(4)$$

$$\therefore V_0 \propto \frac{d}{dt}(V_i) \quad (\because RC \text{ is constant})$$

Thus the output voltage is directly proportional to the derivative of the input voltage.

- (i) The output will be zero for a constant input signal.
- (ii) If the amplitude of the input changes abruptly, the differentiated wave is a sharp narrow pulse.
- (iii) For the less abrupt change in input, the output is a pulse approaching a very narrow rectangular form.
- (iv) For the constant rate of change in the input, the output has a constant value.
- (v) The waveform of the output depends upon the time constant  $RC$  of the circuit, and the frequency and the shape of the input voltage.

#### Output Waveforms

- (i) The input signal with a triangular waveform [Fig. 37.28 (a)] will produce a square wave [Fig. 37.28 (b)] output.

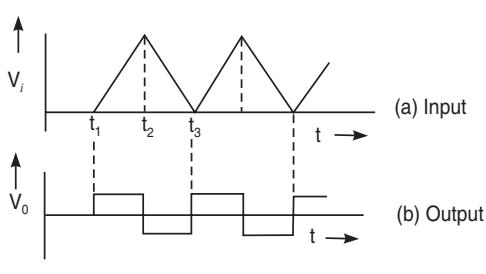


Fig. 37.28

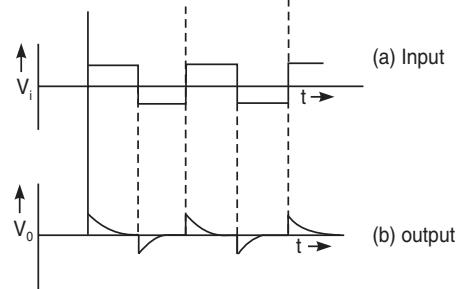


Fig. 37.29

(ii) The square wave input signal will produce a series of equally spaced pulses called spikes [Fig. 37.29].

(iii) The input signal with a sawtooth waveform will produce a rectangular wave output [Fig. 37.30].

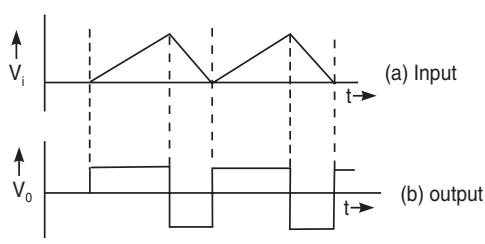


Fig. 37.30

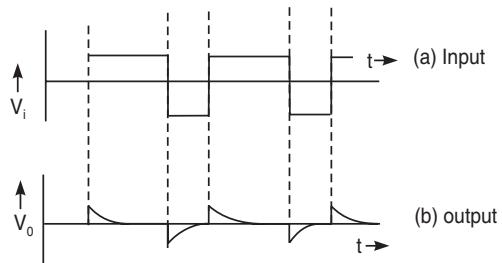


Fig. 37.31

(iv) The rectangular wave input signal will produce a series of spikes with alternately long and short spacing [Fig. 37.31].

### 37.13 INTEGRATING CIRCUIT

A circuit in which the output voltage is directly proportional to the integral of the input voltage is known as integrating circuit, i.e.,

$$\text{output voltage} \propto \int \text{input voltage} .$$

An integrating circuit is a simple  $RC$  circuit in which the output is taken across the capacitor  $C$ , as shown in Fig. 37.32.

The output voltage is directly proportional to the time integral of the input signal, i.e.,

$$V_{out} \propto \int V_{in} dt ,$$

$$\text{or } V_{out} = \frac{1}{RC} \int V_{in} dt .$$

The waveform of the output will depend upon the time constant and frequency and shape of the input signal. For good integration, the time constant  $RC$  of the circuit should be about ten times the interval of the pulses to be integrated. When the time constant of  $RC$  circuit is long, each succeeding pulse adds to the charge on the capacitor until it reaches an average value of the applied voltage. When the time constant is short, the capacitor charges and discharges so rapidly that the output waveform is nearly the same as the input waveform with the exception of the edges being rounded off.

The  $AC$  input voltage can be represented by

$$V_i = V_m \sin \omega t \quad \dots (1)$$

At any instant, let  $i$  be the current in the circuit.

Applying Kirchoff's law to the  $RC$  circuit,

$$V_i = \text{voltage across } R + \text{voltage across } C = iR + \frac{1}{C} \int_0^t i dt$$

$$\text{or } CV_i = RCi + \int_0^t i dt \quad \dots (2)$$

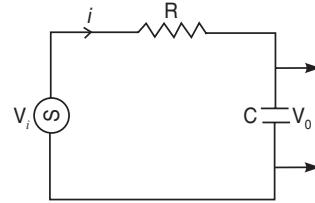


Fig. 37.32

Since  $RC \gg T$  (Time Period), the second term can be neglected.

$$\therefore CV_i = RCi \quad \dots(3)$$

Integrating Eq. (3) w.r.t.  $t$ , we get

$$C \int V_i dt = RC \int_0^t i dt \quad \dots(4)$$

But output voltage  $V_0$  = voltage across capacitor  $C = \frac{1}{C} \int_0^t i dt$

$$\text{or } \int_0^t i dt = C \cdot V_0 \quad \dots(5)$$

From Eqs. (4) and (5),

$$\begin{aligned} C \int V_i dt &= RC \cdot C V_0 \\ \text{or } V_0 &= \frac{1}{RC} \int V_i dt \\ \therefore V_0 &\propto \int V_i dt \quad \dots(6) \quad (\because RC \text{ is constant}) \end{aligned}$$

### Output Waveforms

(i) A square wave input will produce a triangular wave output (Fig. 37.33).

(ii) An input signal with equally spaced pulses will produce a square wave.

(iii) The rectangular wave input signal will produce a sawtooth wave output.

(iv) An input signal with un-equally spaced pulses will produce a rectangular wave output.

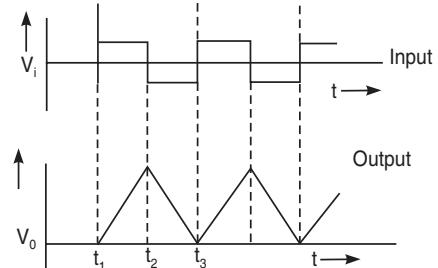


Fig. 37.33

### EXERCISE

1. Discuss Barhausen condition for oscillations. (B.U. 2010)
2. Explain with necessary theory working of a Hartley oscillator. (B.U. 2008)
3. In a Hartley oscillator the tank coil has two sections of inductances 80 mH and 20 mH. The capacitor has a capacitance of 500 pF. Neglecting the mutual inductance of the coil, find its frequency of oscillation.

[Sol.  $L_1 = 80 \times 10^{-3} H$ ,  $L_2 = 20 \times 10^{-3} H$ ,  $C = 500 \times 10^{-12} F$

$$\begin{aligned} f &= \frac{1}{2\pi\sqrt{C(L_1 + L_2)}} = \frac{1}{2 \times 3.14 \sqrt{500 \times 10^{-12} \times 100 \times 10^{-3}}} \\ &= 2.24 \times 10^4 \text{ Hz} \end{aligned}$$

4. Explain the working of a Colpitts oscillator with a circuit. Derive expressions for frequency and the condition for oscillation. (B.U. 2012)
5. In a Colpitts oscillator the inductance and capacitance used in the tuned circuit are 50 mH, and 100 pF and 400 pF. Calculate the frequency of oscillations.

[Sol.  $C_1 = 100 \times 10^{-12} F$ ,  $C_2 = 400 \times 10^{-12} F$ ,  $L = 50 \times 10^{-3} H$ .

$$f = \frac{1}{2\pi\sqrt{\frac{LC_1C_2}{C_1+C_2}}}$$

$$= \frac{1}{2 \times 3.14 \sqrt{\frac{50 \times 10^{-3} \times 100 \times 10^{-12} \times 400 \times 10^{-12}}{100 \times 10^{-12} + 400 \times 10^{-12}}}} = 8 \times 10^4 \text{ Hz}.$$

6. The two capacitors in a Colpitts oscillator have values of 0.2 and 0.02  $\mu\text{F}$ . The frequency of the oscillator is 10 KHz. Find the value of the inductor used in the circuit.  
**(Ans.** 13.93 mH)
7. With a circuit diagram, explain the working of a Wien-Bridge oscillator. Derive an expression for the frequency of oscillation. **(B.U. 2011)**
8. The frequency of a Wien bridge oscillator is 3 kHz. If the value of the resistor in the bridge network is 200 k $\Omega$ , find the value of the capacitors.

$$\text{Sol. } f = \frac{1}{2\pi RC} \text{ or } 3 \times 10^3 = \frac{1}{2\pi \times 200 \times 10^3 C}$$

$$\therefore \text{Capacitance} = C = 265.5 \text{ pF.}$$

9. With relevant circuit diagram discuss the principle of operation of a phase shift oscillator using transistor. Obtain expression for its frequency and for the condition for the circuit to oscillate.  
**(Calicut, 1989)**

10. In the phase shift oscillator,  $R_1 = R_2 = R_3 = 1\text{M}\Omega$  and  $C_1 = C_2 = C_3 = 68 \text{ PF}$ . At what frequency does the circuit oscillate? **(B.U. 2015)**

$$[\text{Ans. } f_0 = \frac{1}{2\pi RC\sqrt{6}} = \frac{1}{2\pi \times 10^6 \times (68 \times 10^{-12})\sqrt{6}} = 954 \text{ Hz}]$$

11. What is piezoelectric effect? Give an equivalent electrical circuit of a vibrating piezoelectric crystal. Draw a circuit of a crystal oscillator. Explain its operation. **(C.U. 1987)**
12. Explain the collector leakage current of a transistor. **(B.U. 2012)**
13. Explain the term ‘saturation collector current’. **(B.U. 2015)**
14. What is a Switching circuit? Explain the switching action of a transistor with the help of output characteristics. **(B.U. 2010)**
15. What are the different types of multivibrators? **(Madras 1995)**
16. Describe the working of astable multivibrator with circuit diagram. Derive the expression for frequency of astable multivibrator. **(B.U. 2010)**
17. In an astable multivibrator, the value of  $R_1 = R_2 = 15 \text{ k}\Omega$  and  $C_1 = C_2 = 0.005 \mu\text{F}$ . Calculate the frequency of oscillation. **(Andhra 1995)**

$$[f = \frac{1}{1.38RC} = \frac{1}{1.38 \times (15 \times 10^3) \times (0.005 \times 10^{-6})} = 9.66 \times 10^3 \text{ Hz}]$$

18. Draw the circuit of monostable multivibrator and explain its working.
19. Draw the circuit diagram of transistor bistable multivibrator and explain its working. **(B.U. 2015)**
20. Describe the action of differentiating circuit. Show that the output from a differentiating circuit is derivative of the input.
21. Sketch the output waveforms from a differentiating circuit when input is (i) a square wave (ii) saw-tooth wave. **(B.U. 2015)**
22. Define integrating circuit. Show that the output voltage is directly proportional to the integral of the input. Draw the output wave form. **(B.U. 2015)**

## JFET, MOSFET, SCR AND UJT

**After reading this chapter, you should be able to**

- ◆ Identify and describe the significant regions of JFET drain curves and transconductance curves.
- ◆ Discuss JFET applications (FET amplifier, FET as VVR)
- ◆ Explain the characteristics and operation of both depletion-mode and enhancement-mode MOSFETs
- ◆ Explain the operation, V-I characteristics, parameters, and circuit applications of SCR
- ◆ Explain the operation of UJT
- ◆ Analyse the operation of a UJT relaxation oscillator

## FIELD-EFFECT TRANSISTORS

### JFET (junction field-effect transistor)

#### 38.1 INTRODUCTION

The field-effect transistor (FET) is a three terminal semiconductor device in which the output current is controlled by an applied electric field. Further the current in a FET is entirely due to the majority carriers whereas in a junction transistor both majority and minority carriers contribute to the current. Thus, while the junction transistor is bipolar, a FET is unipolar. There are two types of field-effect transistors : (i) *junction field effect transistor* (JFET or simply FET). (ii) *metal oxide semiconductor field effect transistor* (MOSFET). A JFET can be either of the *n*-channel type or of the *p*-channel type. We shall here describe an *n*-channel JFET.

**n-channel JFET.** It consists of a channel (*n*-type) into which two *p*-regions are diffused. One end of this symmetric structure is called *source* (*S*) and the other end is called *drain* (*D*). The two *p*-regions are connected together to a third terminal called *gate* (*G*) (Fig. 38.1). The *p*-regions are heavily doped compared to the *n*-region. During operation, majority carriers (electrons in this case) enter the channel through the source *S* and leave it through the drain *D*. The current is controlled by the gate which is always *reverse-biased*.

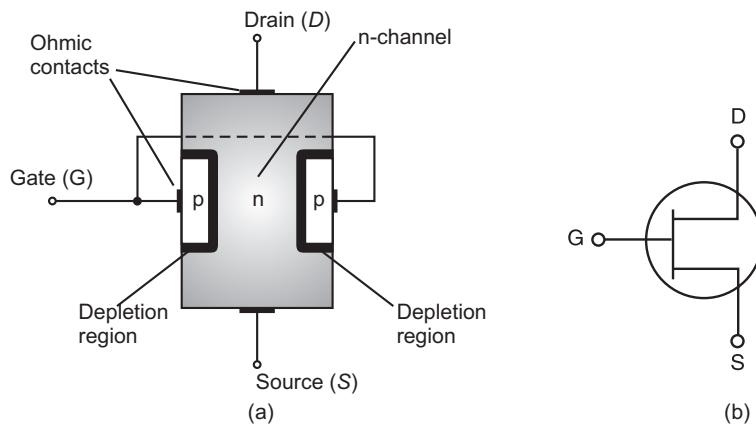


Fig. 38.1

#### 38.2 DRAIN CHARACTERISTICS OF AN N-CHANNEL JFET

A curve plotted between drain current  $i_D$  and drain-to-source voltage  $V_{DS}$  at a fixed gate-to-source voltage  $V_{GS}$  is called the drain characteristics. Fig. 38.2 shows the circuit diagram for determining the output characteristics.  $V_{GG}$  is the gate bias supply and  $V_{DD}$  is the drain voltage source.

Keeping  $V_{GS}$  fixed at some value,  $V_{DS}$  is changed in steps and the corresponding drain current  $I_D$  is noted. A group of such drain characteristics curves are drawn by setting  $V_{GS}$  at different fixed

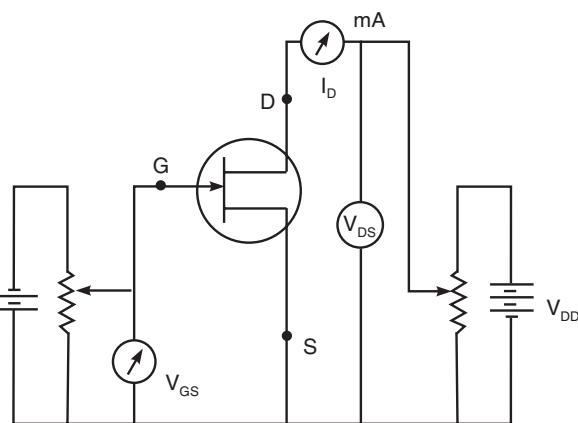


Fig. 38.2

values. Fig. 38.3 shows a family of drain characteristics. There are three distinct regions in the characteristic thus obtained.

(i) When  $V_{DS}$  is small, the channel acts as a resistor. The current increases linearly with the voltage  $V_{DS}$  till point A is reached. This region of the characteristic is called the *ohmic region*.

(ii) When  $V_{DS} = V_P$ , the current  $I_D$  reaches its maximum value,  $I_{DSS}$ . If  $V_{DS}$  is increased beyond  $V_P$ , the current does not increase any further. The region BC is called saturation region or *pinch off region*.

(iii) At a certain voltage  $V_A$ , corresponding to point C, current increases suddenly due to avalanche breakdown. The covalent bonds in the depletion region break up and the current rises. This region is called the *avalanche region*. In actual practice this region is to be avoided.

**Characteristic parameters of FET.** There are three main characteristic parameters of a FET which describe its performance in an electronic circuit.

(i) **Drain resistance  $r_d$ .** It is defined as the ratio of the change in drain-to-source voltage to the corresponding change in drain current at a constant gate-to-source voltage.

$$r_d = \left( \frac{\Delta V_{DS}}{\Delta I_D} \right)_{V_{GS}} \quad \dots(i)$$

It is given by the reciprocal of the slope of the drain characteristic.

(ii) **Transconductance  $g_m$ .** It is defined as the ratio of the change in drain current to the corresponding change in gate-to-source voltage at a constant drain-to-source voltage.

$$g_m = \left( \frac{\Delta I_D}{\Delta V_{GS}} \right)_{V_{DS}} \quad \dots(ii)$$

The transconductance measures the control that the gate voltage has over the drain current.

(iii) **Amplification factor  $\mu$ .** It is defined as the ratio of the change in drain-to-source voltage to the corresponding change in gate-to-source voltage at a constant drain current.

$$\mu = \left( \frac{\Delta V_{DS}}{\Delta V_{GS}} \right)_{I_D} \quad \dots(iii)$$

#### Relation between the three parameters.

We have from the definition of the amplification factor

$$\mu = \frac{\Delta V_{DS}}{\Delta V_{GS}} = \frac{\Delta V_{DS}}{\Delta I_D} \times \frac{\Delta I_D}{\Delta V_{GS}} = r_d \times g_m$$

$$\therefore \mu = r_d \times g_m$$

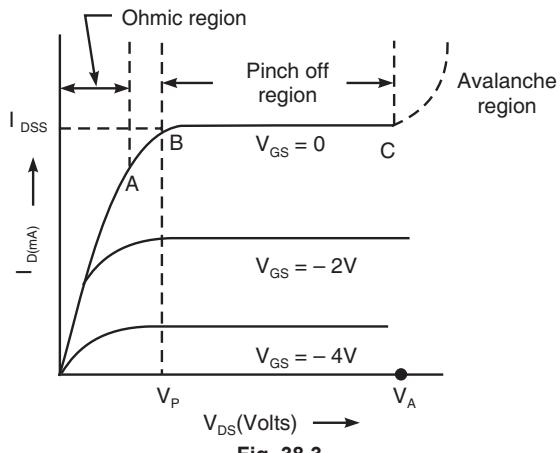


Fig. 38.3

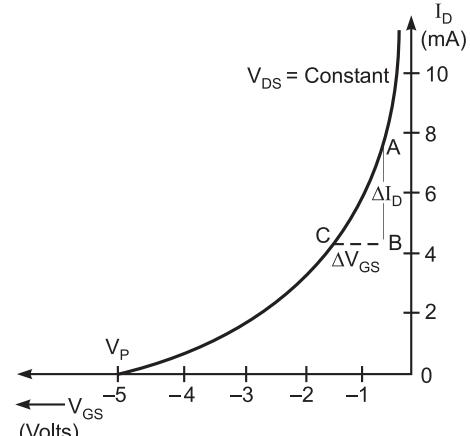


Fig. 38.4

### (ii) Transfer characteristics

The graph drawn between drain current  $I_D$  and gate to source voltage  $V_{GS}$  at constant  $V_{DS}$  is known as *Transfer characteristics* (Fig. 38.4).

For a constant value of  $V_{DS}$ ,  $I_D$  is noted by varying  $V_{GS}$ . The value of  $V_{GS}$  (-ve) is varied till  $I_D$  becomes zero. This particular voltage  $V_P$  where  $I_D = 0$  is called *pinch-off voltage*. The slope of the straight line portion of the graph gives the *transconductance*  $g_m$ .

$$\text{Transconductance } g_m = \left. \frac{\Delta I_D}{\Delta V_{GS}} \right|_{V_{DS}}$$

The transconductance is determined from the transfer characteristic curve.

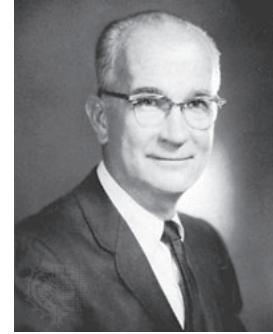
$$g_m = \frac{AB}{BC} = \text{Slope of the curve.}$$

The relationship between  $I_D$  and  $V_{GS}$  is defined by *Shockley's equation*.

$$I_D = I_{DSS} \left(1 - \frac{V_{GS}}{V_p}\right)^2$$

**William Bradford Shockley** (1910-1989), co-inventor of the first transistor and formulator of the "field-effect" theory employed in the development of transistor and the FET.

(Photo courtesy of AT & T Archives.)



The squared term in the equation results in a nonlinear relationship between  $I_D$  and  $V_{GS}$ , producing a curve that grows exponentially with decreasing magnitude of  $V_{GS}$ .

## 38.3 PRINCIPLE OF N-CHANNEL FET OPERATION

### Case (i): When $V_{GS} = 0$ and $V_{DS} = 0$

When no voltages are applied between  $D$  and  $S$  and  $G$  and  $S$ , the thickness of the depletion regions around the  $p-n$  junctions is uniform. So a rectangular channel with uniform cross-section is formed [Fig. 38.1 (a)].

### Case (ii): $V_{GS} = 0$ V, $V_{DS}$ Some Positive Value

A positive voltage  $V_{DS}$  is applied across the channel and the gate is connected directly to the source to establish the condition  $V_{GS} = 0$  V [Fig. 38.5 (a)].

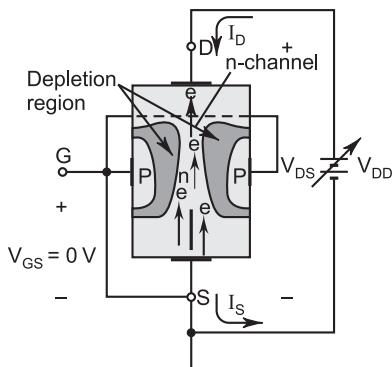


Fig. 38.5 (a)

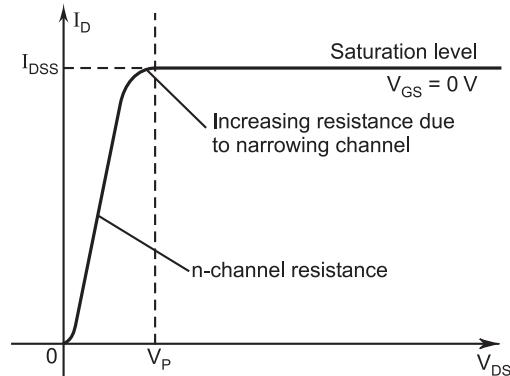


Fig. 38.5 (b)

The result is a gate and a source terminal at the same potential and a depletion region in the low end of each  $p$ -material similar to the distribution of the no-bias conditions of Fig. 38.1 (a). The

instant the voltage  $V_{DD}$  ( $= V_{DS}$ ) is applied, the electrons are drawn to the drain terminal, establishing the conventional current  $I_D$  with the defined direction of Fig. 38.5 (a). The path of charge flow reveals that the drain and source currents are equivalent ( $I_D = I_S$ ). The flow of charge is relatively uninhibited and is limited solely by the resistance of the  $n$ -channel between drain and source. The depletion region is wider near the top of both  $p$ -type materials.

- As the voltage  $V_{DS}$  is increased from 0 V to a few volts, the current will increase as determined by Ohm's law. The plot of  $I_D$  versus  $V_{DS}$  will appear as shown in Fig. 38.5 (b). The relative straightness of the plot reveals that for the region of low values of  $V_{DS}$ , the resistance is essentially constant. As  $V_{DS}$  increases and approaches a level referred to as  $V_P$ , the depletion regions will widen, causing a noticeable reduction in the channel width. The reduced path of conduction causes the resistance to increase. The more horizontal the curve, the higher the resistance, suggesting that the resistance is approaching "infinite" ohms in the horizontal region. If  $V_{DS}$  is increased to a level where it appears

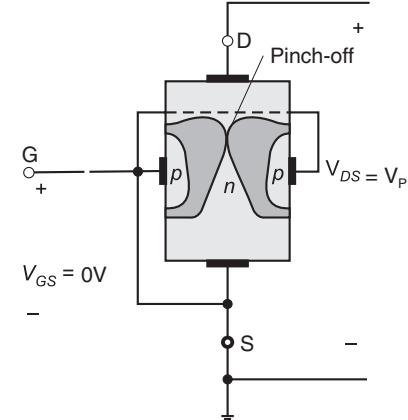


Fig. 38.5 (c)

*Pinch-off* ( $V_{GS} = 0$  V,  $V_{DS} = V_P$ ).

that the two depletion regions would "touch" as shown in Fig. 38.5 (c), a condition referred to as *pinch-off* will result. The level of  $V_{DS}$  that establishes this condition is referred to as the *pinch-off voltage* and is denoted by  $V_P$ .

- As  $V_{DS}$  is increased beyond  $V_P$ , the region of close encounter between the two depletion regions increases in length along the channel, but the level of  $I_D$  remains essentially the same.
- $I_{DSS}$  is the maximum drain current for a JFET and is defined by the conditions

$$V_{GS} = 0 \text{ V} \quad \text{and} \quad V_{DS} > |V_P|.$$

#### Case (iii): $V_{GS} < 0$ V

The controlling voltage  $V_{GS}$  is made more and more negative from its  $V_{GS} = 0$  V level. In other words, the gate terminal will be set at lower and lower potential levels as compared to the source. In Fig. 38.5 (d) a negative voltage of  $-2$  V is applied between the gate and source terminals for a low level of  $V_{DS}$ . The effect of the applied negative-bias  $V_{GS}$  is to establish depletion regions similar to those obtained with  $V_{GS} = 0$  V, but at lower levels of  $V_{DS}$ . Therefore, the result of applying a negative bias to the gate is to reach the saturation level at a lower level of  $V_{DS}$ , as shown in Fig. 38.3 for  $V_{GS} = -2$  V. The resulting saturation level for  $I_D$  has been reduced and will continue to decrease as  $V_{GS}$  is made more and more negative.

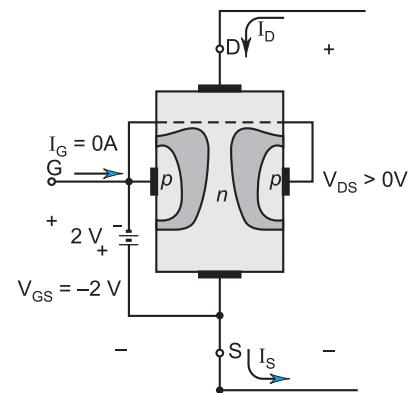


Fig. 38.5 (d)

### FET AMPLIFIER

#### 38.4 COMMON SOURCE FET AMPLIFIER

Fig. 38.6 shows the common source amplifier circuit.  $V_{GG}$  is the gate bias battery. The ac signal to be amplified is applied between the gate and the source. The load resistance  $R_L$  is connected

between the supply battery  $V_{DD}$  and the drain  $D$ . The output is taken across the load  $R_L$  or across the FET. Let the output ac voltage be  $v_{ds}$  for an input signal voltage  $v_{gs}$ .

**Working.** On applying an alternating signal, the fixed reverse bias of the gate changes. A small change in the reverse bias gate voltage, produces a large change in drain current. This fact makes FET capable of amplifying weak signal. The amplifying action of FET can be explained using its transfer characteristics (Fig. 38.7).

When the signal ( $v_{gs}$ ) is not applied, the gate to source voltage is given by  $OA$ . The corresponding drain current is given by  $OF$ . When the signal is applied, during the positive half cycle of the input signal ( $+v_{gs}$ ), the reverse bias on the gate decreases ( $-V_{GG} + v_{gs}$ ). The gate to source voltage decreases from  $OA$  to  $OC$ . Due to this, the channel width increases. Hence, drain current increases from  $OF$  to  $OG$ . During the negative half of the signal ( $-v_{gs}$ ), the reverse voltage on the gate increases ( $-V_{GG} - v_{gs}$ ) to  $OB$ . Due to this the channel width decreases. Hence the drain current decreases to  $OE$ . Thus, a small change in the gate to source voltage, produces a large change in drain current and hence large change in output voltage. Thus, FET works as an amplifier.

When the signal voltage is positive, the gate becomes less negative with respect to the source. The drain current is enhanced, causing a large voltage drop across  $R_L$  which makes the drain terminal less positive with respect to the source. Since an increase of the gate voltage causes a decrease in the drain voltage, there is a phase shift of  $180^\circ$  between the input and output of the FET amplifier.

#### Expression for Voltage Gain

The a.c. voltage source equivalent circuit of FET amplifier is shown in Fig. 38.8. FET can be replaced by a voltage generator  $-\mu v_{gs}$  in series with drain resistance  $r_{ds}$ .  $i_d$  is the ac current flowing through the load  $R_L$ . The output ac voltage is  $v_{ds}$  for an input signal voltage  $v_{gs}$ .

Applying Kirchoff's voltage law to the circuit,

$$-\mu v_{gs} = i_d R_L + i_d r_{ds}$$

$$\therefore i_d = \frac{-\mu v_{gs}}{r_{ds} + R_L}.$$

But

$$v_{ds} = i_d \cdot R_L$$

$$i_d = \frac{v_{ds}}{R_L}$$

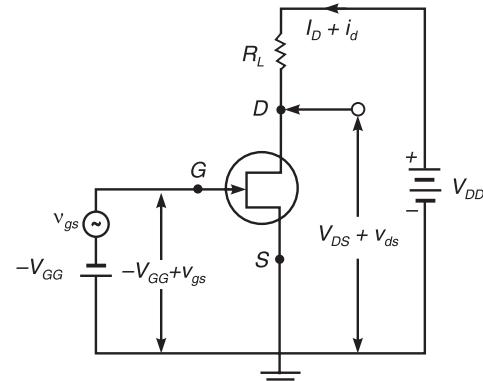


Fig. 38.6

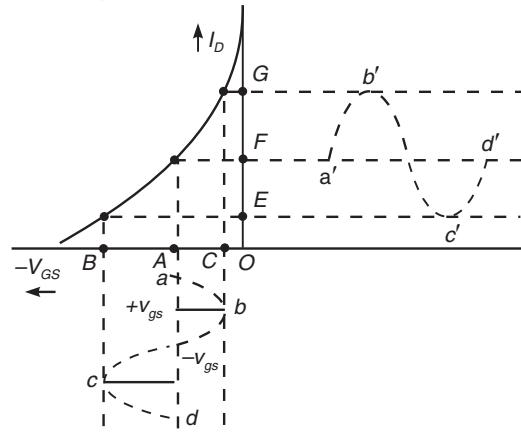


Fig. 38.7

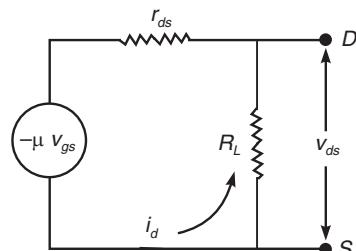


Fig. 38.8

$$\therefore \frac{v_{ds}}{R_L} = \frac{-\mu v_{gs}}{r_{ds} + R_L}$$

$$\frac{v_{ds}}{v_{gs}} = \frac{-\mu R_L}{r_{ds} + R_L}$$

Voltage gain       $A_v = \frac{v_{ds}}{v_{gs}} = \frac{-\mu R_L}{r_{ds} + R_L}$

This is the expression for voltage gain. The negative sign indicates that there is a phase difference of  $180^\circ$  between the input and the output voltages.

#### Advantages of JFET over BJT

BJT	FET
1. Bipolar	Unipolar
2. Input impedance low	Input impedance high
3. Current controlled	Voltage controlled
4. Characterised by current gain	Characterised by transconductance
5. Noise level not small	Noise level very small

**EXAMPLE 1.** A common-source FET amplifier has a load resistance  $R_L = 500 \text{ k}\Omega$ . If the a.c. drain resistance ( $r_d$ ) and amplification factor ( $\mu$ ) of the FET are  $100 \text{ k}\Omega$  and 24, respectively, calculate the voltage gain of the amplifier.

**SOL.**  $|A_v| = \frac{\mu R_L}{r_{ds} + R_L} = \frac{24 \times 500 \times 10^3}{100 \times 10^3 + 500 \times 10^3} = 20.$

**EXAMPLE 2.** For a constant drain-to-source voltage if the gate-source voltage is changed from 0 to  $-2 \text{ V}$ , the corresponding change in drain current becomes  $2 \text{ mA}$ . Calculate the transconductance of the FET. If the a.c. drain resistance is  $100 \text{ k}\Omega$ , calculate also the amplification factor of the FET.

**SOL.** The transconductance  $g_m$  is given by

$$g_m = \left( \frac{\Delta I_D}{\Delta V_{GS}} \right)_{V_{DS}} = \frac{2 \text{ mA}}{2 \text{ V}} = 1 \text{ mA/V.}$$

Again,  $\mu = g_m r_{ds} = 1 \times 10^{-3} \times 100 \times 10^3 = 100.$

#### The Equivalent Circuit for the FET (Low frequency model)

Fig. 38.9 (a) shows the internal elements of a FET arranged as a two-port circuit. We can reduce it to a simpler equivalent form as in (b) by consideration of the relative magnitudes of the resistances and the reactances.

The input element  $r_{gs}$  is the reverse-biased junction resistance of a JFET or the silicon-dioxide insulation resistance of a MOSFET. These resistances exceed  $10^8$  ohms. We will consider the input of a FET as an open circuit.

The series circuit of  $C_c$  and  $r_c$  represents the capacitance between gate and channel and the series resistance of the channel. Being only a few picofarads and a few hundred

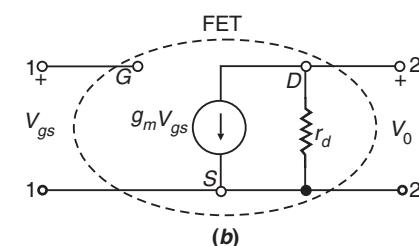
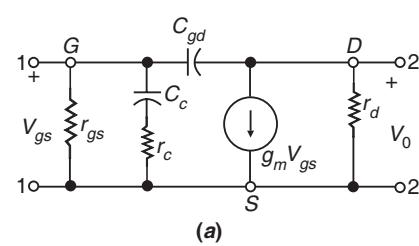


Fig. 38.9

(a) Internal FET circuit;  
(b) Equivalent circuit for a FET.

ohms, the effect of this  $RC$  combination is negligible at low frequencies. So, they are eliminated from the equivalent circuit.

Capacitance  $C_{gd}$  is the gate-to-drain capacitance and includes the capacitance of the mounting. Being only 1 to 3 pF,  $C_{gd}$  may be eliminated from the equivalent circuit.

FET ac equivalent circuit is shown in Fig. 38.9 (b). The control of  $I_d$  by  $V_{gs}$  is included as a current source  $g_m V_{gs}$  connected from drain to source. The arrow points from drain to source to show a phase change of  $180^\circ$  between output and input voltages as will occur in actual operation.

Here gate to source voltage is now represented by  $V_{gs}$  (lower-case subscripts) to distinguish it from d.c. levels. In low frequency model as the gate is reversed biased, the gate current becomes zero. Therefore, the input impedance is represented by open circuit at input terminals. The output impedance is represented by the resistor  $r_d$  from drain to source. In situations, where  $r_d$  is sufficiently large in comparison to other circuit elements, this is ignored. Now, the equivalent circuit is simply a current source whose magnitude is controlled by  $V_{gs}$  and  $g_m$ -clearly a voltage controlled device.

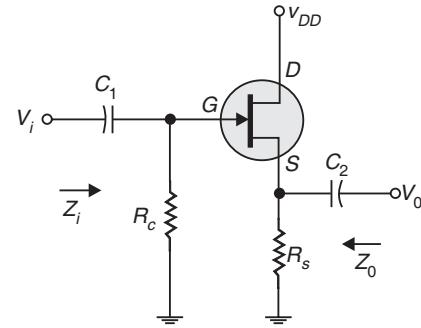


Fig. 38.10

### 38.5 JFET SOURCE-FOLLOWER (COMMON-DRAIN) AMPLIFIER

Fig. 38.10 shows the JFET source-follower configuration.

The input is applied between gate and source. The output is taken off the source terminal. When the d.c. supply is replaced by its short-circuit equivalent, the drain is grounded (hence, the terminology common-drain).

Fig. 38.11 shows the low frequency equivalent model for the common drain amplifier circuit shown in Fig. 38.10.

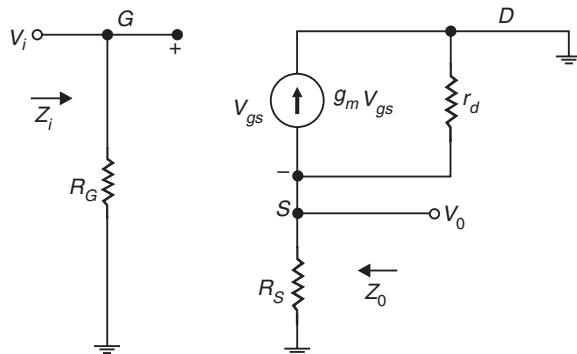


Fig. 38.11

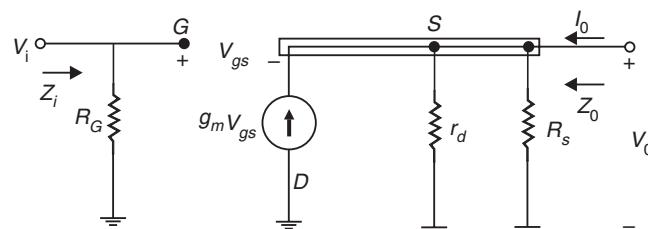


Fig. 38.12

Fig. 38.11 can be simplified as shown in Fig. 38.12.

**Input Impedance  $Z_i$ :** Fig. 38.12 reveals that  $Z_i$  is defined by

$$Z_i = R_G \quad \dots(1)$$

**Output Impedance  $Z_0$ :** Fig. 38.13 shows the circuit for  $Z_0$  determination. The output impedance of the circuit is found by making the independent signal source,  $V_i$ , equal zero. Setting  $V_i = 0$  V will result in the gate terminal being connected directly to ground.

$V_{gs}$  and  $V_0$  are across the same parallel network.

$$\therefore V_0 = -V_{gs} \quad \dots(2)$$

Applying Kirchhoff's current law at node  $S$ ,

$$I_0 + g_m V_{gs} = I_{rd} + I_{Rs} = \frac{V_0}{r_d} + \frac{V_0}{R_s}$$

$$I_0 = V_0 \left[ \frac{1}{r_d} + \frac{1}{R_s} \right] - g_m V_{gs}$$

$$= V_0 \left[ \frac{1}{r_d} + \frac{1}{R_s} \right] - g_m [-V_0] = V_0 \left[ \frac{1}{r_d} + \frac{1}{R_s} + g_m \right]$$

$$Z_0 = \frac{V_0}{I_0} = \frac{V_0}{V_0 \left[ \frac{1}{r_d} + \frac{1}{R_s} + g_m \right]} = \frac{1}{\frac{1}{r_d} + \frac{1}{R_s} + g_m}$$

$$= \frac{1}{\frac{1}{r_d} + \frac{1}{R_s} + \frac{1}{g_m}}$$

$$\therefore Z_0 = r_d \parallel R_s \parallel 1/g_m \quad \dots(3)$$

$$\text{For } r_d \geq 10 R_s, \quad Z_0 \cong R_s \parallel 1/g_m \quad \dots(4)$$

### Expression for Voltage Gain

The output voltage  $V_0$  is determined by

$$V_0 = g_m V_{gs} (r_d \parallel R_s) \quad \dots(5)$$

Applying KVL around the perimeter of the network of Fig. 38.12,

$$V_i = V_{gs} + V_0$$

$$\text{or} \quad V_{gs} = V_i - V_0$$

$$\text{Eq. (5) becomes} \quad V_0 = g_m(V_i - V_0)(r_d \parallel R_s)$$

$$\text{or} \quad V_0 = g_m V_i (r_d \parallel R_s) - g_m V_0 (r_d \parallel R_s)$$

$$\text{or} \quad V_0 [1 + g_m(r_d \parallel R_s)] = g_m V_i (r_d \parallel R_s)$$

$$\therefore A_V = \frac{V_0}{V_i} = \frac{g_m (r_d \parallel R_s)}{1 + g_m (r_d \parallel R_s)} \quad \dots(6)$$

$$\text{If } r_d \geq 10 R_s, \quad A_V = \frac{V_0}{V_i} \cong \frac{g_m R_s}{1 + g_m R_s} \quad \dots(7)$$

If  $g_m R_s \gg 1$        $A_V = 1$ , but it is always less than one.

We observe that common drain circuit does not provide voltage gain.

**Phase Relationship:** Since  $A_V$  of Eq. (6) is a positive quantity,  $V_0$  and  $V_i$  are in phase for the JFET source-follower configuration.

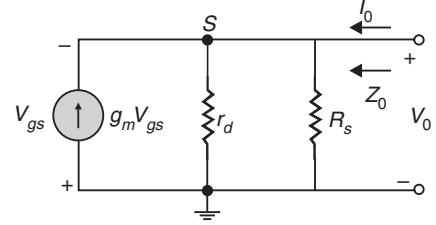


Fig. 38.13

### 38.5.1. Voltage-Series Feedback FET Amplifier

Fig. 38.14 shows an FET amplifier stage with voltage-series feedback. A part of the output signal ( $V_0$ ) is obtained using a feedback network of resistors  $R_1$  and  $R_2$ . The feedback voltage  $V_f$  is connected in series with the source signal  $V_s$ .

Their difference is the input signal  $V_i$ .

Without feedback the amplifier gain is

$$A = \frac{V_0}{V_i} = -g_m R_L \quad \dots(1)$$

Here  $R_L$  is the parallel combination of resistors:

$$R_L = R_D R_0 (R_1 + R_2) \quad \dots(2)$$

The feedback network provides a feedback factor of

$$\beta = \frac{V_f}{V_0} = \frac{-R_2}{R_1 + R_2} \quad \dots(3)$$

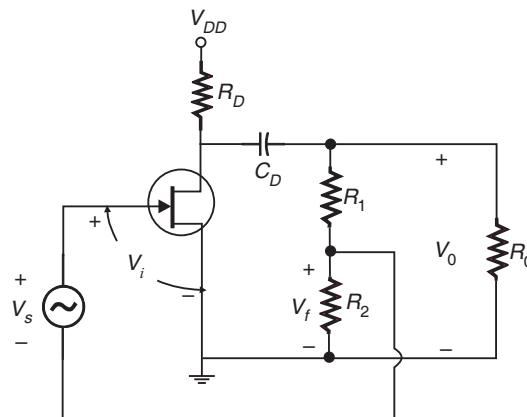


Fig. 38.14

The gain with negative feedback for the FET amplifier is

$$A_f = \frac{A}{1 + \beta A} = \frac{-g_m R_L}{1 + [R_2 R_L / (R_1 + R_2)] g_m} \quad \dots(4)$$

If  $\beta A \gg 1$ , we have

$$A_f \approx \frac{1}{\beta} = -\frac{R_1 + R_2}{R_2} \quad \dots(5)$$

## 38.6 FET AS VOLTAGE-VARIABLE RESISTOR (VVR)

**Principle.** JFET is operated in the constant-current portion of its output characteristics *i.e.*, in saturation region for the linear applications. FET can also be used in the region before pinch-off where  $V_{DS}$  is small. FET when used in the region before pinch-off, it works as variable resistance device *i.e.*, the channel resistance  $r_d$  is controlled by the gate bias voltage ( $V_{GS}$ ).

In JFET, the drain-to-source conductance  $g_d = \frac{I_D}{V_{DS}}$  for small values of  $V_{DS}$ . It may also be

expressed as,

$$g_d = g_{d0} \left[ 1 - \left( \frac{V_{GS}}{V_P} \right)^{\frac{1}{2}} \right]$$

where  $g_{d0}$  is the value of drain conductance when the bias voltage  $V_{GS}$  is zero.

The variation of the  $r_d$  with  $V_{GS}$  can be closely approximated by the empirical expression,

$$r_d = \frac{r_0}{1 - KV_{GS}}$$

where  $r_0$  = drain resistance at zero gate bias, and  $K$  = a constant, dependent upon FET type.

Thus, small signal FET drain resistance  $r_d$  varies with applied gate voltage  $V_{GS}$  and FET acts like a variable passive resistor.

FET finds wide applications where VVR property is useful. For example, the VVR can be used in *Automatic Gain Control* (AGC) circuit of a multistage amplifier.

The VVR is used to vary the voltage gain of a multistage amplifier. If the signal is low then voltage gain of the stages can be increased and when the signal becomes high, the gain can be reduced automatically. In this way, the general level of amplification is maintained fairly constant.

**Circuit.** Fig. 38.15 shows the circuit arrangement of AGC amplifier using the FET as *Voltage Variable Resistor* (VVR) or *Voltage Dependent Resistor* (VDR).

**Working.** The input signal  $v_i$  is amplified by amplifier. It is then rectified and filtered to produce a d.c. voltage proportional to output signal level. This voltage is applied to the gate of the FET so that the a.c. resistance between drain and source changes.

The capacitor  $C$  isolates the transistor from FET so that the bias conditions of transistor are not affected. So, when output increases,  $V_{GS}$  also increases and  $R_{DS}$  changes so that the gain of the transistor decreases. Thus automatically the gain is controlled.

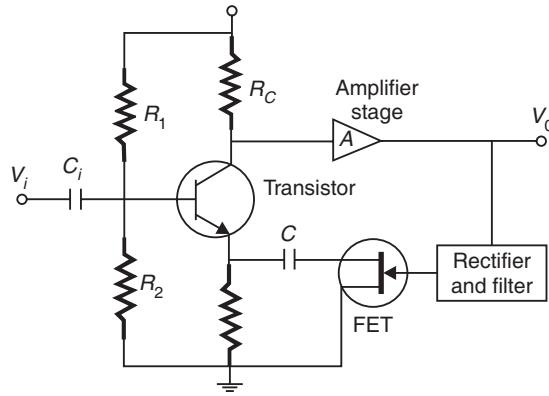


Fig. 38.15

### METAL OXIDE SEMICONDUCTOR FIELD EFFECT TRANSISTOR (MOSFET)

There are two types of MOSFETs.

- (i) The Depletion MOSFET.
- (ii) The Enhancement MOSFET.

In both types, *p*-channel and *n*-channel versions exist.

## 38.7 THE DEPLETION MOSFET

### Construction

Fig. 38.16 shows the construction of an *n*-channel depletion MOSFET. It consists of a lightly doped *p*-type substrate into which two heavily-doped *n*<sup>+</sup>-regions are diffused. These two *n*<sup>+</sup> regions act as source *S* and drain *D*.

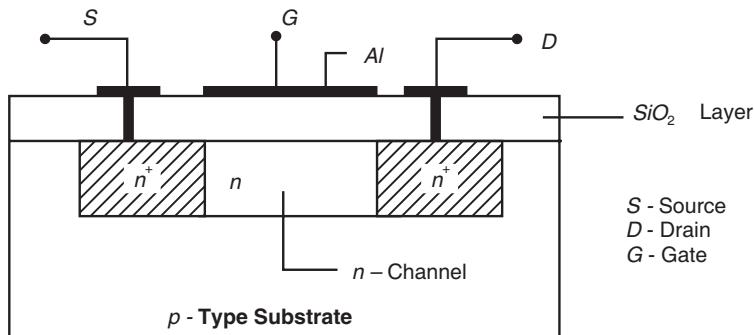


Fig. 38.16

They are separated by about 5  $\mu m$ . A lightly-doped *n*-channel is diffused between the source and the drain. A thin ( $\sim 0.1 \mu m$ ) layer of insulating silicon dioxide ( $SiO_2$ ) is grown over the surface of the structure. Holes are cut into the oxide layer to make metallic contacts with the source and the drain. Then a metallic layer (say, of aluminium) is overlaid on the oxide layer, covering the entire channel region. This aluminium layer acts as gate (*G*). Simultaneously, aluminium contacts are made to the source and the drain.

Fig. 38.17 shows the symbol of *n*-channel depletion-type MOSFET. Usually, the substrate terminal *SS* is internally connected to the source terminal *S*.

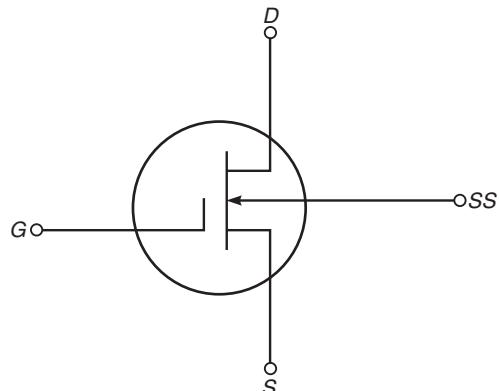


Fig. 38.17

### Operation

Suppose a potential is applied to the gate such that it is positive with respect to source. Then, the positive charge on the gate attracts additional free electrons into the channel from the source. This enhancement of mobile carriers decreases the channel resistance.

Suppose the applied voltage makes the gate negative with respect to source (Fig. 38.18). Then, the negative charge on the gate forces free electrons out of the channel.

A carrier-depletion region is formed on the surface of the silicon at the oxide-silicon interface. The channel is thus constricted by a negative gate voltage. The resistance of the channel increases. When the gate is sufficiently negative, the depletion region extends completely across the channel and joins with depletion region of the *p-n* junction on the other side of the channel. Under this condition, the channel cannot conduct current between drain and source. Thus drain and source become cut off. The negative gate voltage when channel becomes nonconducting is the pinch off voltage  $V_p$ .

A positive voltage  $V_{DS}$  is also applied across the drain *D* and the source *S* (Fig. 38.18). There is a voltage drop along the channel, with the drain end of the channel positive relative to the source end. This further increases the depletion region at the surface.

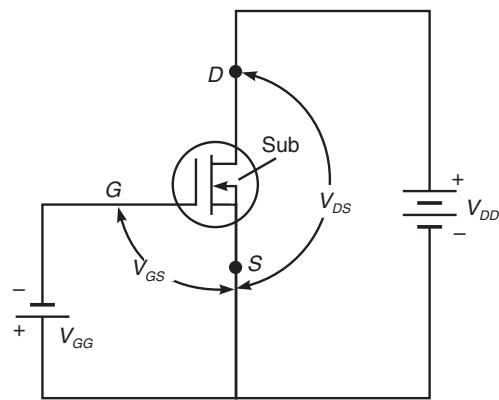


Fig. 38.18

### 38.7.1. Static Characteristics of Depletion MOSFET

**1. Output or Drain Characteristics.** Fig. 38.19 shows the drain characteristic curves ( $i_D$  versus  $V_{DS}$  at constant  $V_{GS}$ ) for an *n*-channel MOSEFT.

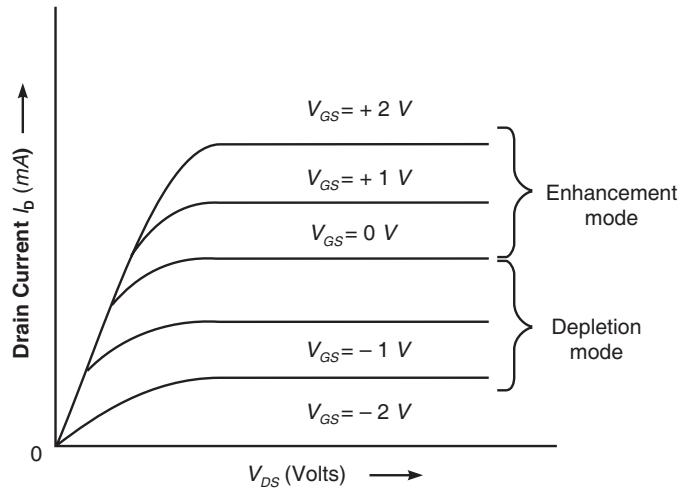


Fig. 38.19

An *n*-channel MOSFET may be operated in either the enhancement mode or the depletion mode. The enhancement mode occurs for positive values of  $V_{GS}$  while the depletion mode occurs for negative values of  $V_{GS}$ .

When no gate-to-source voltage is applied ( $V_{GS} = 0$ ), a significant drain current  $i_D$  flows due to the flow of majority carriers (electrons) in the *n*-channel from source to drain under  $V_{DS}$ . As  $V_{DS}$  increases,  $i_D$  increases to a saturation value.

When the gate is made negative, positive charges are induced in the channel through the dielectric  $\text{SiO}_2$ . This causes depletion of electrons (majority carriers) in the channel. Therefore, the channel becomes less conductive. Consequently, the drain current progressively decreases to almost zero as  $V_{GS}$  is made more and more negative, at all values of  $V_{DS}$ .

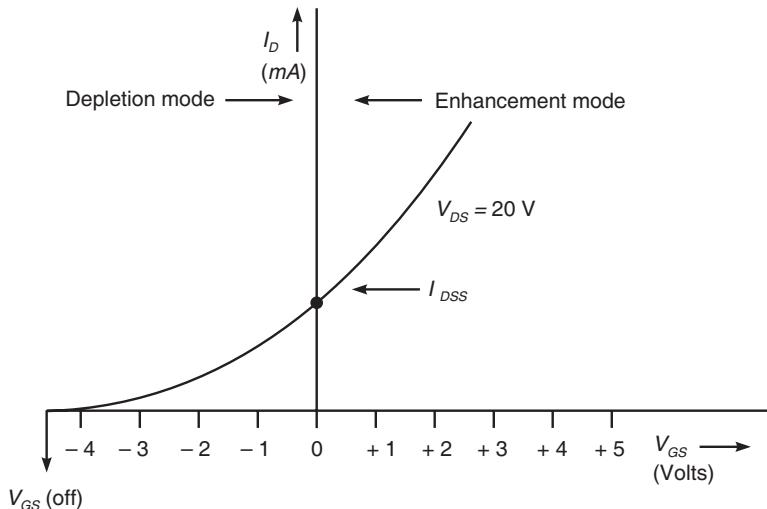


Fig. 38.20

When the gate is made positive, negative charges are induced in the channel, thus enhancing the majority-carriers. Hence the drain current increases above its value at  $V_{GS} = 0$ , for all  $V_{DS}$ .

Thus, this MOSFET can be operated in depletion as well as in enhancement mode although it is termed depletion type.

**2. Transfer Characteristics.** The variation of drain current  $I_D$  with gate voltage  $V_{GS}$  at a constant drain voltage  $V_{DS}$  is represented by the ‘transfer curve’ for the MOSFET (Fig. 38.20). It may be seen that drain current  $I_D$  flows even when gate-bias  $V_{GS} = 0$ . When  $V_{GS}$  is made more and more negative,  $I_D$  goes on decreasing.

## 38.8 THE ENHANCEMENT MOSFET

### Construction

Fig. 38.21 shows the construction of an  $n$ -channel enhancement type MOSFET. It consists of a lightly doped substrate of  $p$ -type silicon. Two highly doped  $n^+$  regions are diffused in the substrate. These two  $n^+$  regions act as source ( $S$ ) and drain ( $D$ ). A thin layer of silicon dioxide ( $\text{SiO}_2$ ) is deposited over the substrate. Then a thin film of metal aluminium is deposited over  $\text{SiO}_2$ . This aluminium layer acts as gate ( $G$ ).

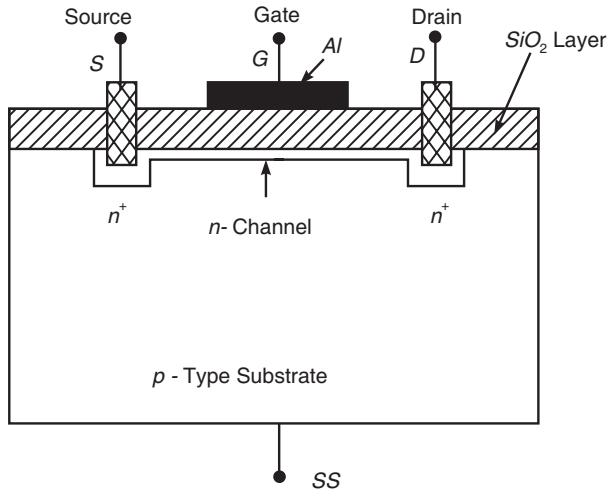


Fig. 38.21

Fig. 38.22 shows the symbol of  $n$ -channel E-MOSFET. The vertical line (representing channel) is broken to represent that there is no continuous channel in an enhancement MOSFET.

### Operation

The metallic layer of  $Al$  and the upper surface of the substrate act as the parallel plates of a capacitor. The insulating layer of  $\text{SiO}_2$  acts as the dielectric medium. When a *positive* voltage is applied to the gate  $G$  relative to source  $S$ , the capacitor begins to charge (Fig. 38.23). Consequently negative charges appear in the substrate between drain and source. Thus in effect an  $n$ -channel is created which allows the current to flow in source-to-drain circuit. Thus the MOSFET conducts only when gate is made positive relative to the source.

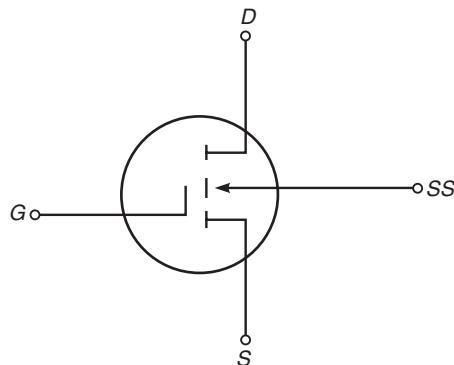


Fig. 38.22

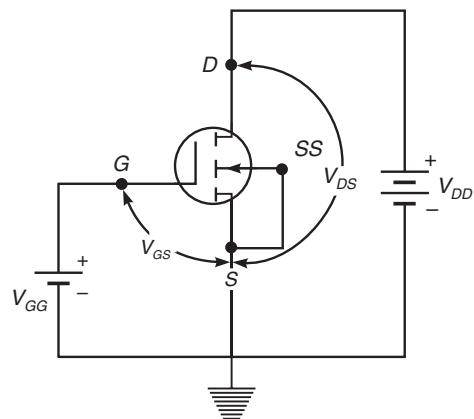


Fig. 38.23

The MOSFET can never operate with a negative gate voltage. The MOSFET is cut-off when  $V_{GS} = 0$ .

### 38.8.1. Characteristics of Enhancement MOSFET

#### 1. Output or Drain Characteristics.

Fig. 38.24 shows the drain characteristics of an *n*-channel enhancement MOSFET. Each curve shows the variation of drain current  $I_D$  with the drain-to-source voltage ( $V_{DS}$ ) for a fixed value of gate-to-source voltage ( $V_{GS}$ ).

The magnitude of the drain current  $I_D$  is a function of  $V_{DS}$  and  $V_{GS}$ . When  $V_{GS} = 0$ ,  $I_D = 0$  because two back to back *p-n* junctions between *S* and *D* (one between source *S* and the substrate and the other between the substrate and the drain *D*) are reverse biased regardless of the value of  $V_{DS}$ .

For a fixed positive  $V_{GS}$ , the drain current first increases rapidly with  $V_{DS}$  and then saturates. It increases for higher values of  $V_{GS}$ .

The minimum positive value of  $V_{GS}$  at which the drain current is established is called the 'gate-source threshold voltage'  $V_{GST}$  or  $V_T$ .

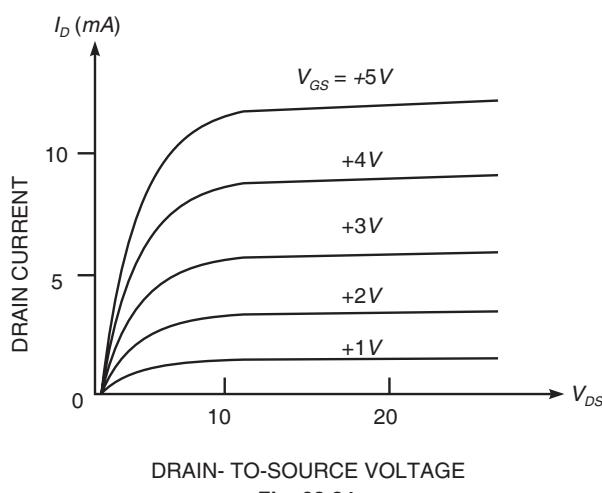
**2. Transfer Characteristics.** The curve showing the variation of drain current  $I_D$  with gate-to-source voltage  $V_{GS}$  for a fixed value of  $V_{DS}$  is called the transfer characteristic. The transfer curve for an *n*-channel enhancement MOSFET is shown in Fig. 38.25.  $I_D$  flows only when  $V_{GS}$  exceeds gate-to-source threshold voltage  $V_T$ . With increase in  $V_{GS}$ ,  $I_D$  increases slowly at first and then rapidly.

#### MOSFET Biasing

An enhancement MOSFET requires forward biasing of the gate-to-source junction. Fig. 38.26 shows a circuit arrangement for biasing an enhancement MOSFET. It provides *forward biasing of the gate-to-source junction* whose magnitude is given by

$$V_{GS} = \frac{R_1}{R_1 + R_f} V_{DS}. \text{ Here the}$$

feedback resistor  $R_f$  provides the operating point stability.



DRAIN- TO-SOURCE VOLTAGE

Fig. 38.24

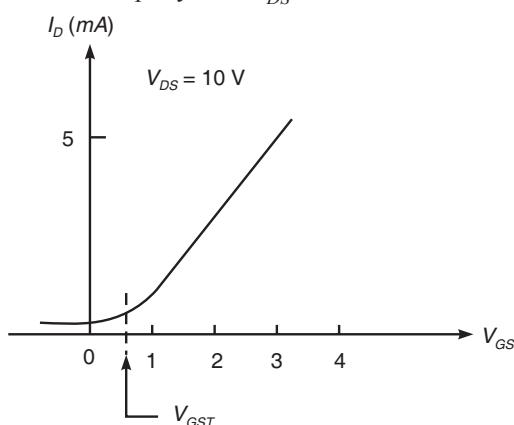


Fig. 38.25

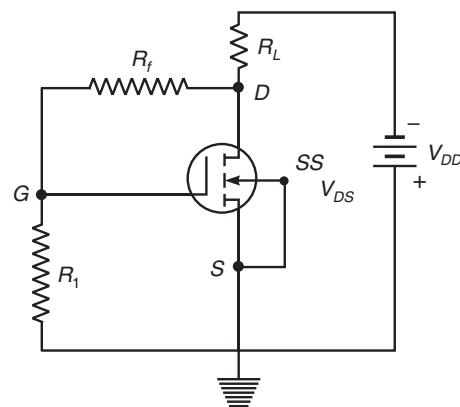
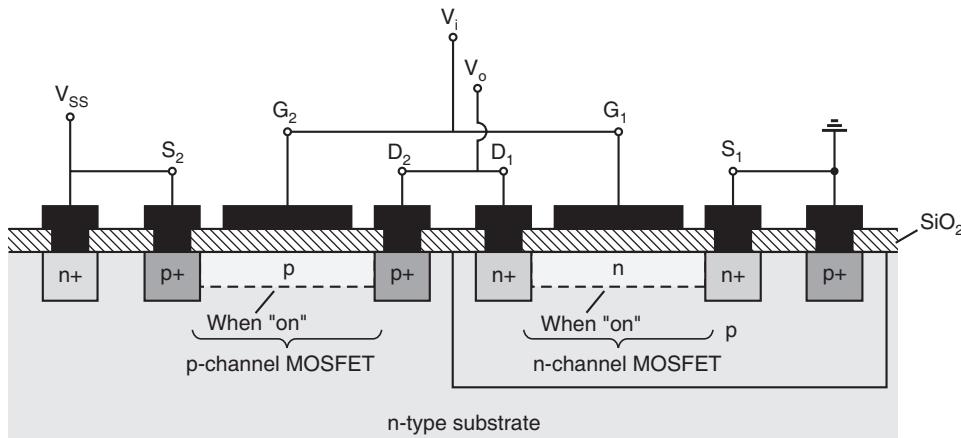


Fig. 38.26

38.9 CMOS

A very effective logic circuit can be established by constructing a *p*-channel and an *n*-channel MOSFET on the same substrate. (Fig. 38.27).



**Fig. 38.27**

- Note the induced *p*-channel on the left and the induced *n*-channel on the right for the *p*- and *n*-channel devices, respectively.
  - It has fast switching speeds, and lower operating power levels.

## Applications

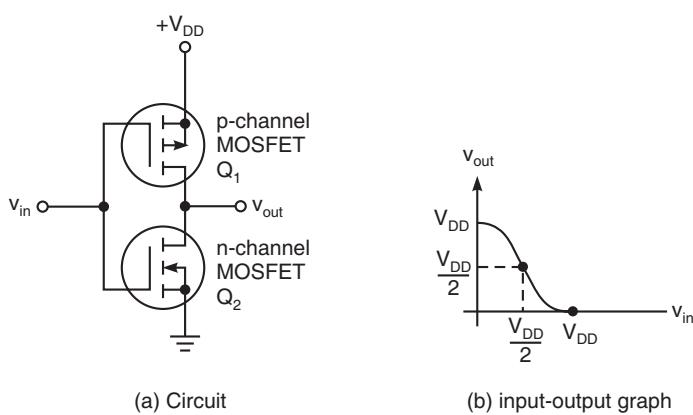
- Because of its ease of fabrication and lower power dissipation, CMOS is ideally suited for LSI (Large Scale Integration) and VLSI (Very Large Scale Integration) for dedicated applications such as large memories, calculator chips, large microprocessors, etc.

- Ideally suited for applications requiring battery power. One very effective use of the CMOS is as an inverter.

### CMOS inverter

- An inverter is a logic element that “inverts” the applied signal. That is, if the logic levels of operation are 0V(0-state) and 5V(1-state), an input level of 0V will result in an output level of 5V, and vice versa.

**Circuit.** Fig. 38-28 a shows the circuit



**Fig. 38.28**

- $Q_1$  is a *p*-channel MOSFET.  $Q_2$  is an *n*-channel MOSFET. These two devices are complementary. They have equal and opposite values of  $V_{GS(th)}$ ,  $V_{GS(on)}$ ,  $I_{D(on)}$ , and so on. One MOSFET conducts while the other is off.

The input voltage is either high ( $+V_{DD}$ ) or low (0V).

- When the input voltage is high,  $Q_1$  is off and  $Q_2$  is on. The shorted  $Q_2$  pulls the output voltage down to ground.

- When the input voltage is low,  $Q_1$  is on and  $Q_2$  is off. The shorted  $Q_1$  pulls the output voltage up to  $+V_{DD}$ .

- Since the output voltage is inverted, the circuit is called a *CMOS inverter*.

**Input-output graph.** Fig. 38.28 *b* shows how the output voltage varies with the input voltage. When the input voltage is zero, the output voltage is high. When the input voltage is high, the output voltage is low.

Between the two extremes, there is a crossover point where the input voltage equals  $V_{DD}/2$ . At this point, both MOSFETs have equal resistances and the output voltage equals  $V_{DD}/2$ .

- The main *advantage* of CMOS is its extremely low power consumption. The average power consumption is so small that CMOS circuits are often used for battery-powered applications such as calculators, digital watches, and hearing aids.

### SILICON-CONTROLLED RECTIFIER (SCR)

## 38.10 SILICON-CONTROLLED RECTIFIER

The SCR is a rectifier constructed of silicon material. Silicon is chosen because of its high temperature and power capabilities.

**Construction:** It is a four-layered device. It has three terminals. The graphic symbol for the SCR is shown in Fig. 38.29 with the corresponding connections to the four-layer semiconductor structure.

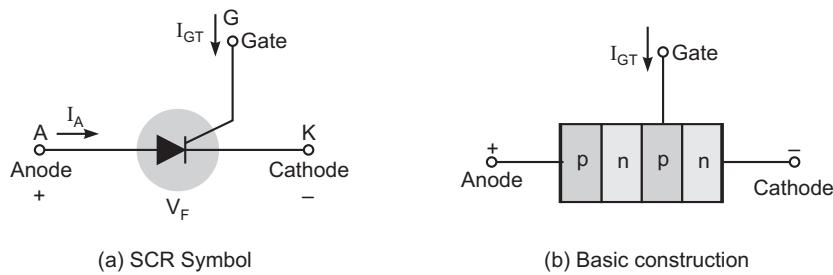


Fig. 38.29

As indicated in Fig. 38.29 *a*, if forward conduction is to be established, the anode must be positive with respect to the cathode. This is not, however, a sufficient criterion for turning the device on. A pulse of sufficient magnitude must also be applied to the gate to establish a turn-on gate current, represented symbolically by  $I_{GT}$ .

### Basic operation of an SCR

The four-layer *pnpn* structure of Fig. 38.29 *b* can be split into two three-layer transistor structures [Fig. 38.30 *(a)*].

Fig. 38.30 *(b)* shows the *SCR two-transistor equivalent circuit*.

$Q_1$  is a *pnp* transistor.

$Q_2$  is an *npn* transistor.

The signal shown in Fig. 38.31 *(a)* is applied to the gate of the circuit of Fig. 38.30 *(b)*.

## (i) "Off" state of the SCR. (When Gate is open)

During the interval  $0 \rightarrow t_1$ ,  $V_{\text{gate}} = 0$  V, the circuit of Fig. 38.30 (b) will appear as shown in Fig. 38.31 (b) ( $V_{\text{gate}} = 0$  V is equivalent to the gate terminal being grounded).

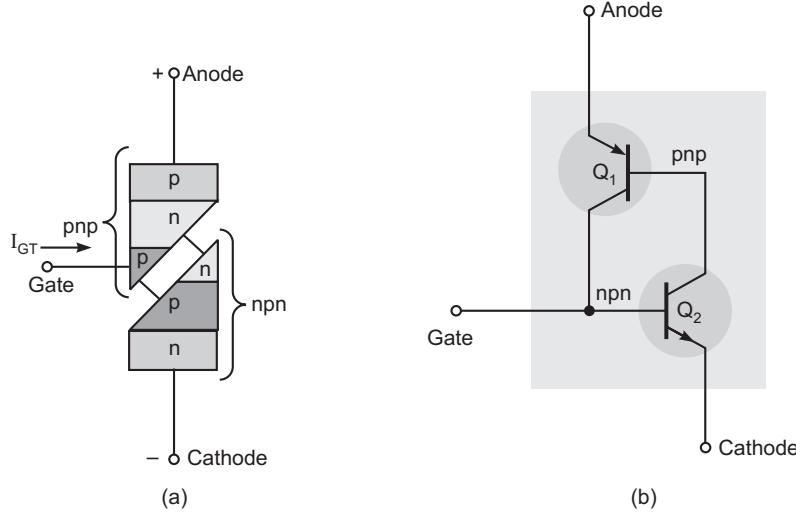


Fig. 38.30

- For  $V_{BE_2} = V_{\text{gate}} = 0$  V, the base current  $I_{B_2} = 0$ , and  $I_{C_2}$  will be approximately  $I_{CO}$ .
- The base current of  $Q_1$ ,  $I_{B_1} = I_{C_2} = I_{CO}$ , is too small to turn  $Q_1$  on. Both transistors are therefore in the "off" state, resulting in a high impedance between the collector and the emitter of each transistor.

Fig. 38.31 (c) shows open-circuit representation for the SCR.

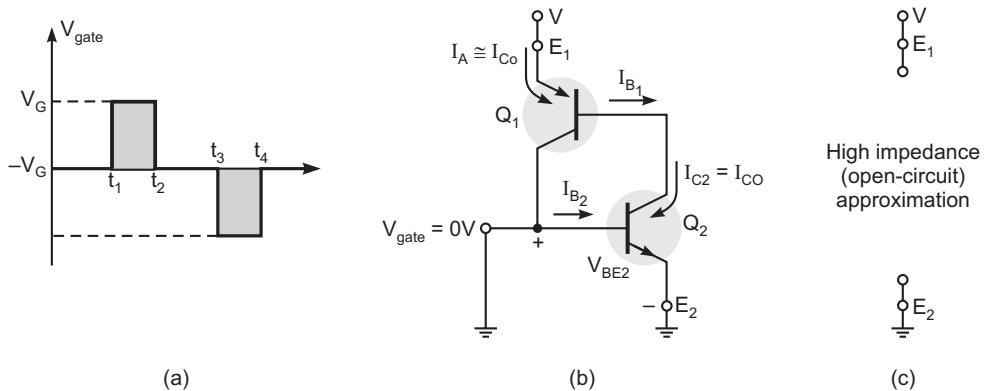
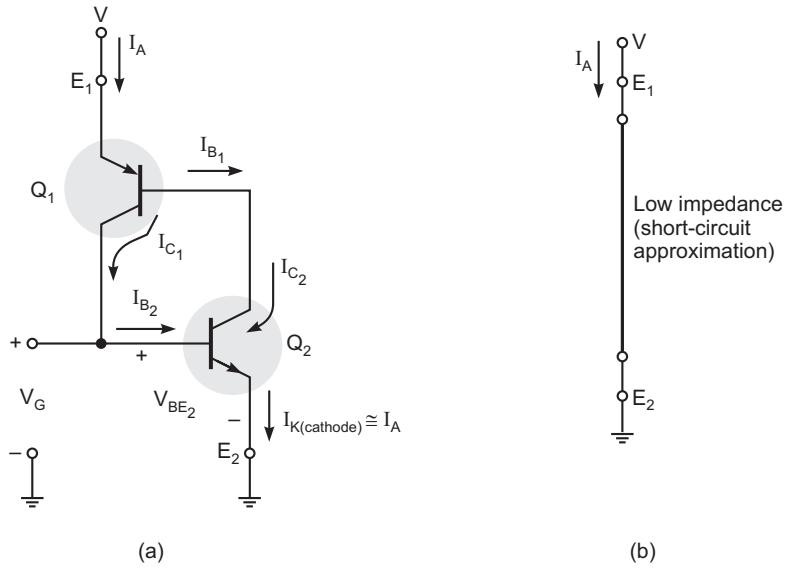


Fig. 38.31

## (ii) "On" state of the SCR (When gate is applied +ve voltage w.r.t Cathode)

- At  $t = t_1$ , a pulse of  $V_G$  volts will appear at the SCR gate [Fig. 38.32 (a)]. The potential  $V_G$  is chosen sufficiently large to turn  $Q_2$  on ( $V_{BE_2} = V_G$ ). The collector current of  $Q_2$  will then rise to a value sufficiently large to turn  $Q_1$  on ( $I_{B_1} = I_{C_2}$ ).
- As  $Q_1$  turns on,  $I_{C_1}$  will increase, resulting in a corresponding increase in  $I_{B_2}$ . The increase in base current for  $Q_2$  will result in a further increase in  $I_{C_2}$ .
- The net result is a regenerative increase in the collector current of each transistor. The resulting anode-to-cathode resistance ( $R_{SCR} = V/I_A$ ) is then small because  $I_A$  is large. Fig. 38.32 (b) shows the short-circuit representation for the SCR.



**Fig. 38.32**

- An SCR *cannot* be turned off by simply removing the gate signal. Only a special few can be turned off by applying a negative pulse to the gate terminal at  $t = t_3$  [Fig. 38.31 (a)].

### **Volt-ampere (V - I) characteristics of an SCR**

The characteristics of an SCR are provided in Fig. 38.33 for various values of gate current. It is the graph drawn between anode - cathode voltage (supply voltage)  $V_F$  and the anode current  $I_A$  for various values of gate current  $I_G$ .

**(i) Forward characteristics:**

- As the supply voltage increases, the anode current remains very small at first and SCR is in the “OFF” state.
  - As the supply voltage increases further, the anode current also increases. At a particular voltage ( $V_{(BR)F^*}$ ), called *Forward breakdown voltage*, SCR is turned “ON” (fired). SCR conducts heavy current.

1. *Forward breakdown voltage*  $V_{(BR)F^*}$  is the voltage above which the SCR enters the conduction. The asterisk (\*) denotes the letter to be added, which is dependent on the condition of the terminal as follows:

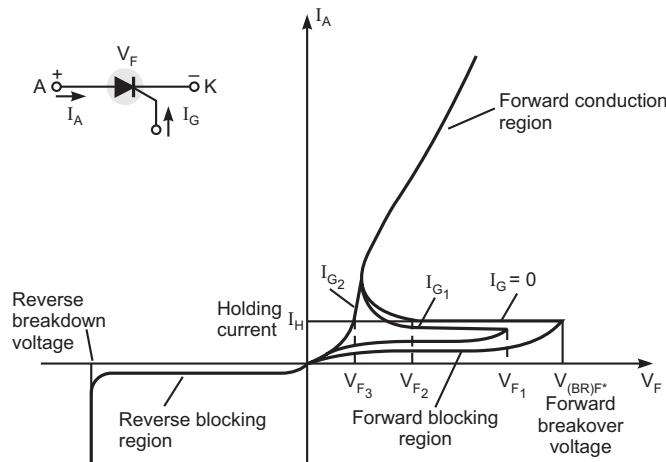


Fig. 38.33

$O$  = open circuit from  $G$  to  $K$

$S$  = short circuit from  $G$  to  $K$

$R$  = resistor from  $G$  to  $K$

$V$  = fixed bias (voltage) from  $G$  to  $K$

2. *Holding current*  $I_H$  is the value of current below which the *SCR* switches from the conduction state to the forward blocking region under stated conditions.

3. *Forward blocking region* is the region corresponding to the open-circuit condition for the controlled rectifier that *blocks* the flow of charge (current) from anode to cathode.

#### Curves for varying values of $I_G$

- For the characteristic having  $I_G = 0$ ,  $V_F$  must reach the largest required breakdown voltage ( $V_{(BR)F^*}$ ) before the “collapsing” effect results and the *SCR* can enter the conduction region corresponding to the *on* state.
- If the gate current is increased to  $I_{G1}$ , by applying a bias voltage to the gate terminal, the value of  $V_F$  required for the conduction ( $V_{F1}$ ) is considerably less. Note also that  $I_H$  drops with increase in  $I_G$ .
- If the gate current is increased to  $I_{G2}$ , the *SCR* will fire at very low values of voltage ( $V_{F3}$ ). The characteristics will begin to approach those of the basic  $p-n$  junction diode.

#### (ii) Reverse characteristics:

When anode is given negative voltage w.r.t cathode, the curve between voltage and current is known as reverse characteristic.

- As the reverse voltage increases, the increase in anode current is very small. At a particular reverse voltage, avalanche breakdown occurs and *SCR* conducts heavily in the reverse direction. This maximum reverse voltage at which *SCR* starts conducting heavily is known as “reverse breakdown voltage”.

*Reverse breakdown voltage* is equivalent to the Zener or avalanche region of the fundamental two-layer semiconductor diode.

- Reverse blocking region* is the region corresponding to the open-circuit condition for the controlled rectifier that *blocks* the flow of charge (current) from anode to cathode.

### 38.11 SCR APPLICATIONS

#### 1. Series Static Switch :

Fig. 38.34 shows a half wave *series static switch*.

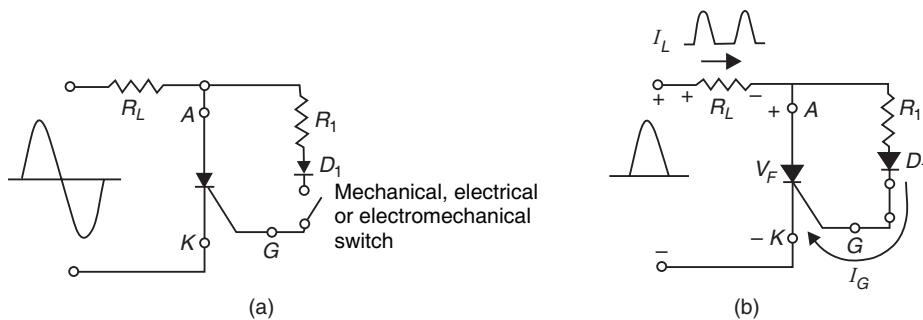


Fig. 38.34

If the switch is closed as shown in Fig. 38.34 (b), a gate current will flow during the positive portion of the input signal, turning the *SCR* on. Resistor  $R_1$  limits the magnitude of the gate current.

When the SCR turns on, the anode to cathode voltage ( $V_F$ ) will drop to the conduction value, resulting in a greatly reduced gate current and very little loss in the gate circuitry. For the negative region of the input signal, the SCR will turn off since the anode is negative with respect to the cathode. The diode  $D_1$  is included to prevent a reversal in gate current.

The waveforms for the resulting load current and voltage are shown in Fig. 38.34 (b). The result is a half-wave-rectified signal through the load. If less than  $180^\circ$  conduction is desired, the switch can be closed at any phase displacement during the positive portion of the input signal. The switch can be electronic, electromagnetic, or mechanical, depending on the application.

## 2. SCR half wave rectifier:

Fig. 38.35 shows the circuit diagram of SCR half wave rectifier and its waveform.

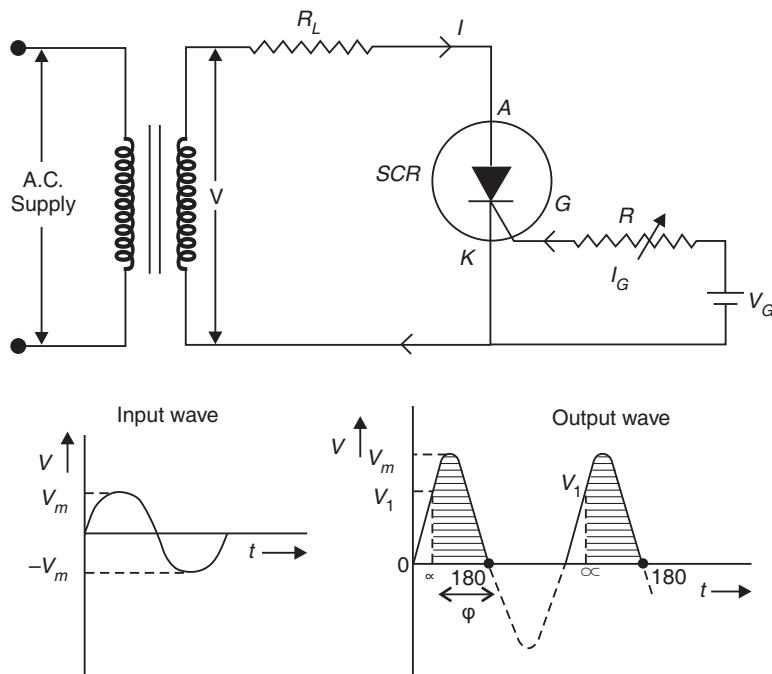


Fig. 38.35

The a.c. supply to be converted into d.c. supply is applied to the primary of the transformer. The load resistance  $R_L$  is connected in series with the anode. To control the gate current, a variable resistance  $R$  is inserted in the gate circuit. Output is taken across the load  $R_L$ .

### Operation

(i) Suppose the gate current is adjusted to such a value that SCR closes at a positive voltage  $V_1$ . So, when the secondary a.c. voltage becomes  $V_1$  in the positive half cycle, SCR conducts. Beyond this voltage SCR continues to conduct till the voltage becomes zero (at which it is turned OFF). SCR conducts current during the conduction angle  $\phi = 180 - \alpha$  where  $\alpha$  is the firing angle. So we get output voltage not during the entire positive half cycle but from  $V_1$  only.

(ii) During the negative half cycle of a.c. voltage appearing across secondary, the SCR will not conduct. Hence output voltage is zero.

Again at the start of the next positive half cycle, when the secondary a.c. voltage becomes  $V_1$ , the SCR conducts.

**Expression for average current :**

Let the alternating voltage appearing across the secondary of the transformer be  $V_m \sin \theta$ . Suppose the firing angle is  $\alpha$ . Then the rectifier will conduct from  $\alpha$  to  $180^\circ$  during the positive half cycle.

$\therefore$  Average output voltage,

$$V_{av} = \frac{1}{2\pi} \int_{\alpha}^{180^\circ} V_m \sin \theta \, d\theta = \frac{V_m}{2\pi} \int_{\alpha}^{180^\circ} \sin \theta \, d\theta = \frac{V_m}{2\pi} [\cos \alpha - \cos 180^\circ] = \frac{V_m}{2\pi} [1 + \cos \alpha]$$

$$\text{The average current } I_{av} = \frac{V_{av}}{R_L} = \frac{V_m}{2\pi R_L} (1 + \cos \alpha)$$

$$\text{When } \alpha = 0^\circ, \text{ then } I_{av} = \frac{V_m}{\pi R_L}.$$

$$\text{When } \alpha = 90^\circ, \text{ then } I_{av} = \frac{V_m}{2\pi R_L}$$

Thus as the firing angle  $\alpha$  increases,  $I_{av}$  decreases.

**3. SCR full wave rectifier**

Fig. 38.36 shows the circuit diagram of SCR full wave rectifier.

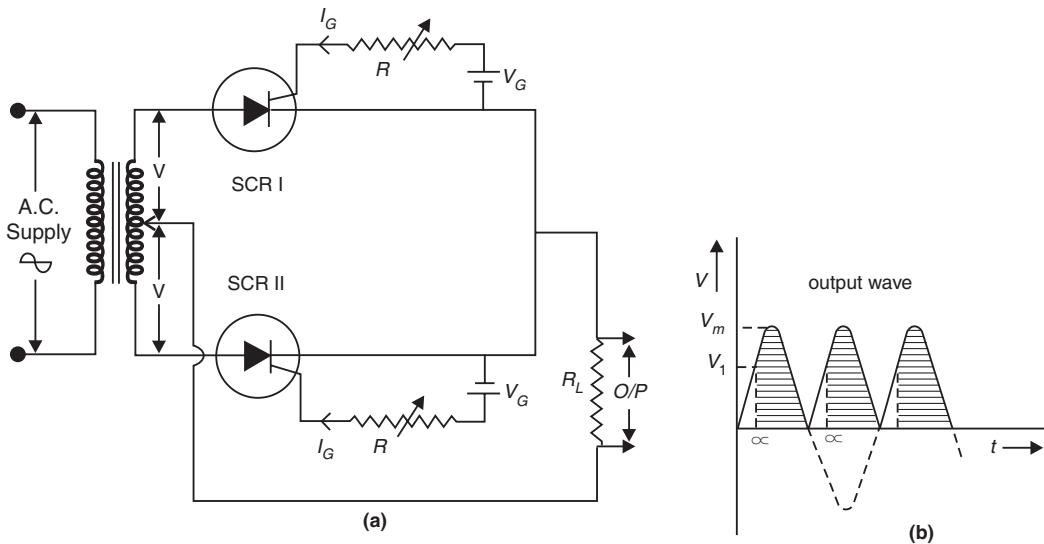


Fig. 38.36

The gates of both SCRs get their supply from two gate controls.

**Operation.** Suppose the gate currents are so adjusted that SCRs conduct as the secondary voltage (across half winding) becomes  $V_1$ . During the +ve half cycle, SCR I conducts. The conduction will start only when the voltage across the upper half of the secondary becomes  $V_1$ . During the -ve half cycle, SCR II conducts. Current through the load is in the same direction (d.c.) on both half-cycles of input a.c.

**Expression for average current in full wave rectifier**

Let  $V_m \sin \theta$  be the alternating voltage across one end of the secondary and centre tap. Let  $\alpha$  be the firing angle.

Average output voltage,

$$V_{av} = \frac{1}{\pi} \int_{\alpha}^{180^\circ} V_m \sin \theta d\theta = \frac{V_m}{\pi} [\cos \alpha - \cos 180^\circ] = \frac{V_m}{\pi} (1 + \cos \alpha)$$

$$\therefore \text{Average current, } I_{av} = \frac{V_{av}}{R_L} = \frac{V_m}{\pi R_L} (1 + \cos \alpha)$$

Thus  $I_{av}$  in full wave rectifier is two times that of half wave rectifier.

### UNIJUNCTION TRANSISTOR (UJT)

## 38.12 UNIJUNCTION TRANSISTOR (UJT)

### Construction

Fig. 38.37 (a) shows the basic construction of a unijunction transistor (UJT).

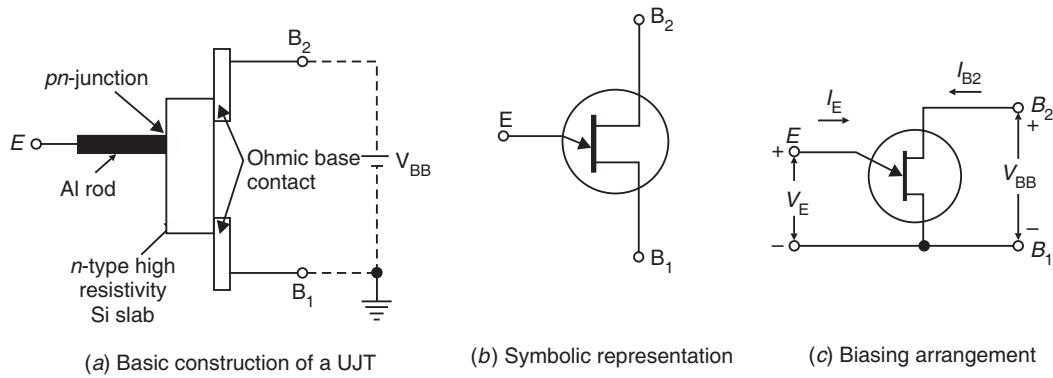
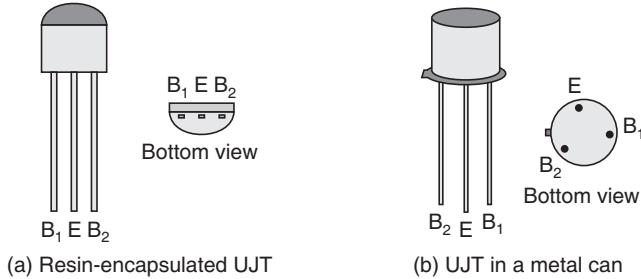


Fig. 38.37

It is a three-terminal device. A slab of lightly doped (increased resistance characteristic) *n*-type silicon material has two base contacts attached to both ends of one surface and an aluminium rod alloyed to the opposite surface. The *p-n* junction of the device is formed at the boundary of the aluminium rod and the *n*-type silicon slab. The single *p-n* junction accounts for the terminology *unijunction*. The aluminium rod is alloyed to the silicon slab at a point closer to the base 2 contact than the base 1 contact. Further, base 2 terminal is made positive with respect to the base 1 terminal by  $V_{BB}$  volts.

Fig. 38.37 (b) shows the circuit symbol of a UJT. Note that the emitter leg is drawn at an angle to the vertical line representing the slab of *n*-type material. The arrowhead is pointing in the direction of conventional current (hole) flow when the device is in the forward-biased, active, or conducting state.



The basic biasing arrangement for the unijunction transistor is provided in Fig. 38.37 (c).

### Equivalent Circuit of a UJT

The circuit equivalent of the UJT is shown in Fig. 38.38. Note the relative simplicity of this equivalent circuit: two resistors (one fixed, one variable) and a single diode.

The resistance of the silicon bar is represented by two series resistors:  $R_{B_2}$  is the resistance of the base-2 portion.

The resistance  $R_{B_1}$  is shown as a variable resistor since its magnitude will vary with the current  $I_E$ . For a representative UJT,  $R_{B_1}$  may vary from  $5 \text{ k}\Omega$  down to  $50 \Omega$  for a corresponding change of  $I_E$  from 0 to  $50 \mu\text{A}$ .

The interbase resistance  $R_{BB}$  is the resistance between terminals  $B_1$  and  $B_2$  when  $I_E = 0$ . In equation form,

$$R_{BB} = (R_{B_1} + R_{B_2})|_{I_E=0} \quad \dots(1)$$

Typically,  $R_{BB}$  ranges from 4 to  $10 \text{ k}\Omega$ .

The position of the aluminium rod will determine the relative values of  $R_{B_1}$  and  $R_{B_2}$  with  $I_E = 0$ .

A battery  $V_{BB}$  is connected across  $B_2 B_1$  of equivalent circuit of UJT. Part of  $V_{BB}$  is dropped over  $R_{B_2}$  and part on  $R_{B_1}$ .

The magnitude of  $V_{R_{B_1}}$  (with  $I_E = 0$  i.e., emitter open) is determined by the voltage-divider rule.

$$V_{R_{B_1}} = \frac{R_{B_1} V_{BB}}{R_{B_1} + R_{B_2}} = \eta V_{BB} \Bigg|_{I_E=0} \quad \dots(2)$$

Here,  $\eta$  is called the *intrinsic stand-off ratio* of the device. It is defined by

$$\eta = \frac{R_{B_1}}{R_{B_1} + R_{B_2}} \Bigg|_{I_E=0} = \frac{R_{B_1}}{R_{BB}} \quad \dots(3)$$

For applied emitter potentials ( $V_E$ ) greater than  $V_{R_{B_1}} (= \eta V_{BB})$  by the forward voltage drop of the diode  $V_D$  ( $0.35 \rightarrow 0.70 \text{ V}$ ), the diode will fire. Assume the short-circuit representation (on an ideal basis), and  $I_E$  will begin to flow through  $R_{B_1}$ . In equation form, the emitter firing potential is given by

$$V_P = \eta V_{BB} + V_D \quad \dots(4)$$

### OPERATION OF UJT

Fig. 38.39 shows the basic circuit operation of UJT.

Normally the base 1 ( $B_1$ ) is grounded. A positive voltage  $V_{BB}$  is applied at the base 2 ( $B_2$ ). When there is no emitter current, this voltage  $V_{BB}$  produces a uniform drop across the 5000 to 10,000 ohms internal resistance of silicon bar. The resultant current through the bar will produce a voltage drop  $\eta V_{BB}$  between points  $A$  and  $B_1$  of the bar where  $A$  denotes the point, where the  $P$  region is formed. So long as the voltage  $V_E$  applied to the emitter with respect to  $B_1$  is less than  $V_A$  or  $V_P (= \eta V_{BB})$ , the  $p-n$  junction remains reverse

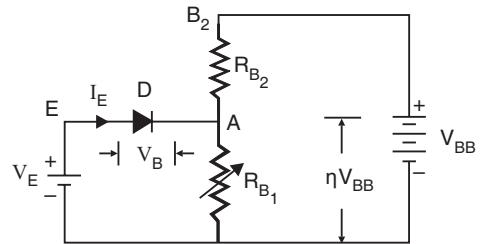


Fig. 38.38

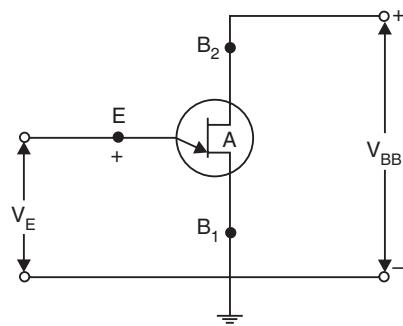


Fig. 38.39

biased and only a very small reverse current  $I_E$  flows in the emitter circuit. However, at the instant the emitter voltage  $V_E$  becomes greater than  $\eta V_{BB}$ , the junction becomes forward biased and the emitter current  $I_E$  increases. The increase is quite steep because this current is due to the holes injected into the silicon, which move from the emitter to base 1. The presence of these holes in the silicon bar attracts a considerable number of electrons to this region and, in consequence, causes the bar resistance between emitter and base 1 to drop sharply. The upshot is that an increase in emitter current causes a decrease in the resistance and in the emitter voltage, which is known as a negative resistance characteristic.

### STATIC Emitter – CHARACTERISTIC CURVE FOR A UJT

Fig. 38.40 shows the characteristics of a representative unijunction transistor for  $V_{BB} = 10\text{ V}$ .

If  $V_{BB}$  is made zero, and a voltage is applied to  $V_E$ , the resulting current  $I_E$  that flows gives the emitter to base-1 diode characteristic ( $I_{B2} = 0$ ).

Note that for emitter potentials to the left of the peak point, the magnitude of  $I_E$  is never greater than  $I_{E0}$  (measured in microamperes). The current  $I_{E0}$  corresponds very closely with the reverse leakage current  $I_{C0}$  of the conventional bipolar transistor. This region, is called the *cut off region*. Conduction is established at  $V_E = V_P$ . After  $V_P$ , an attempt to increase  $V_E$  is followed by a sudden increase in emitter current  $I_E$  with a corresponding decrease in  $V_E$ . This corresponds to a *negative resistance region*. After the negative resistance region, the valley point is reached. Further increase in  $I_E$  places the device in the *saturation region* (with characteristic approaching that of a semiconductor diode).

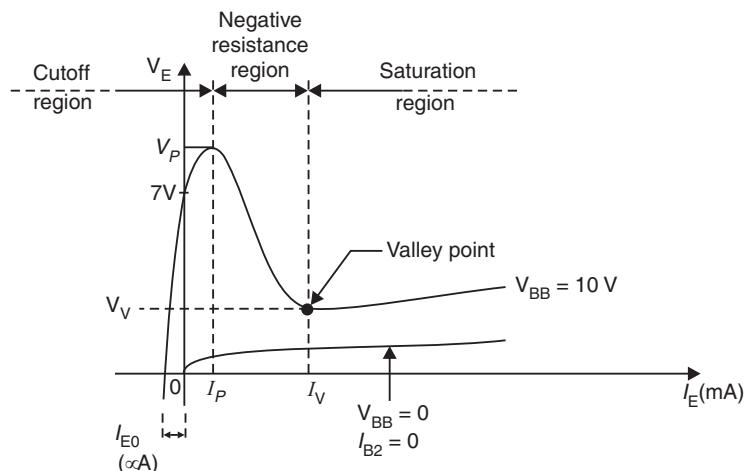


Fig. 38.40

The decrease in resistance in the active region is due to the holes injected into the *n*-type slab from the aluminium *p*-type rod when conduction is established. The increased hole content in the *n*-type material will result in an increase in the number of free electrons in the slab, producing an increase in conductivity ( $G \uparrow$ ) and a corresponding drop in resistance ( $R \downarrow = 1/G \uparrow$ ). Three other important parameters for the UJT are  $I_P$ ,  $V_V$  and  $I_V$ . Each is indicated on Fig. 38.40.

Fig. 38.41 shows the typical family of *static emitter characteristic curves* of a UJT for various  $V_{BB}$  (voltage between the bases) values.

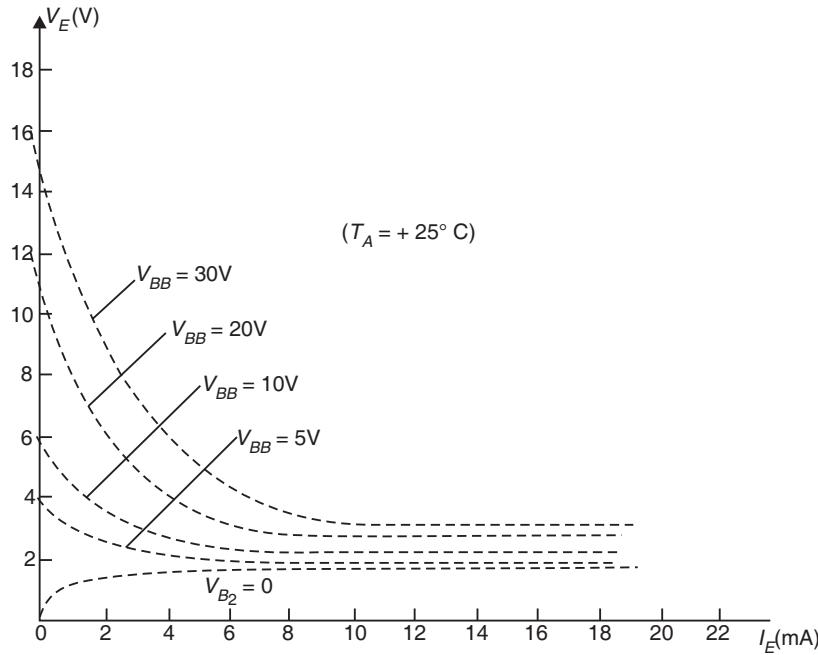


Fig. 38.41

Note that  $I_{E0}$  ( $\mu A$ ) is not in evidence since the horizontal scale is in milliamperes.

The intersection of each curve with the vertical axis is the corresponding value of  $V_P$ . For fixed values of  $\eta$  and  $V_D$ , the magnitude of  $V_P$  will vary as  $V_{BB}$ , that is,

$$V_P \uparrow = \underbrace{\eta V_{BB} \uparrow}_{\text{fixed}} + V_D$$

### 38.13 APPLICATIONS OF UJT

#### (1) UJT relaxation oscillator

The UJT can be used in a single stage oscillator circuit to provide a pulse signal suitable for digital-circuit applications.

#### Circuit Details

Fig. 38.42 shows the circuit of the UJT relaxation oscillator.

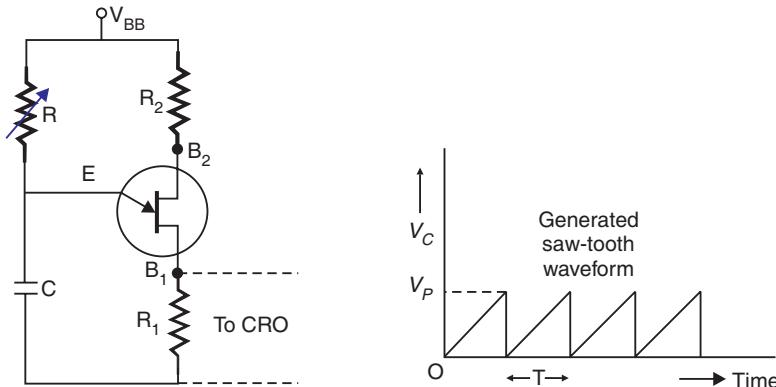


Fig. 38.42

It consists of a UJT and a capacitor  $C$  which is charged through a variable resistance  $R$  when  $V_{BB}$  is switched on.  $R_1$  and  $R_2$  are the external resistances such that  $R_1 \ll R_2$ .  $R_1$  is less than  $50\ \Omega$ . The inclusion of these resistances provide spike waveforms. Here the negative resistance region of the UJT is used. The discharge of the capacitor through UJT develops a saw-tooth output. Resistor  $R$  and capacitor  $C$  are the timing components that set the circuit oscillating rate.

### Circuit operation

When the circuit is switched on, the capacitor  $C$  starts charging through  $R$ . During charging period, the voltage across capacitor increases in an exponential relation till it reaches the peak point voltage  $V_p$ . At this instant of time, UJT switches to its low resistance conducting mode. So, the capacitor  $C$  discharges suddenly through UJT and  $R_1$ . Discharge time constant is very small due to small value of  $R_1$ . So  $C$  discharges abruptly. As the capacitor voltage moves back to zero, the emitter ceases to conduct and UJT is switched off. The capacitor starts charging again and the next cycle begins. The output saw-tooth waveform frequency can be varied by changing the value of  $R$ .

### Frequency of oscillation

The frequency of saw-tooth wave can be calculated as follows :

The voltage across capacitor during charging at any instant  $t$  is given by

$$v_c = V_{BB} (1 - e^{-t/\tau})$$

where  $\tau$  is charging time constant  $= R C$ .

This voltage is applied to the emitter. So

$$v_c = V_E = V_{BB} (1 - e^{-t/\tau})$$

When capacitor charges to peak voltage  $V_p$ , UJT triggers. The UJT triggers when  $V_p = \eta V_{BB}$ .

If the capacitor takes a time  $T$  to charge to firing potential  $V_p$ , then

$$V_p = V_{BB} (1 - e^{-T/CR})$$

or  $\frac{V_p}{V_{BB}} = \eta = (1 - e^{-T/CR})$

or  $e^{-T/CR} = 1 - \eta$

or  $T = CR \log_e \left( \frac{1}{1 - \eta} \right)$

Frequency of oscillation,  $f = \frac{1}{T} = \frac{1}{RC \log_e \left( \frac{1}{1 - \eta} \right)}$ .

The oscillating frequency may be calculated using the above equation, which includes the UJT intrinsic stand-off ratio  $\eta$  as a factor (in addition to  $R$  and  $C$ ) in the oscillator operating frequency.

### (2) UJT as over voltage detector

Fig. 38.43 shows a simple d.c. over voltage detector. A warning lamp  $L$  is connected between emitter  $E$  and  $B_1$  circuit.

#### Principle

As long as the input voltage is less than peak-point voltage  $V_p$  of UJT, the device remains switched off. When the input voltage exceeds  $V_p$ , the device is switched on.

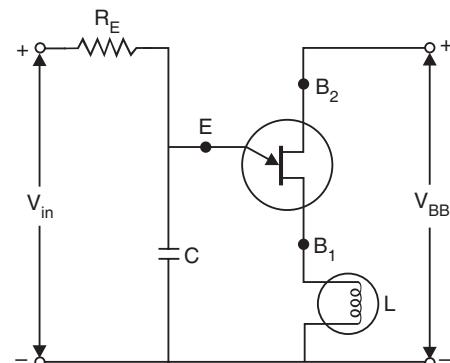


Fig. 38.43

### Working

Suppose we want to keep certain voltage constant say at value  $V_i$ . Now we choose  $V_{BB}$  in such a way that this is slightly less than  $V_i$ . In case  $V_i$  exceeds  $V_P$ , then UJT is switched on. As a result, the capacitor starts discharging through low resistance path between  $E$  and  $B_1$ . The current flowing in pilot lamp  $L$  lights it. This gives an indication of over voltage in the circuit.

### EXERCISE

1. Describe a FET. What are its advantages over a junction transistor? Sketch the characteristic curves of a FET. What is pinch off voltage? Define the three parameters of FET. *(Madras 1992)*
2. How can the static characteristics of a field effect transistor be studied? Give the necessary circuit diagram. *(Kolkata 82)*
3. Discuss FET as an amplifier having common source configuration. Obtain an expression for its voltage gain. *(Bangalore, 2000)*
4. Find an expression for the voltage gain and output resistance of an common drain FET amplifier. *(H.P.U. 2003)*
5. Sketch and explain the construction of an N-channel enhancement and depletion MOSFET. Draw the static drain characteristic and transfer characteristic curves for an N-channel MOSFET which may be operated in either the depletion or enhancement mode. *(Rohilkhand 88)*
6. Draw a biasing circuit for an enhancement MOSFET.
7. Describe the structure of SCR. Draw its volt-amp characteristics. Explain its switching action (a) when the gate is open (b) when the gate is +ve. *(Madras 1992)*
8. Explain the action of an SCR as a switch. What are the advantages of SCR switch over a mechanical or electro-mechanical switch?
9. Discuss some important applications of SCR.
10. Describe the construction of UJT. Draw its characteristic curve. Explain its -ve resistance region. Explain how it can be used as a relaxation oscillator. *(Madras 1994)*
11. The intrinsic stand-off ratio for a UJT is determined to be 0.6. If the inter-base resistance is  $10\text{ k}\Omega$ , what are the values of  $R_{B1}$  and  $R_{B2}$ ? *(B.U. 2011)*

[Ans.  $R_{B1} = 6\text{ k}\Omega$ ,  $R_{B2} = 4\text{ k}\Omega$ ]

## OPERATIONAL AMPLIFIERS

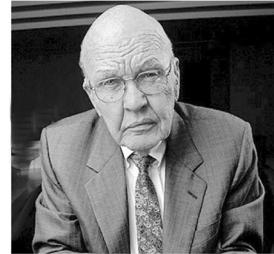
**After reading this chapter, you should be able to**

- ◆ Discuss manufacturing methods for the production of integrated circuits
- ◆ Discuss op-amp basics, parameters and op-amp characteristics
- ◆ Compare the voltage gain of inverting and noninverting amplifiers
- ◆ Analyse the operation of summing amplifier
- ◆ Analyse the operation of integrators and differentiators
- ◆ Analyse the noninverting voltage-follower
- ◆ Design op-amp series regulator circuits to fulfill a given specification.

### 39.1 INTEGRATED CIRCUIT (IC)

An integrated circuit is just a package of electronic circuit in which both the active and passive components are fabricated on a small semiconductor chip.

Thus in an IC, a number of circuits containing many diodes, transistors, resistors, capacitors etc., are formed and connected within an extremely tiny chip of semiconductor material. The typical size of an IC is  $0.2 \text{ mm} \times 0.2 \text{ mm} \times 0.001 \text{ mm}$ . The individual circuit components of an IC cannot be removed or replaced because each one of them is an integral part of the same semiconductor chip. The different components are isolated from each other by isolation diffusion within the crystal chip and are interconnected by an aluminium layer which serves as wires. Monolithic IC is the most commonly used one.



Op-Amps are the most versatile and widely used of all linear integrated circuits (ICs). In 1958, **Jack S. Kilby** made the first integrated circuit for which he was awarded a Nobel prize in 2000. This invention created the modern computer age.

#### Fabrication of a Monolithic IC

In monolithic IC, the entire circuit (all circuit components and their inter-connections) is formed on a single thin wafer, called the substrate.

Planar technology for IC fabrication consists of the following processes:

**1. Wafer (or substrate) preparation.** A single and pure *p*-type silicon crystal is grown and cut into wafers which are cleaned and polished to a mirror finish. The typical thickness of a wafer is  $\approx 200 \mu\text{m}$ . This wafer serves as the base for hundreds of ICs.

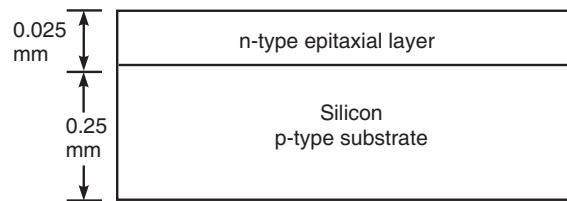
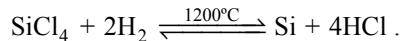


Fig. 39.1

**2. Epitaxial growth.** The epitaxial process is used to grow a thin layer ( $\approx 25 \mu\text{m}$ ) of low resistivity *n*-type silicon on the high resistivity *p*-type substrate (Fig. 39.1). Epitaxial growth is performed in a special furnace called a *reactor* into which the finished silicon wafers are inserted and heated to  $1200^\circ\text{C}$ . Hydrogen reduction of  $\text{SiCl}_4$  is used as the source of silicon to be grown. The chemical reaction is



For producing epitaxial film of particular impurity concentration, phosphine ( $\text{PH}_3$ ) is added for *n*-type doping and diborane ( $\text{B}_2\text{H}_6$ ) for *p*-type doping to the silicon tetrachloride-hydrogen gas stream.

The active and passive components are built within the thin *n*-type 'epitaxial layer' on top.

**3. Oxidation.** A thin ( $\approx 1 \mu\text{m}$ ) layer of  $\text{SiO}_2$  is formed over the entire wafer (Fig. 39.2) by heating the substrate to about  $1000^\circ\text{C}$  in an oxygen atmosphere.  $\text{SiO}_2$  has the property of preventing the diffusion of impurities through it.

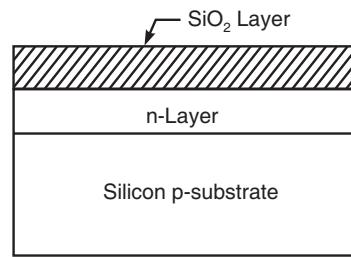


Fig. 39.2

**4. Photolithographic etching process.** Here the  $\text{SiO}_2$  layer is removed at different selected places on the surface of the wafer by a photolithographic etching process. It forms openings through which impurities may be diffused.

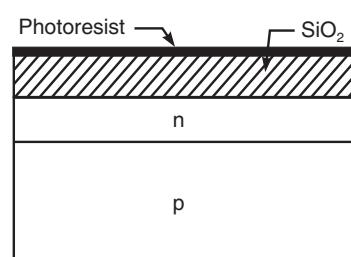


Fig. 39.3(i)

Fig. 39.3 illustrates the photoetching method.

(i) The wafer is coated with a uniform film of a photosensitive emulsion called ‘photoresist’ [Fig. 39.3 (i)].

(ii) A large black-and-white layout of the desired pattern of openings is made and reduced photographically. This negative is placed as a “mask” over the photoresist which is then exposed to ultraviolet light [Fig. 39.3 (ii)].

The photoresist under the transparent regions of the mask becomes polymerized. The mask is now removed, and the wafer is “developed” by a chemical (like trichloroethylene) which dissolves the unexposed (unpolymerized) portions of the photoresist coating and leaves the surface pattern as in Fig. 39.3 (b).

(iii) The  $\text{SiO}_2$  not covered by polymerized photoresist is then removed by immersing the chip in an etching solution of hydro-fluoric acid. It does not affect the portions of  $\text{SiO}_2$  which are protected by photoresist.

The exposed photoresist is then removed by a ‘stripping process’. It leaves the wafer ready for diffusion process, as shown in Fig. 39.3 (c).

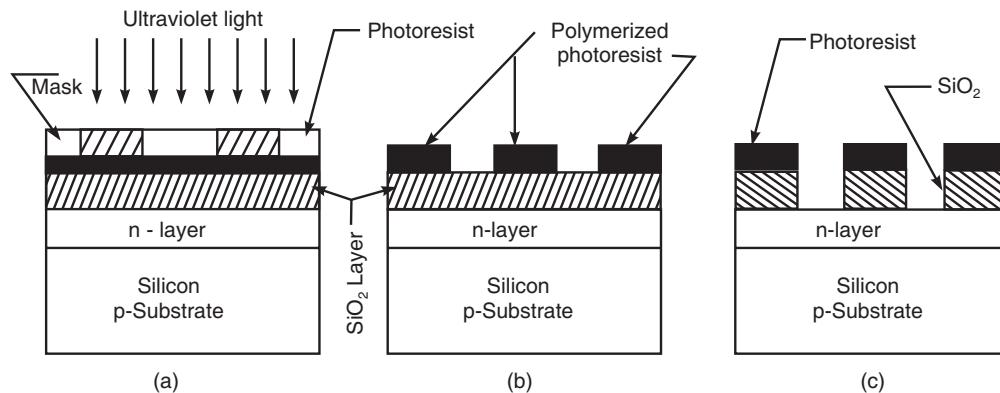


Fig. 39.3 (ii)

**5. Diffusion.** The introduction of controlled impurity concentrations is performed in a diffusion furnace at a temperature of about  $1000^\circ \text{C}$  over a 1-to 2-h period. Gaseous impurities used are generally the hydrides of boron and phosphorus. An inert gas (nitrogen) transports the impurity atoms to the surface of the wafers where they diffuse into the silicon.

In a bipolar transistor, two diffusions of impurities are often used. For an  $n-p-n$  device, the first is the diffusion of the  $p$ -type base into the  $n$ -type epitaxially grown collector; the second is the  $n$ -type emitter region into the  $p$ -type base. Fig. 39.4 shows the cross-section of an IC transistor. The collector junction is formed at a distance of  $2.7 \mu\text{m}$  from the surface. The emitter diffusion (phosphorus) starts from the surface and the emitter junction is formed at a depth of  $2 \mu\text{m}$ . This junction corresponds to the intersection of the base and emitter distributions of impurities. We now see that the base thickness for this monolithic transistor is  $0.7 \mu\text{m}$ .

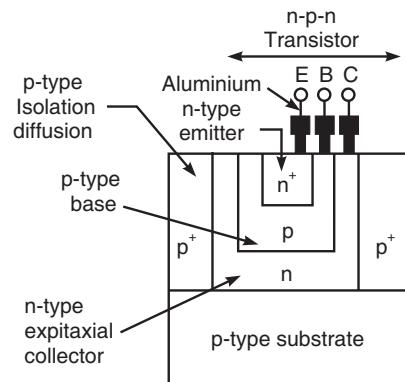


Fig. 39.4

**6. Metallization.** Interconnections between the various components of the integrated circuit are made by aluminium metallization. For this, a thin coating of aluminium is deposited over the

entire wafer by vacuum evaporation of aluminium, and the undesired aluminium areas are then etched away. This leaves the desired pattern of ohmic contacts and interconnections, as shown in Fig. 39.4 in which an *n-p-n* transistor has been fabricated on a Si chip.

## 39.2 INTRODUCTION TO OP AMPS

*An operational amplifier is a high gain direct coupled amplifier with high input impedance and low output impedance to which feedback is added to regulate overall response.*

### Block diagram of an op amp

Fig. 39.5 shows a block diagram of an op amp.

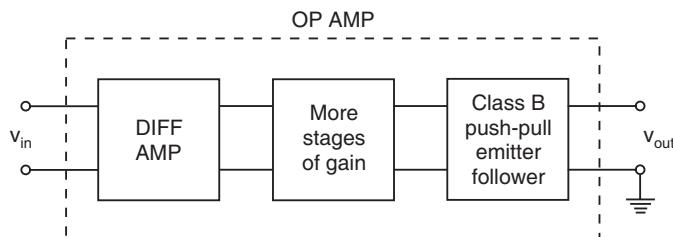


Fig. 39.5

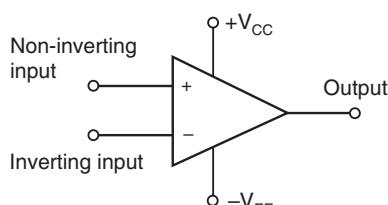
A typical op-amp is made up of three types of amplifier circuit: a *differential amplifier*, a *voltage amplifier*, and a *push-pull amplifier*.

The input stage is a diff amp, followed by more stages of gain, and a class B push-pull emitter follower.

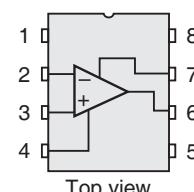
- (i) Because a diff amp is the first stage, it determines the input characteristics of the op amp. The differential amplifier can accept two input signals and amplifies the difference between these two input signals.
- (ii) The voltage amplifier is usually a class A amplifier that provides additional op-amp gain. Some op-amps may have more than one voltage amplifier stage.
- (iii) A push-pull class B amplifier is used for the output stage.

In most op amps the output is single-ended, as shown. With positive and negative supplies, the single ended output is designed to have a quiescent value of zero. This way, zero input voltage ideally results in zero output voltage.

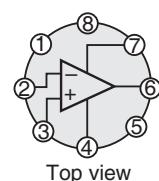
- Two important features of op amps are: the differential input and the single-ended output. Fig. 39.6 shows the op-amp circuit symbol and terminal connections. The + and - signs identify the non-inverting and inverting input terminals, respectively.



(a) Op-amp circuit symbol



(b) Terminal connections for DIP package



(c) Terminal connections for metal can package

Fig. 39.6

Two typical op-amp packages are illustrated in Fig. 39.6b and c. For both the dual-in-line plastic (DIP) package and the metal can package, terminals 2 and 3 are the inverting and non-inverting inputs, respectively, terminal 6 is the output, and terminals 4 and 7 are the + and - supply terminals.

### Basic differential Amplifier

Figure 39.7 shows an op amp connected as a differential amplifier.

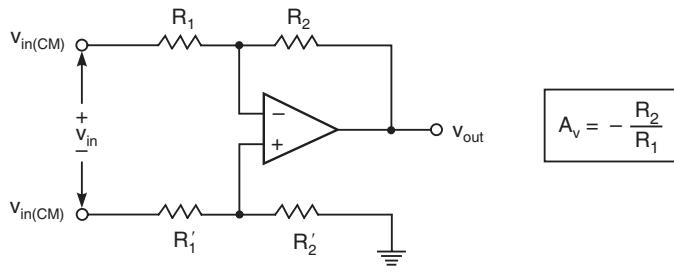


Fig. 39.7

- The resistor  $R'_1$  has the same nominal value as  $R_1$  but differs slightly in value because of tolerances.
- Resistors  $R_2$  and  $R'_2$  are nominally equal but may differ slightly because of tolerances.
- The desired input voltage  $v_{in}$  is called the **differential input voltage** to distinguish it from the common-mode input voltage  $v_{in(CM)}$ .
- The circuit provides operation as a constant-gain multiplier. The circuit amplifies the differential input voltage  $v_{in}$  to get an output voltage of  $v_{out}$ . Using the superposition theorem, it can be shown that:

$$v_{out} = A_v v_{in}$$

Voltage gain       $A_v = -\frac{R_2}{R_1}$  ... (1)

This voltage gain is called the **differential voltage gain** to distinguish it from the common-mode voltage gain  $A_{v(CM)}$ . By using precision resistors, we can build a differential amplifier with a precise voltage gain.

### Circuit Symbol of an OP AMP

Fig. 39.8 shows the circuit symbol of an OP AMP.

Terminals  $a$  and  $b$  are the input terminals. The terminal  $c$  is the output terminal. Terminal  $a$  (marked ' $-$ ') is called the *inverting input terminal*. The negative sign indicates that a signal applied at the terminal  $a$  will appear at the terminal  $c$  with a polarity opposite to that at the terminal  $a$ . Terminal  $b$  (marked ' $+$ ') is the *noninverting input terminal*. This means that the output signal at  $c$  is always of the same polarity as that of the signal applied at the terminal  $b$ .

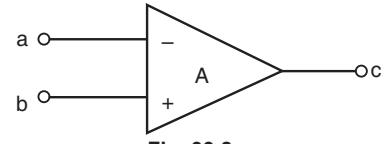


Fig. 39.8

The output signal voltage is proportional to the *difference* between the two signal voltages applied at the two input terminals.

$$V_0 = AV_1 - AV_2 = A(V_1 - V_2).$$

Here, voltages  $V_1$  and  $V_2$  are measured with respect to ground (Fig. 39.9).

$A$  is the differential voltage gain. The quantity  $A$  is a real constant approaching infinity in the ideal case for all frequencies.

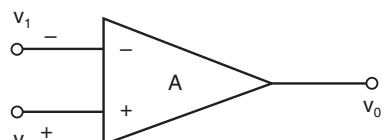


Fig. 39.9

**Common mode gain ( $A_C$ )**. The operation of the op amp depends not only on the difference signal ( $V_d$ ), but also on the average of the two input signals,

$$V_c = \frac{V_1 + V_2}{2}$$

$V_c$  is called the *common mode input signal*.

If we supply two input voltages which are equal in all the respects, i.e.,  $V_1 = V_2$ , then ideally the output voltage  $V_0 = (V_1 - V_2) A_d$  must be zero. But the output voltage not only depends on the difference voltage but also depends on the average common level of the two inputs. Such an average level of the two input signals is called **common mode signal** denoted by  $V_c$ .

Practically, the op-amp produces the output voltage proportional to such common mode signal also.

- The gain with which an op amp amplifies the common mode signal is called **common mode gain** denoted by  $A_c$ .

Thus,

$$V_0 = A_c V_c$$

Hence, there exists some finite output for  $V_1 = V_2$  due to such common mode gain  $A_c$  in case of a practical op-amp.

∴

$$V_0 = A_d V_d + A_c V_c$$

- For an ideal op amp,  $A_d$  must be infinite while  $A_c$  must be zero. This ensures zero output for  $V_1 = V_2$ . But due to mismatch in the internal circuitry, there is some output available for  $V_1 = V_2$  and gain  $A_c$  is not practically zero. The value of  $A_c$  is very small while the value of  $A_d$  is very large.

For ideal performance we like to have  $A_d$  large and  $A_c$  very small (ideally zero).

### Characteristics of an OP AMP

An ideal operational amplifier has the following characteristics.

1. Infinite input impedance *i.e.*,
2. Zero output impedance
3. Infinite voltage gain

$$Z_i = \infty$$

$$Z_o = 0$$

$$A = -\infty$$

Even a feeble differential input signal gets amplified to a large extent.

4. Infinite bandwidth

$$B_w = \infty$$

The voltage gain of the amplifier is constant at all frequencies of the input signals.

5. Perfect balance

$$V_0 = 0 \text{ when } V_1 = V_2.$$

Any signal, common to both the inputs, is rejected at the output.

6. Zero drift, *i.e.*, characteristics do not change with temperature.

7. Common mode rejection ratio should tend to  $\infty$ .

8. Slew rate should tend to  $\infty$ .

### The Common Mode Rejection Ratio (CMRR)

It is defined as the ratio between the differential gain,  $A_d$  to the common mode gain,  $A_c$ , *i.e.*,

$$CMRR = \rho = \frac{A_d}{A_c} = \frac{\text{Differential-mode gain}}{\text{Common-mode gain}}.$$

Alternately, *CMR* may be expressed in decibels as

$$CMR = 20 \log CMRR = 20 \log A_d - 20 \log A_c.$$

*CMRR* is infinity for a differential amplifier.

### The Slew Rate

*It is defined as the maximum rate of change in output voltage per unit of time*

$$SR = \left. \frac{dV_0}{dt} \right|_{\max}.$$

It is expressed in volts per microsecond.

Slew rate indicates how fast the output of an op-amp can change in response to changes in the input frequency.

**EXAMPLE.** The op-amp 741 has a slew rate of  $0.5 \text{ V}/\mu\text{s}$ . Its implication is explained by considering Fig. 39.10.

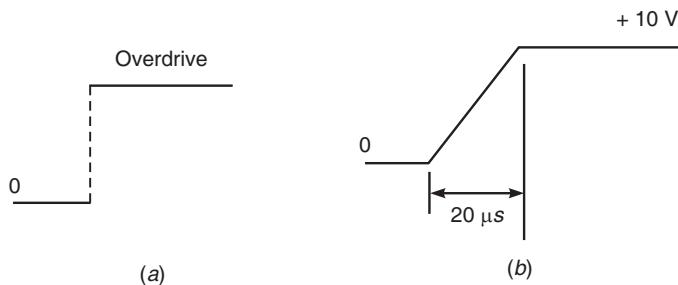


Fig. 39.10

If we overdrive a 741 with a large step input (Fig. 39.10 a), the output slews as shown in Fig. 39.10 b. It takes 20 microseconds for the output voltage to change from 0 to 10 V. It is impossible for the 741 to change faster than this.

### 39.3 INVERTING AMPLIFIER

If only one input is applied to the inverting input terminal, then it is called inverting amplifier. Fig 39.11 shows the circuit of an inverting amplifier.

The differential input resistance  $R_1$  is infinite for an ideal OP-AMP. But for a practical amplifier it must be much larger than output resistance  $R_0$ . So we can assume that  $I_0 = I_1$ . It means that an extremely negligible amount of current flows in the OP-AMP. We may write

$$I_1 = \frac{V_1 - V_s}{R_1} \quad \dots(1)$$

and

$$I_0 = \frac{V_s - V_o}{R_0} \quad \dots(2)$$

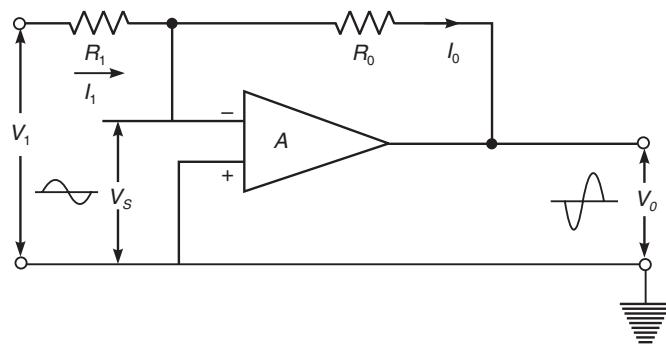


Fig. 39.11

Neglecting the current entering into the OP-AMP, we may write

$$\begin{aligned} I_1 &= I_0 \\ \frac{V_1 - V_s}{R_1} &= \frac{V_s - V_o}{R_0} \end{aligned}$$

On rearranging       $\frac{V_0}{R_0} = -\frac{V_1}{R_1} + \frac{V_s}{R_1} + \frac{V_s}{R_0}$  ... (3)

Again       $V_0 = -AV_s$  or  $V_s = -\frac{V_0}{A}$

$$\frac{V_0}{R_0} = -\frac{V_1}{R_1} - \frac{V_0}{A} \left( \frac{1}{R_0} + \frac{1}{R_1} \right) \quad \dots (4)$$

On the right hand side of Eq. (4), the second term may be neglected compared with the first term. Therefore, Eq. (4) can be written as

$$\frac{V_0}{R_0} = -\frac{V_1}{R_1}$$

or voltage gain,       $\frac{V_0}{V_1} = -\frac{R_0}{R_1}$  ... (5)

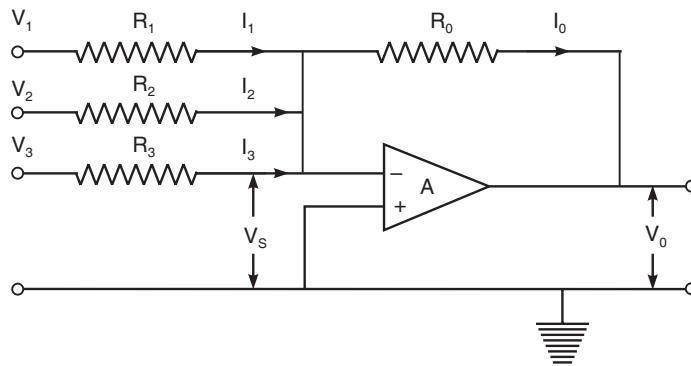
Thus the resulting amplification is entirely determined by feedback return *i.e.*, by resistors  $R_0$  and  $R_1$ . The negative sign in the voltage gain equation indicates a  $180^\circ$  phase shift.

## OP- AMP APPLICATIONS

### 39.4 ADDER OR SUMMING AMPLIFIER

An adder is a circuit whose output is proportional to the algebraic sum of the input voltages.

Consider an inverting amplifier with 3 inputs at the inverting terminal. (Fig. 39.12)



**Fig. 39.12**

The inverting terminal is virtually grounded by the feedback resistor  $R_0$ .

So the sum of the currents through  $R_1$ ,  $R_2$ ,  $R_3$  is equal to the current through  $R_0$ .

$$I_0 = I_1 + I_2 + I_3 \quad \dots (1)$$

$$\frac{V_s - V_0}{R_0} = \frac{V_1 - V_s}{R_1} + \frac{V_2 - V_s}{R_2} + \frac{V_3 - V_s}{R_3}$$

or       $-\frac{V_0}{R_0} = \frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3} - V_s \left( \frac{1}{R_0} + \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right)$  ... (2)

We have       $V_0 = -AV_s$ .

$$-\frac{V_0}{R_0} = \left( \frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3} \right) + \frac{V_0}{A} \left( \frac{1}{R_0} + \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right) \quad \dots (3)$$

Since the value of  $A$  is very large, the second term on the R.H.S. is neglected as compared with the first term. Eq. (3) becomes

$$-\frac{V_0}{R_0} = \frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3} \quad \dots (4)$$

If  $R_0 = R_1 = R_2 = R_3$ , we get

$$V_0 = -(V_1 + V_2 + V_3) \quad \dots (5)$$

output voltage =  $-(\text{sum of the input voltages})$ .

### 39.5 OP-AMP AS INTEGRATOR

The circuit shown in Fig. 39.13 produces an output voltage that is proportional to the time integral of the input voltage. Hence, this circuit is called an integrator.

The input signal source of voltage  $v(t)$  is connected to the inverting input terminal through resistance  $R$ . The negative feedback is given using a capacitor  $C$ . The non-inverting terminal is earth-connected [Fig. 39.13(a)]

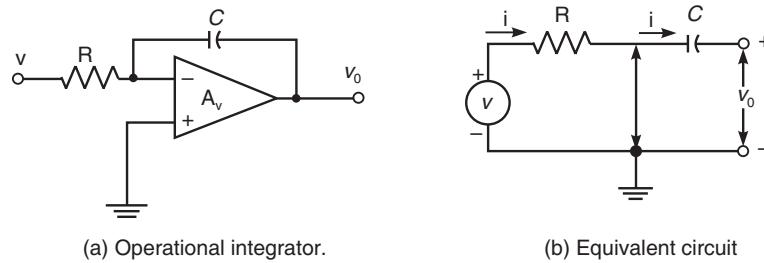


Fig. 39.13

In Fig. 39.13 b, the double-headed arrow represents a virtual ground. Hence  $i = v/R$ .

$$v_0 = -\frac{1}{C} \int i dt = -\frac{1}{RC} \int v dt \quad \dots (1)$$

The amplifier therefore provides an output voltage proportional to the integral of the input voltage.

**Case I.** If the input voltage is a constant,  $v = V$ , then the output will be a ramp,  $V_0 = -Vt/RC$ . Such an integrator makes an excellent sweep circuit for a cathode-ray-tube oscilloscope, and is called a *Miller integrator*, or *Miller sweep*.

**Case II.** If the input voltage is a square wave, then the output will be a triangle wave (Fig. 39.14).

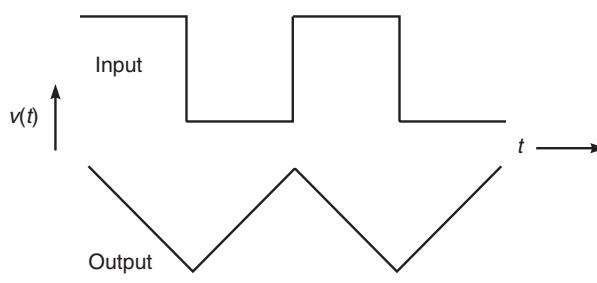


Fig. 39.14

The integrating circuits are used to solve differential equations in analog computers.

### 39.6 OP-AMP AS DIFFERENTIATOR

Fig. 39.15 (a) shows a circuit that performs the operation of differentiation.

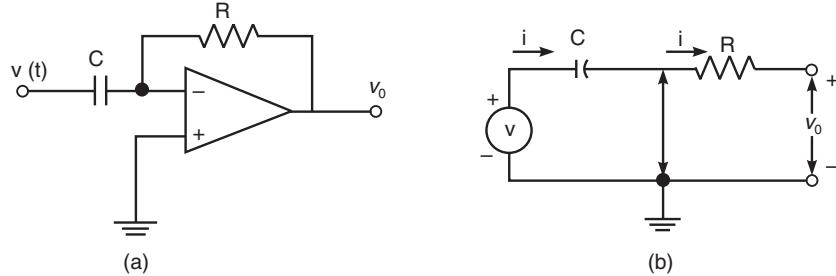


Fig. 39.15

The input signal source of voltage  $v(t)$  is connected to the inverting input terminal through a capacitor  $C$ . The non-inverting input terminal is earth-connected. Negative feedback is given through a resistance  $R$ .

Let  $v(t)$  be the signal voltage given as the input, which drives varying current through the capacitance  $C$ .

We see from the equivalent circuit of Fig. 39.15 (b) that

$$i = C \frac{dv(t)}{dt} \quad \dots(1)$$

$$v_0 = -Ri = -RC \frac{d}{dt} v(t) \quad \dots(2)$$

Hence, the output voltage is proportional to the differential of the input voltage.

### 39.7 DIFFERENCE AMPLIFIER OR SUBTRACTOR

The function of a subtracter is to provide an output proportional to or equal to the difference of two input signals.

We have to apply the inputs at the inverting and non-inverting terminals (Fig. 39.16).

Both inverting and non-inverting operations take place simultaneously.

The output can be derived using the superposition theorem.

Thus the resultant output,

$$v_0 = v_{01} + v_{02} \quad \dots(1)$$

Here,  $v_{01}$  and  $v_{02}$  are the outputs due to  $v_1$  and  $v_2$  inputs respectively.

**Output  $v_{01}$ :** The output  $v_{01}$  due to  $v_1$  alone with the other input earthed is the same as that of inverting circuit.

$$v_{01} = -(R_F/R_1)v_1 \quad \dots(2)$$

**Output  $v_{02}$ :** The output  $v_{02}$  due to  $v_2$  alone with the other input earthed is the same as that of non-inverting circuit.

$$v_{02} = (1 + R_F/R_1)v_A$$

$$v_A = R_3 v_2 / (R_2 + R_3)$$

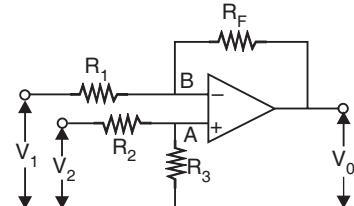


Fig. 39.16

$$v_{02} = \left(1 + \frac{R_F}{R_l}\right) \frac{R_3}{R_2 + R_3} v_2 \quad \dots(3)$$

The resultant  $v_0 = -\frac{R_F}{R_l} v_1 + \left(1 + \frac{R_F}{R_l}\right) \left(\frac{R_3}{R_2 + R_3}\right) v_2 \quad \dots(4)$

For  $R_2 = R_1$  and  $R_3 = R_F$ , we get

$$\begin{aligned} v_0 &= -\frac{R_F}{R_l} v_1 + \left(1 + \frac{R_F}{R_l}\right) \left(\frac{R_F}{R_l + R_F}\right) v_2 \\ &= (R_F/R_l) (v_2 - v_1) \end{aligned} \quad \dots(5)$$

Further, if  $R_F = R_l$ , then,

$$v_0 = (v_2 - v_1) \quad \dots(6)$$

Output voltage = difference of the two input voltages

### 39.8 LOGARITHMIC AMPLIFIER

**Principle.** A logarithmic amplifier has an output voltage which is proportional to the logarithm of the input voltage, i.e.,

$$v_0 \propto \log_e v_i$$

The linear OPAMP can be combined with a nonlinear element such as a diode or transistor to achieve this. The output gets greatly compressed. Therefore the response of a meter across the output will be like that of a decibel meter.

(i) Fig. 39.17 shows the circuit of a logarithmic amplifier using a diode.

The branch point  $S$  acts as the virtual ground. When the input voltage source is switched on, current flowing through the resistance  $R$  flows through the diode  $D$ .

Let

$I$  = forward current through the diode and

$V$  = forward voltage across the diode.

The relation between the current and voltage is

$$I = I_0(e^{qV/\eta kT} - 1) \quad \dots(1)$$

Here,  $q$  is the electronic charge,  $I_0$  is the reverse saturation current of the diode,  $k$  is the Boltzmann constant,  $T$  is the kelvin temperature of the diode.  $\eta = 1$  for germanium and 2 for silicon diode.

Let  $\frac{kT}{q} = x$ . Then,  $I = I_0 (e^{V/\eta x} - 1) = I_0 e^{V/\eta x} \left(\because \frac{V}{\eta x} \gg 1\right)$

By virtual ground property of the input terminal,  
current through  $R$  = current through the diode,

$$\frac{V_S}{R} = I$$

$$\therefore (V_S/R) = I_0 e^{V/\eta x}$$

$$\frac{V_S}{RI_0} = e^{V/\eta x}$$

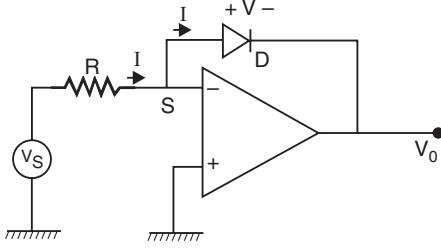


Fig. 39.17

$$\text{Taking logarithm, } \log_e \left( \frac{V_S}{RI_0} \right) = \frac{V}{\eta x}$$

$$\therefore V = \eta x \log_e(V_S/RI_0) \quad \dots(2)$$

The forward voltage across the diode  $V = 0 - V_0$

or  $V_0 = -V$

$$\text{The output voltage } V_0 = -\eta x \log_e \left( \frac{V_S}{RI_0} \right) \quad \dots(3)$$

All the quantities on the right hand side of the above equation (except  $V_S$ ) are constants at constant temperature. So, the output voltage  $V_0$  is proportional to the logarithm of the input voltage  $V_S$ , and the circuit works as logarithmic amplifier.

(ii) Though a diode can be used for realising a log-amplifier, the grounded base transistor is preferred because the exponential relation between current and voltage extends over a much wider voltage range. A typical circuit is shown in Fig. 39.18.

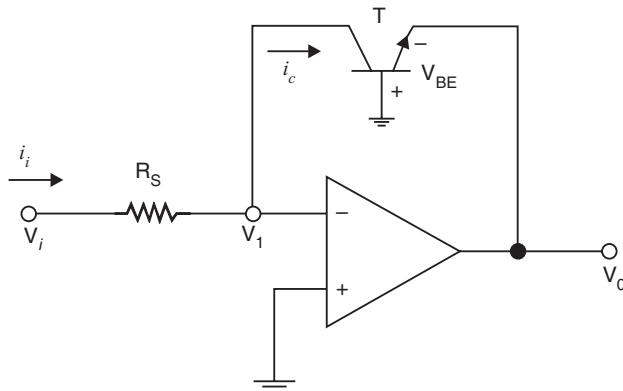


Fig. 39.18

Since  $v_1 = 0$  and the base is grounded,  $v_{CB} = 0$ .

The collector current  $i_C = \beta i_B$  ...(1)

Here, the base current is given by

$$i_B = I_S \exp(qv_{BE}/kT)$$

which is the same as the diode current of the forward-biased emitter-base  $p-n$  junction. Hence

$$i_C = \beta I_S \exp(qv_{BE}/kT) \quad \dots(2)$$

$$= K \exp(qv_{BE}/kT) \quad \dots(3)$$

$$\therefore v_{BE} = \frac{kT}{q} \ln \left( \frac{i_c}{K} \right) \quad \dots(4)$$

Since the emitter of the transistor is connected to the output of the OPAMP,

$$v_0 = v_{EB} = -v_{BE}$$

$$\text{Also, } i_C = i_i = \frac{v_i}{R_S} \quad \dots(5)$$

$$\text{Hence } v_0 = -\frac{kT}{q} \ln \left( \frac{v_i}{KR_S} \right) \quad \dots(6)$$

Thus  $v_0$  is proportional to  $\log v_i$ .

### 39.9 OP-AMP AS COMPARATOR

A comparator is a circuit with two inputs. It is used to compare a voltage in one input with a fixed reference voltage at the other input.

We can examine the operation of a comparator using a 741 op-amp, as shown in Fig. 39.19.

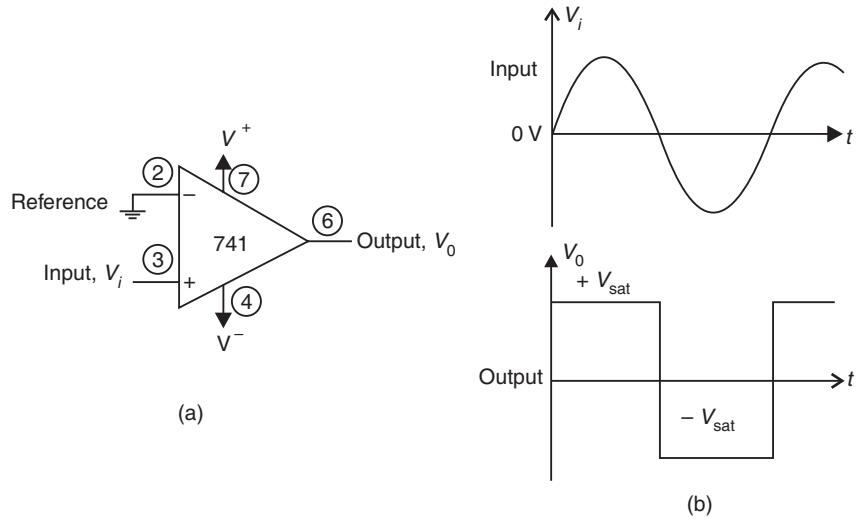


Fig. 39.19

With reference input (at pin 2) set to  $0V$ , a sinusoidal signal applied to the noninverting input (pin 3) will cause the output to switch between its two output states, as shown in Fig. 39.19 (b). The input  $V_i$  going even a fraction of a millivolt above the  $0-V$  reference level will be amplified by the very high voltage gain (typically over 100,000). So the output rises to its positive output saturation level and remains there while the input stays above  $V_{\text{ref}} = 0 V$ .

When the input drops just below the  $0-V$  reference level, the output is driven to its lower saturation level and stays there while the input remains below  $V_{\text{ref}} = 0 V$ .

The input signal is linear. The output signal is digital.

In general use, the reference level need not be  $0 V$  but can be any desired positive or negative voltage. Also, the reference voltage may be connected to either plus or minus input and the input signal then applied to the other input.

## 39.10 NON-INVERTING OPERATIONAL AMPLIFIER

Fig. 39.20 shows the circuit diagram of an ideal op-amp in the non-inverting mode. The input signal is applied to the non-inverting input terminal (+). The feed back is applied to the inverting input terminal (-) through  $R_f$ . The resistors  $R_f$  and  $R_i$  form the feed back voltage divider circuit.

### To find the voltage gain

The node  $B$  is at potential  $V_{in}$ . Hence the potential of point  $A$  is the same as that of  $B$  which is  $V_{in}$ .

Applying KCL,

$$\frac{0 - V_{in}}{R_i} = \frac{V_{in} - V_0}{R_f}$$

$$\therefore -\frac{V_{in}}{R_f} - \frac{V_{in}}{R_i} = -\frac{V_0}{R_f}$$

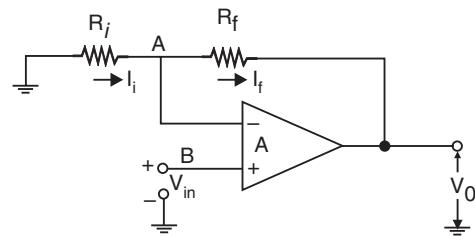


Fig. 39.20

$$\therefore V_{in} \left[ \frac{1}{R_i} + \frac{1}{R_f} \right] = \frac{V_0}{R_f}$$

$$V_0 = V_{in} \left[ \frac{R_f}{R_i} + 1 \right]$$

$$\therefore \frac{V_0}{V_{in}} = 1 + \frac{R_f}{R_i}$$

The output voltage is in phase with the input voltage (Fig. 39.21).

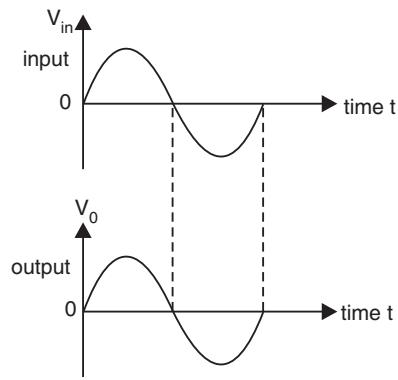


Fig. 39.21

### 39.11 VOLTAGE FOLLOWER (UNITY GAIN BUFFER)

A circuit in which the output voltage follows the input voltage is called voltage follower circuit.

Fig. 39.22 shows the voltage follower circuit using op-amp.

$$V_A = V_B = V_{in} \quad \dots(1)$$

The node  $B$  is at potential  $V_{in}$ . Now node  $A$  is also at the same potential as  $B$ , i.e.,  $V_{in}$ .

Node  $A$  is connected directly to the output.

$$\therefore V_0 = V_A \quad \dots(2)$$

From Eqs. (1) and (2),

$$V_0 = V_{in} \quad \dots(3)$$

i.e., output voltage = input voltage both in magnitude and phase.

The voltage gain of the circuit becomes unity. The circuit is then referred to as a *voltage follower* or a *unity-gain buffer*. As the circuit has a high input impedance and a low output impedance, it can be used as an impedance matching device between a high-impedance source and a low-impedance load.

A voltage buffer circuit provides a means of isolating an input signal from a load by using a stage having unity voltage gain, with no phase or polarity inversion, and acting as an ideal circuit with very high input impedance and low output impedance.

**EXAMPLE 1.** Show the connection of a 741 as a unity-gain circuit.

**SOL.** Fig. 39.23 shows the connection.

#### Multistage Op-Amp Circuit

- When a number of op amp stages are connected in series, the overall voltage gain is equal to the product of individual stage gains.

- Fig. 39.24 shows a three stage op-amp circuit.

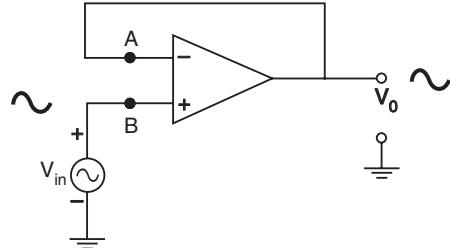


Fig. 39.22

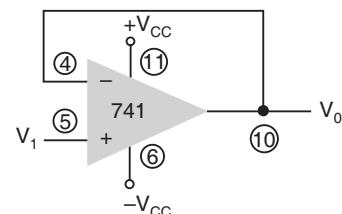


Fig. 39.23

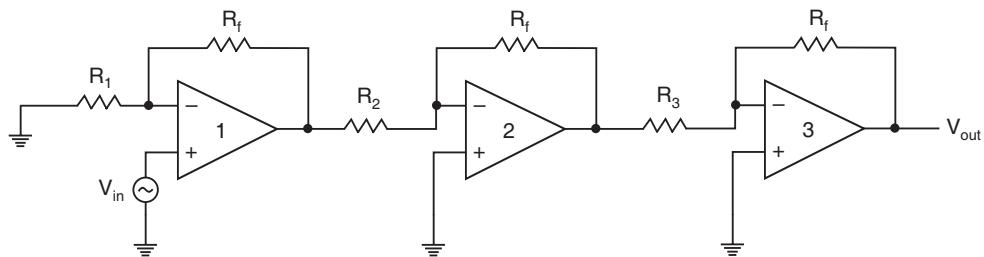


Fig. 39.24

The first stage is connected to provide noninverting gain. The next two stages provide inverting gains.

The overall voltage gain  $A$  of this circuit is given by:

$$A = A_1 A_2 A_3$$

Here,

$$A_1 = \text{Voltage gain of first stage} = 1 + (R_f/R_1)$$

$$A_2 = \text{Voltage gain of second stage} = -R_f/R_2$$

$$A_3 = \text{Voltage gain of third stage} = -R_f/R_3$$

The overall voltage gain is positive. So the circuit behaves as a noninverting amplifier.

#### IC VOLTAGE REGULATORS

### 39.12 INTRODUCTION

- A voltage regulator is a circuit that gives a constant output voltage irrespective of the variations in input voltage and load current.

**A voltage regulator provides a constant dc output voltage that is essentially independent of the input voltage, output load current, and temperature.**

*Voltage regulation is defined as the ability of a power supply source to maintain a constant output voltage in spite of ac input voltage fluctuations and change in load resistance.*

The two basic types of voltage regulation are : (i) Line regulation (ii) Load regulation.

$$\text{Line regulation} = \left( \frac{\Delta V_{OUT}}{\Delta V_{IN}} \right) 100\%$$

When the dc input (line) voltage changes, the voltage regulator must maintain a nearly constant output voltage.

$$\text{Load regulation} = \left( \frac{V_{NL} - V_{FL}}{V_{FL}} \right) 100\%$$

$V_{NL}$  = no-load output voltage,

$V_{FL}$  = full-load output voltage.

When the amount of current through a load changes due to a varying load resistance, the voltage regulator must maintain a nearly constant output voltage across the load.

#### Functional block diagram of three-terminal IC regulator

- IC regulators have three pins; one for the unregulated input voltage, one for the regulated output voltage and one for ground. Fig. 39.25 shows the functional block diagram for the 78 XX series.

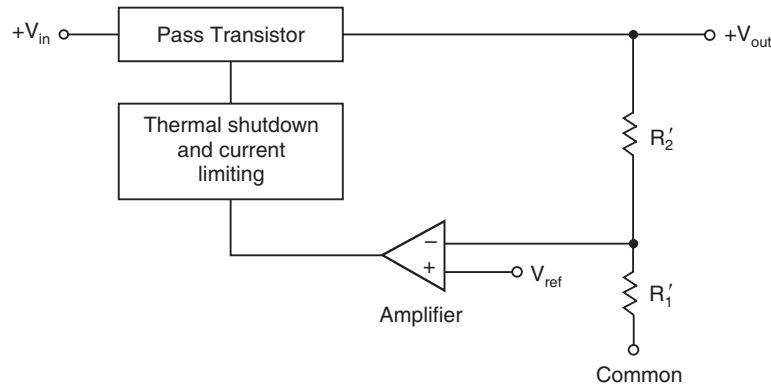


Fig. 39.25

(i) **Reference voltage.** A built in reference voltage  $V_{\text{ref}}$  drives the noninverting input of an amplifier.

(ii) A voltage divider consisting of  $R_1'$  and  $R_2'$  samples output voltage and returns a feedback voltage to the inverting input of a high-gain amplifier. The output voltage is given by:

$$V_{\text{out}} = \frac{R_1' + R_2'}{R_1'} V_{\text{ref}}.$$

The primes attached to  $R_1'$  and  $R_2'$  indicate that these resistors are inside the IC itself, rather than being external resistors. These resistors are factory-trimmed to get the different output voltages (5 to 24 V) in the 78XX series. The tolerance of the output voltage is  $\pm 4$  percent.

(iii) The LM78XX includes a pass transistor that can handle 1 A of load current. A heat sink must be used when high power dissipation is involved.

(iv) The IC includes current limiting and a *thermal shutdown* circuit that protects against excessive internal power dissipation. **Thermal shutdown** means that the chip will shut itself off when the internal temperature becomes too high, around 175°C. Because of thermal shutdown and current limiting, devices in the 78XX series are almost indestructible.

Some important types of linear IC voltage regulators are:

- (1) Fixed positive/negative output voltage regulators.
- (2) Adjustable output voltage regulators.

### 39.13 FIXED VOLTAGE REGULATORS

(i) The LM78XX series (where XX = 05, 06, 08, 10, 12, 15, 18 or 24) are three terminal, positive fixed voltage regulators. In 78XX, the last two numbers (XX) indicate the output voltage.

The 7805 produces an output of + 5 V, the 7806 produces + 6 V, the 7808 produces + 8 V, and so on, up to the 7824, which produces an output of + 24 V.

Fig. 39.26 (a) shows an LM7805 connected as a fixed voltage regulator. Pin 1 is the input, pin 2 is the output, and pin 3 is ground. The LM7805 has an output voltage of + 5 V and a maximum load current over 1A.

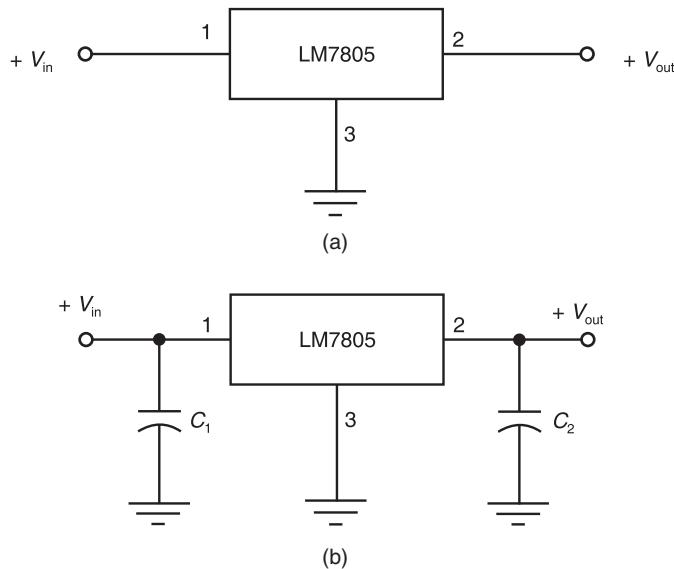


Fig. 39.26

The inductance of the connecting wire may produce oscillations inside the IC. The input capacitor  $C_1$  on pin 1 (Fig. 39.26 b) prevents oscillations. To improve the transient response of the regulated output voltage, a bypass capacitor  $C_2$  is used on pin 2.

(ii) The LM79XX series is a group of negative voltage regulators with preset voltages of  $-5$ ,  $-6$ ,  $-8$ ,  $-10$ ,  $-12$ ,  $-15$ ,  $-18$ , or  $-24$  V. For instance, an LM7905 produces a regulated output voltage of  $-5$  V. At the other extreme, an LM7924 produces an output of  $-24$  V.

### Regulated Dual Supplies

By combining an LM78XX and an LM79XX, as shown in Fig. 39.27, we can regulate the output of a dual supply. The LM78XX regulates the positive output, and the LM79XX handles the negative output. The input capacitors prevent oscillations, and the output capacitors improve transient response. The two diodes ensure that both regulators can turn on under all operating conditions.

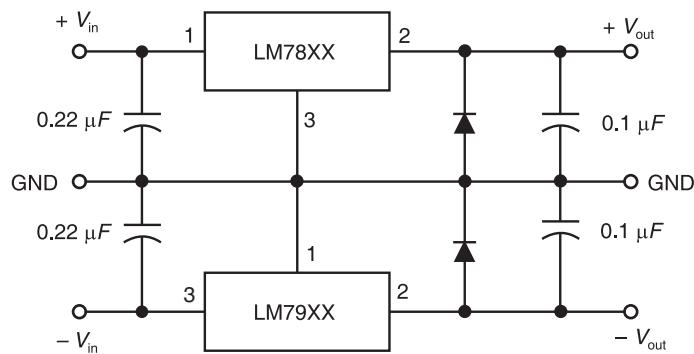


Fig. 39.27

## 39.14 ADJUSTABLE VOLTAGE REGULATOR

LM 723 C is the general purpose adjustable voltage regulator. The output voltage is adjustable from 2 to 37 V.

We can understand the regulating function of the chip LM 723C by considering its internal circuit (Fig. 39.28).

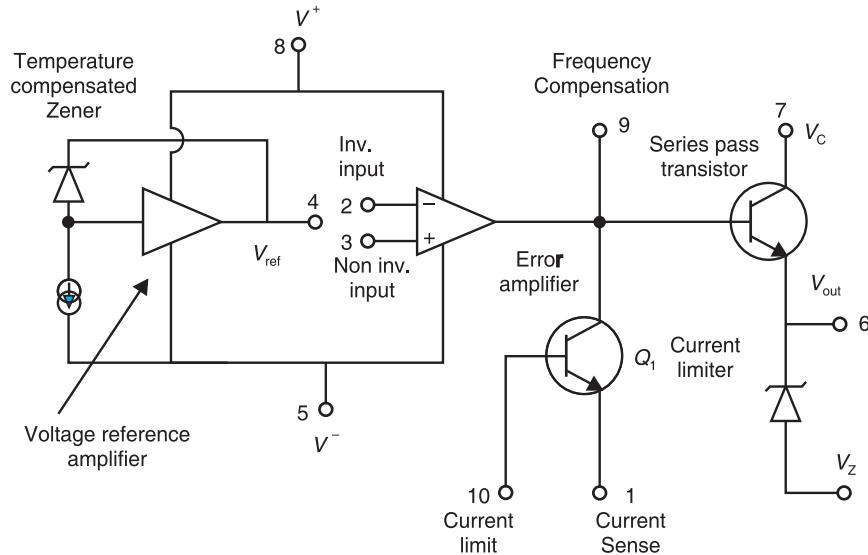


Fig. 39.28

A reference voltage is developed across the zener diode which is temperature compensated. The reference amplifier acts as a buffer. So the constant reference voltage is available at its output. The control amplifier has two inputs, one inverting and the other non-inverting. This amplifier compares the reference with a fixed part of the feedback output voltage and the resultant error is amplified. This error voltage controls the series transistor such that the output voltage remains at a constant level. Transistor  $Q_1$  is used for current limiting.

The load current can be limited by providing a small resistance  $R_{SC}$  between the current limit and current sense terminals. The voltage across  $R_{SC}$  is used to bias the current limiting transistor, present inside the chip.

Hence,  $V_{BE} = I_{\text{limit}} R_{SC}$ , where  $V_{BE}$  is the bias voltage and  $I_{\text{limit}}$  is the limiting current.

$$R_{SC} = \frac{V_{BE}}{I_{\text{limit}}}.$$

**Regulated power supplies.** Almost all electronic circuits require a direct voltage supply. This is usually derived from the standard industrial or domestic ac supply by transformation, rectification, and filtering. The resultant *raw dc* is not stable enough for most purposes. It usually contains an unacceptably large ac ripple waveform. Voltage regulator circuits are employed to render the voltage more constant and to attenuate the ripple.

### EXERCISE

1. What is an integrated circuit? Give an account of the fabrication of a monolithic integrated circuit.
2. Give the block diagram of an operational amplifier.
3. Draw the basic circuit of a differential amplifier and obtain expression for the differential mode voltage gain.

4. Explain the terms “Common mode gain” “differential gain” and “CMRR” of an OP – Amp.  
*(Madras 1994)*
5. What are the different characteristics of an ideal operational amplifier?  
*(Osmania 1993)*
6. Draw the circuit diagram of an inverting amplifier and explain its operation. Derive an expression for its voltage gain.  
*(Andhra, 1995)*
7. What is an Op-Amp? Explain with necessary theory how an Op-Amp can be used as (a) differentiator (b) integrator and (c) comparator.  
*(Madras 1995)*
8. Explain how an op amp can be used as a summing amplifier.
9. Draw a diagram to illustrate how an op-amp can be used as a noninverting amplifier. Explain how it functions.
10. What is a voltage follower? What are its important characteristics?
11. Explain the principles of obtaining regulated power supply.

# COMMUNICATIONS ELECTRONICS

**After reading this chapter, you should be able to**

- ◆ Design analogue and digital electronic circuits at block level
- ◆ Discuss the fundamentals of amplitude modulation and frequency modulation
- ◆ Discuss basic superheterodyne receivers
- ◆ Explain the basic principles and fundamental concepts of TV system
- ◆ Explain the working of CRO and mention its uses
- ◆ Apply Thevenin's theorem and Norton's theorem to a circuit

**RADIO COMMUNICATION**

## 40.1 INTRODUCTION

In radio broadcasting, an A.F. signal (speech, music, etc.) has to be transmitted to distant places. It is done by superimposing the given A.F. signal on an R.F. signal which will serve as a *carrier*. This process of superimposing an A.F. signal on an R.F. carrier wave is called *modulation*. The resulting wave is known as a modulated wave. At the receiving end, the original signal is extracted from the modulated wave. This is called *demodulation* or detection.

**Definition.** *Modulation is defined as the process by which some characteristic property of a high frequency wave is varied in accordance with the instantaneous value of the signal.*

The high frequency wave on which the signal is superimposed is called the carrier wave. Let the carrier wave be represented by

$$e_c = E_c \cos(\omega_c t + \theta).$$

Here,  $E_c$ ,  $\omega_c$  and  $\theta$  are its peak amplitude, angular frequency, and phase respectively. In commercial transmitters, modulation is achieved by varying any one of the above three parameters in accordance with the signal, while keeping the remaining two parameters strictly constant. Thus, there are three types of modulation:

1. Amplitude modulation in which amplitude  $E_c$  is varied keeping  $\omega_c$  and  $\theta$  constant.
2. Frequency modulation in which the frequency  $\omega_c$  is varied keeping  $E_c$  and  $\theta$  constant.
3. Phase modulation in which the phase  $\theta$  is varied keeping  $E_c$  and  $\omega_c$  constant.

### Need for modulation

(1) **Practical antenna length.** For efficient radiation of waves from the antenna, the length of the transmitting antenna should be approximately equal to the wavelength of the wave.

(i) For example, if the frequency of the audio signal is 1000 hertz, the length of the antenna must be

$$\frac{3 \times 10^8}{1000} = 3 \times 10^5 \text{ metres} = 300 \text{ kilometers.}$$

This is practically impossible.

(ii) If the electromagnetic waves radiated by the antenna is of high frequency. (Radio frequency greater than 20 kilohertz), the length of the antenna required will be conveniently small.

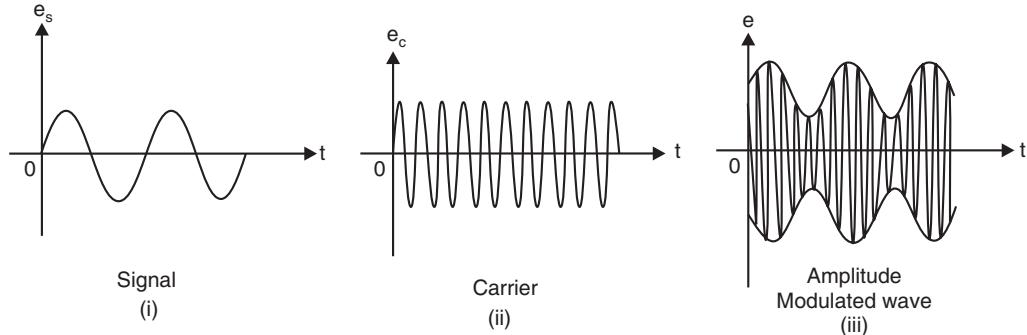
For example, if a carrier wave of 1000 kHz is used to carry the signal, we need an antenna length of 300 metres only. This size can be easily constructed.

(2) **Power.** With high frequency electromagnetic wave, more power can be transmitted into space. The audio signal frequencies are small. So they cannot be transmitted over large distances if radiated directly into space. The only practical solution is to modulate a high frequency carrier wave with audio signal and permit the transmission to occur at this high frequency (i.e., carrier frequency).

(3) **Wireless communication.** One desirable feature of radio transmission is that it should be carried without wires i.e., radiated into space. At audio frequencies, radiation is not practicable because the efficiency of radiation is poor. However, efficient radiation of electrical energy is possible at high frequencies ( $> 20$  kHz). For this reason, modulation is always done in communication systems.

## 40.2 AMPLITUDE MODULATION

In amplitude modulation, the amplitude of the carrier is varied in accordance with the instantaneous values of the signal voltage. Here frequency and phase of the carrier remain constant. The waveforms for audio signal (modulating signal), carrier and amplitude modulated signal are shown in Fig. 40.1.



**Fig. 40.1**

**Modulation index or modulation factor.** It is defined as the ratio of the peak voltage (amplitude) of the signal to the peak voltage (amplitude) of the carrier, i.e.,

$$\text{Modulation Index } m = \frac{E_s}{E_c}.$$

### Mathematical analysis of amplitude modulated wave

A carrier wave is represented by

$$e_c = E_c \cos \omega_c t \quad \dots(1)$$

Here,

$e_c$  = instantaneous carrier voltage,

$E_c$  = amplitude of carrier,

$\omega_c = 2\pi f_c$  ( $f_c$  = frequency of carrier)

The audio signal (modulating signal) is represented by

$$e_s = E_s \cos \omega_s t \quad \dots(2)$$

Here,

$e_s$  = instantaneous voltage of signal

$E_s$  = amplitude of audio signal.

$\omega_s = 2\pi f_s$ , ( $f_s$  = frequency of audio signal).

But,

$E_s = mE_c$ , where  $m$  is the modulation factor.

Therefore, the signal can be represented by

$$e_s = m E_c \cos \omega_s t \quad \dots(3)$$

AM wave is obtained by varying  $E_c$  in accordance with  $e_s$ .

The amplitude of carrier wave varies at signal frequency  $f_s$ . Therefore, the amplitude of AM wave is given by

$$\begin{aligned} \text{amplitude} &= E_c + mE_c \cos \omega_s t \\ &= E_c (1 + m \cos \omega_s t) \end{aligned}$$

The instantaneous voltage of AM wave is

$$\begin{aligned} e &= \text{amplitude} \times \cos \omega_c t \\ &= E_c (1 + m \cos \omega_s t) \cos \omega_c t \\ &= E_c \cos \omega_c t + \frac{m E_c}{2} (2 \cos \omega_c t \cos \omega_s t) \end{aligned}$$

$$\begin{aligned}
 &= E_c \cos \omega_c t + \frac{m E_c}{2} [\cos(\omega_c + \omega_s)t + \cos(\omega_c - \omega_s)t] \\
 &= E_c \cos \omega_c t + \left[ \frac{m E_c}{2} \right] \cos(\omega_c + \omega_s)t + \left[ \frac{m E_c}{2} \right] \cos(\omega_c - \omega_s)t.
 \end{aligned}$$

1. The AM wave is equivalent to the summation of three sinusoidal waves.

- (a) One having amplitude  $E_c$  and frequency  $f_c$ .
- (b) The second having amplitude  $m E_c/2$  and frequency  $(f_c + f_s)$ .
- (c) The third having amplitude  $m E_c/2$  and frequency  $(f_c - f_s)$ .

2. The AM wave contains three frequencies  $f_c$ ,  $f_c + f_s$  and  $f_c - f_s$ . The first frequency is the carrier frequency. Thus, the process of modulation does not change the original carrier frequency, but produces two new frequencies  $(f_c + f_s)$  and  $(f_c - f_s)$ . They are called sideband frequencies.

3. The sum of carrier frequency and signal frequency, i.e.,  $(f_c + f_s)$  is called upper sideband frequency. The difference between the carrier and the signal frequencies is called the lower sideband frequency, i.e.,  $(f_c - f_s)$ .

From the frequency spectrum (Fig. 40.2), it is clear that USF and LSF are symmetrical w.r.t. the carrier frequency.

**Example 1.** In a broadcasting studio a 100 MHz carrier having an amplitude of 50 V is amplitude modulated by a 5 kHz audio signal having an amplitude of 20 V. Find the modulation index, percentage modulation, frequency components of the modulated wave and their amplitudes.

**Solution.**

$$\begin{aligned}
 E_c &= 50 \text{ V}, \quad E_s = 20 \text{ V} \\
 E_{\max} &= E_c + E_s = 50 + 20 = 70 \text{ V} \\
 E_{\min} &= E_c - E_s = 50 - 20 = 30 \text{ V}
 \end{aligned}$$

The modulation index is given by,

$$m = \frac{E_{\max} - E_{\min}}{E_{\max} + E_{\min}} = \frac{70 - 30}{70 + 30} = 0.4$$

Percentage of modulation =  $0.4 \times 100 = 40\%$

carrier frequency,  $f_c = 100 \times 10^6 \text{ Hz}$

signal frequency,  $f_s = 5 \times 10^3 \text{ Hz}$ .

The AM wave contains three frequencies,  $f_c$ ,  $f_c + f_s$ ,  $f_c - f_s$

$f_c$  – carrier frequency = 100 MHz

$f_c + f_s$  = upper sideband frequency

$$= 100 \text{ MHz} + 5 \text{ kHz} = 1,00,005 \text{ kHz}$$

$f_c - f_s$  = lower sideband frequency

$$= 100 \text{ MHz} - 5 \text{ kHz} = 99995 \text{ kHz}$$

carrier amplitude =  $E_c = 50 \text{ V}$

$$\text{Amplitude of sidebands} = 2 \times m \cdot \frac{E_c}{2} = 2 \times 0.4 \times \frac{50}{2} = 20 \text{ V}.$$

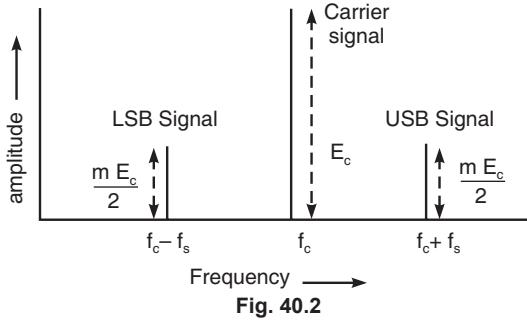


Fig. 40.2

#### 40.2.1. Power in an Amplitude Modulated Wave

The AM wave has three components:

(a) carrier wave of amplitude  $E_c / \sqrt{2}$ . (r.m.s. value)

(b) USF wave of amplitude  $mE_c / (2\sqrt{2})$ . (r.m.s. value)

(c) LSF wave of amplitude  $mE_c / (2\sqrt{2})$ . (r.m.s. value)

The power is calculated by using the formula,

$$\text{power} = \frac{E_{rms}^2}{R}$$

$$\text{Carrier power} = P_c = \frac{(E_c / \sqrt{2})^2}{R} = \frac{E_c^2}{2R} \quad \dots (1)$$

$$\text{Total sideband power} = P_s = \frac{(mE_c / 2\sqrt{2})^2}{R} + \frac{(mE_c / 2\sqrt{2})^2}{R}$$

$$P_s = \frac{m^2 E_c^2}{4R} \quad \dots (2)$$

Total power of AM wave,

$$\begin{aligned} P_T &= P_c + P_s \\ &= \frac{E_c^2}{2R} + \frac{m^2 E_c^2}{4R} = \frac{E_c^2}{2R} \left[ 1 + \frac{m^2}{2} \right] \end{aligned}$$

or

$$P_T = \frac{E_c^2 [2 + m^2]}{2R} \quad \dots (3)$$

$$\therefore \text{carrier power} = P_c = P_T \left( \frac{2}{2 + m^2} \right) \quad \dots (4)$$

Dividing Eq. (2) by Eq. (3), we get

$$\frac{P_s}{P_T} = \frac{m^2}{2 + m^2} \quad \dots (5)$$

$$\text{Total sideband power} = P_s = P_T \left( \frac{m^2}{2 + m^2} \right).$$

**Example 1.** A carrier wave of 500 watt is subjected to 100% amplitude modulation. Determine (1) Power of modulated wave, and (2) Power in sidebands.

**Solution.** Power of modulated waves is given by

$$P_T = P_c [1 + m^2/2]$$

Here,

$$P_c = \text{power in carrier wave} = 500 \text{ watt}$$

$$m = \text{modulation index} = 1 \text{ (for 100% modulation)}$$

$$P_T = 500 [1 + 1/2] = 750 \text{ watts}$$

$$\text{Power in sidebands, } P_s = P_T - P_c = 750 - 500 = 250 \text{ watt.}$$

There is 125 watt power in upper sideband and 125 watt in lower sideband.

**Example 2.** The total power content of an AM wave is 1500 watt for a hundred percent modulation. Determine (i) power transmitted by carrier; (ii) power transmitted by each sideband. What is your inference?

**Solution.** Carrier power  $P_C = P_T \left( \frac{2}{2 + m^2} \right)$

When modulation is 100%, then  $m = 1$ .

$$\therefore \text{Power transmitted by carrier} = 1500 \left( \frac{2}{2+1} \right) = 1000 \text{ W.}$$

$$\begin{aligned} \text{Power transmitted by sidebands } P_S &= P_T \left( \frac{m^2}{2 + m^2} \right) \\ &= 1500 (1/3) = 500 \text{ W.} \end{aligned}$$

$$\therefore \text{Power transmitted by each sideband} = 500/2 = 250 \text{ W.}$$

The carrier power is that required for transmission and sideband power represents the signal content. Of these two, the power due to carrier is much greater than the other. But the message or signal is present only in the latter, i.e., in the sidebands. Hence most of the power from the modulating source goes to a component which does not contain the signal. If we can suppress the unwanted, large power consuming carrier component, then the power drawn from the source will be less and whatever is drawn from the source will go to the useful sidebands. This type of transmission is economical and is called carrier suppressed sideband transmission.

### 40.3 FREQUENCY MODULATION

In frequency modulation, the frequency of the carrier is varied in accordance with the instantaneous values of signal voltage. Here the amplitude and phase of the carrier remain constant.

The frequency of the carrier before modulation is called the centre frequency or resting frequency usually denoted by  $f_c$ . The variation of frequency either above or below the centre frequency is known as frequency deviation  $\Delta f$ . The frequency deviation is proportional to amplitude of the modulating signal. The maximum permitted frequency deviation is 75 kHz. So bandwidth for FM is  $2 \times 75$  or 150 kHz. Bandwidth is also referred to as channel width.

In addition, a guardband of 25 kHz is allowed on either side of the channel width to eliminate the interference between the adjacent channels. Therefore in FM, total channel width is 200 kHz. But in AM the channel width is only 10 kHz.

Frequency modulation is illustrated in Fig. 40.3.

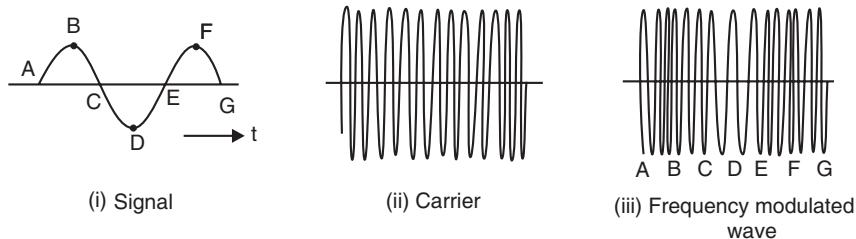


Fig. 40.3

- The amplitude of the modulated wave remains the same as the carrier amplitude.
- The frequency of the carrier wave varies in proportion to the instantaneous value of amplitude of the signal.

- When the signal voltage is zero as at A, C, E and G, the carrier frequency is not changed.
- When the signal approaches its positive peaks as at B and F, the carrier frequency is increased to maximum as shown by the closely spaced cycles.
- During the negative peak of the signal, as at D, the carrier frequency is reduced to minimum as shown by the widely spaced cycles.

At present, frequency modulation is employed for audio signal and amplitude modulation for video signal.

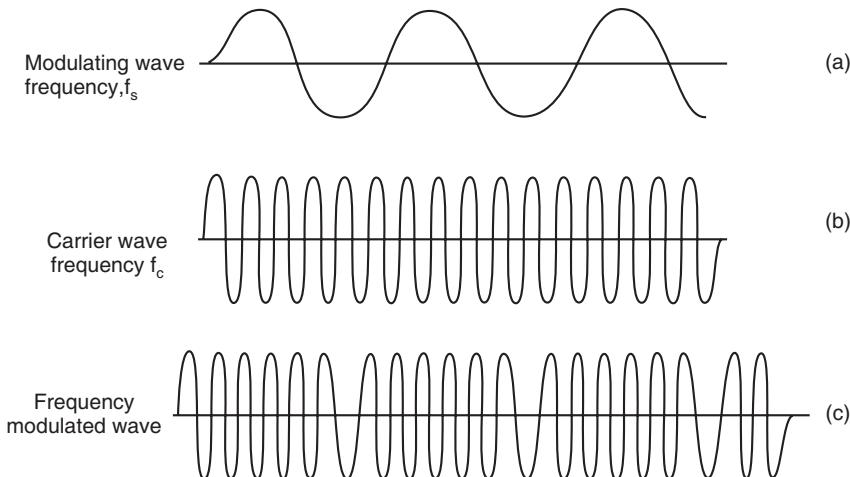
- The standard FM broadcast band consists of carrier frequencies from 88 MHz to 108 MHz, which is significantly higher than AM.
- In amplitude modulation, the parameter of amplitude is varied. In frequency modulation (FM), the frequency of a carrier is varied above and below its normal or at-rest value by a modulating signal.

### Theory of Frequency modulation

- In frequency modulation (FM), the amplitude of the carrier is kept constant but the frequency  $f_c$  of the carrier is varied by the modulating signal.
- The carrier frequency  $f_c$  varies at the rate of the signal frequency  $f_s$ . The frequency deviation is proportional to the instantaneous amplitude of the modulating signal. Maximum frequency deviation is  $(f_{c(\max)} - f_c)$ . It occurs at the peak voltage of the modulating signal.

### Equation of FM wave

The frequency of carrier wave [Fig. 40.4 (b)] is changed in accordance with the amplitude of modulating wave [Fig. 40.4(a)]. The amplitude of modulated wave [Fig. 40.4 (c)] remains constant as that of carrier wave.



**Fig. 40.4**

The carrier frequency is varied sinusoidally at signal frequency.

The instantaneous deviation in frequency from the carrier is proportional to the instantaneous amplitude of the modulating signal.

The instantaneous angular frequency of FM is

$$\omega_i = \omega_c + \Delta\omega_c \cos \omega_s t \quad \dots (1)$$

Here,  $f_c = \frac{\omega_c}{2\pi}$  is called centre frequency.

Total phase angle  $\theta = \omega t$  so that if  $\omega$  is variable, then,

$$\begin{aligned}\theta &= \int_0^t \omega_i dt \\ &= \int_0^t (\omega_c + \Delta\omega_c \cos \omega_s t) dt \\ \therefore \theta &= \omega_c t + \frac{\Delta\omega_c}{\omega_s} \sin \omega_s t\end{aligned}\dots(2)$$

The term  $\frac{\Delta\omega_c}{\omega_s}$  is called *modulation index*  $m_f$ .

$$\therefore \theta = \omega_c t + m_f \sin \omega_s t \dots(3)$$

The instantaneous value of FM voltage wave is given by:

$$\begin{aligned}e &= E_c \cos \theta \\ \text{or } e &= E_c \cos (\omega_c t + m_f \sin \omega_s t)\end{aligned}\dots(4)$$

Eq. (4) is the general voltage equation of a frequency modulated wave. Such a wave is shown in Fig. 40.4 (c).

**Modulation index.** The modulation index  $m_f$  is the ratio of maximum frequency deviation ( $\Delta f$ ) to the frequency ( $= f_s$ ) of the modulating signal i.e.,

$$\text{Modulation index, } m_f = \frac{\Delta\omega_c}{\omega_s} = \frac{f_{c(\max)} - f_c}{f_s} = \frac{\Delta f}{f_s} \dots(5)$$

**Frequency spectrum.** If  $f_c$  and  $f_s$  are the carrier and signal frequencies respectively, then FM spectrum will have the following frequencies:

- $f_c, f_c \pm f_s, f_c \pm 2f_s, f_c \pm 3f_s$  and so on
- $f_c + f_s, f_c + 2f_s, f_c + 3f_s \dots$  are the upper sideband frequencies.
  - $f_c - f_s, f_c - 2f_s, f_c - 3f_s \dots$  are the lower sideband frequencies.

## 40.4 PHASE MODULATION

Let the carrier wave be represented by:

$$e_c = E_C \cos (\omega_c t + \theta) \dots(1)$$

Here,  $\omega_c$  is the angular frequency of the carrier wave,  $E_C$  is its amplitude and  $\theta$  is the phase. In phase modulation, the phase  $\theta$  is varied keeping  $E_C$  and  $\omega_c$  constant.

**Definition.** In the process of phase modulation, the instantaneous phase of the carrier is varied by an amount that is proportional to the instantaneous amplitude of the signal and at a rate that is proportional to modulating frequency. The amplitude of carrier remains unaltered.

### Equation of a phase modulated wave

- The phase angle  $\theta$  is modulated depending on the incoming signal.

The modulating signal has the form

$$e_s = E_s \cos \omega_s t \dots(2)$$

The instantaneous phase for phase modulated wave is

$$\phi(t) = \omega_c t + K_p E_s \cos \omega_s t$$

Here, proportionality factor  $K_p$  determines the maximum variation in phase for a given signal strength.

Therefore, the expression for phase modulated wave can be written as

$$e_c = E_c \cos (\omega_c t + K_p E_s \cos \omega_s t) \quad \dots (3)$$

The product  $K_p E_s$  is called peak phase deviation.

$$K_p E_s = \Delta\phi$$

Eq. (3) can be written as

$$e_c = E_c \cos (\omega_c t + \Delta\phi \cos \omega_s t) \quad \dots (4)$$

- $\Delta\phi$  determines the characteristics of phase modulated wave in the same manner as the factor  $m_f$  determines characteristics of frequency modulated wave.  
 $\Delta\phi$  is called the *phase modulation index*  $m_p$ .

## 40.5 BLOCK DIAGRAM OF AM TRANSMITTER

Fig. 40.5 shows the block diagram of amplitude-modulated radio-transmitter.

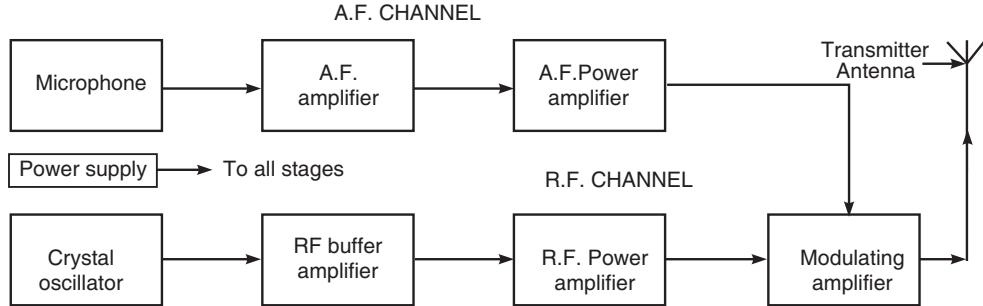


Fig. 40.5

- Radio transmitters produce radio waves and transmit them. The information is contained in the audio signal. AM transmitter consists of (i) RF channel, and (ii) AF channel.

**(i) RF Channel.** The production, amplification and modulation of the carrier wave takes place in this part.

- **RF master oscillator.** The high frequency carrier wave is generated by a crystal controlled master oscillator.  
It is used to operate the transmitter at desirable fixed radio frequency.
- **RF buffer amplifier.** It isolates the master oscillator from succeeding stages so that variations of coupling and antenna loading do not affect oscillator frequency generated by it.
- **RF power amplifier.** The output of the crystal controlled oscillator is power amplified by RF power amplifier. The buffer is a low gain amplifier which isolates the RF power amplifier from the oscillator. This arrangement keeps the frequency of the crystal controlled oscillator constant.
- **Modulating amplifier.** The last stage of the RF channel is the modulating amplifier. In this amplifier, the *RF* wave and the modulating *AF* signal are mixed to produce the *AM* wave. Its function is to modulate the amplitude of RF carrier wave in accordance with the amplified audio signal.

Finally, modulated waves are fed to transmitter antenna.

The transmitter antenna radiates modulated waves into space.

(ii) **AF Channel.** Modulating signal is produced and amplified in this channel.

- **Microphone.** The microphone converts sound waves into AF electrical signals.
- **Audio amplifier.** The AF signals are amplified first by an AF voltage amplifier and then by an AF power amplifier. The signal from this part is used as modulating signal. The output of the AF power amplifier is fed into the modulating amplifier.

## 40.6 DEMODULATION OR DETECTION (LINEAR DIODE DETECTOR)

Demodulation is the process of recovering the audio signal from the modulated wave. The process of detecting an AM carrier wave consists of two steps:

*Step 1.* The negative half-cycles of the modulated carrier wave have to be removed by rectification.

*Step 2.* The carrier frequency must be removed by using suitable filters. Fig. 40.6 illustrates the process of detecting an AM carrier wave.

**Circuit Details.** Fig. 40.7 shows the diode detector circuit.

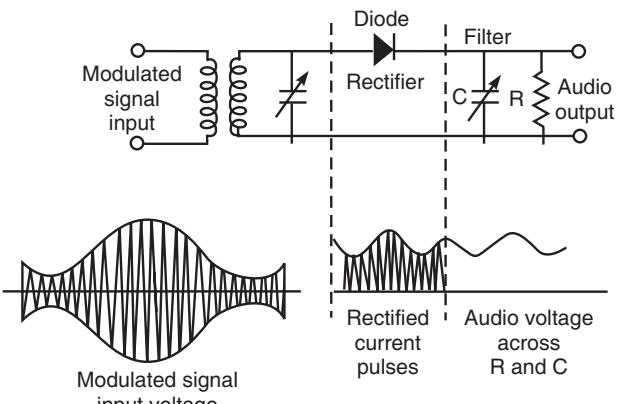


Fig. 40.6

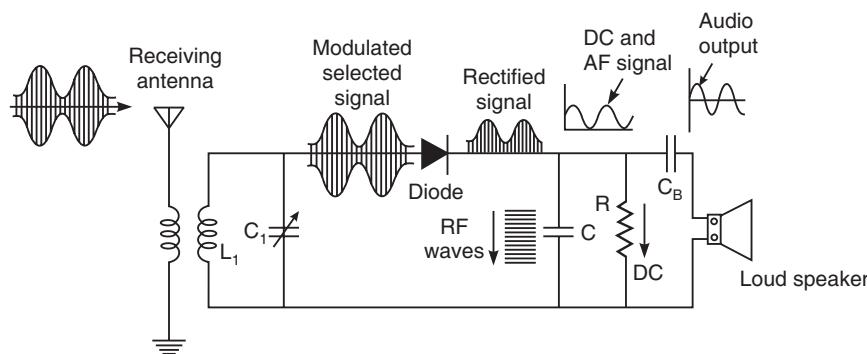


Fig. 40.7

The input to the detector is an amplitude modulated carrier voltage. Tuned circuit  $L_1C_1$  selects the modulated wave of desired frequency. Capacitor  $C$  offers a low impedance to the carrier while  $R$  is a relatively high resistance.  $C_B$  is a blocking capacitor which can pass low frequency *AF* signals only, but blocks *DC* from appearing in the output.

**Circuit operation.** By varying the capacitor  $C_1$ , the resonant frequency of tuned circuit  $L_1C_1$  can be varied and *RF* signal of any desired frequency can be tuned in. Thus the modulated wave of desired frequency is selected by the parallel tuned circuit  $L_1C_1$  and is applied to the *p-n* junction diode. During the positive half cycle of modulated wave, the diode conducts but not during negative half cycles. As a result of this rectifying action, we get only the positive half cycles of the modulated wave in the output of diode.

This rectified modulated wave consists of radio frequency and the signal and cannot be applied directly to the speaker for sound production. The *RF* carrier wave is filtered out by the capacitor  $C$  which presents a low reactance path to *RF* components but a relatively high reactance to the audio signals. The *DC* component of the remaining signal can not pass through blocking capacitor  $C_B$ .

and is shunted out through  $R$ . The low frequency audio signal can easily pass through  $C_B$ . Thus modulating signal is recovered at the detector output and passed on to the speaker for reproduction of sound.

## 40.7 BLOCK DIAGRAM OF SUPERHETERODYNE AM RECEIVER

Fig. 40.8 shows the block diagram of a superheterodyne AM receiver.

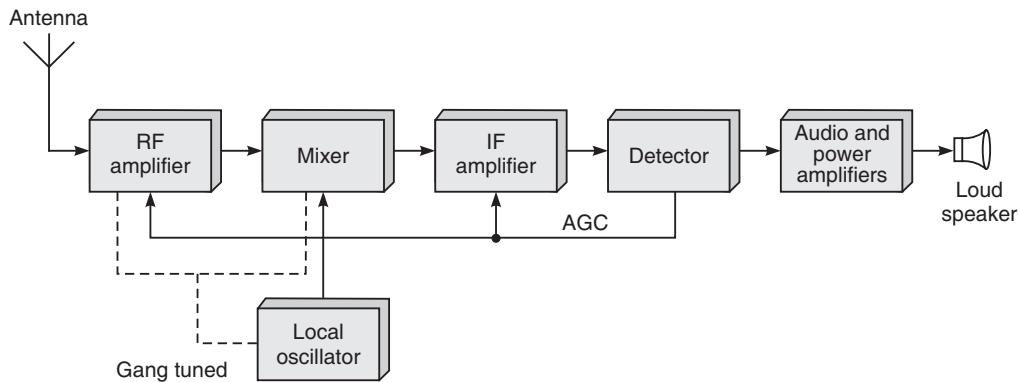


Fig. 40.8

- It consists of an antenna, an RF (radio frequency) amplifier, a mixer, a local oscillator (LO), an IF (Intermediate frequency) amplifier, a detector, an audio amplifier, a power amplifier, and a speaker.

**Principle.** In this receiver the incoming R.F. modulated signal, after amplification, is mixed with an unmodulated R.F. signal generated in the receiver itself by a local oscillator. This results in a new signal with the same modulation but a *lower* carrier frequency called ‘intermediate frequency’ (I.F.). This new signal is then amplified and detected. The intermediate frequency is usually 455 kHz.

### Working of different stages

**Antenna.** The antenna picks up all radiated signals and feeds them into the RF amplifier. These signals are very small (usually only a few microvolts).

**RF amplifier.** This circuit can be adjusted (tuned) to select and amplify any carrier frequency within the AM broadcast band. Only the selected frequency and its two side bands pass through the amplifier.

**Local oscillator.** This circuit generates a steady sine wave at a frequency 455 kHz above the selected RF frequency.

**Mixer.** This circuit accepts two inputs, the amplitude modulated RF signal from the output of the RF amplifier and the sinusoidal output of the local oscillator (LO). These two signals are then “mixed”.

- When mixed, the sum ( $f_0 + f_s$ ) and difference ( $f_0 - f_s$ ) frequencies are formed. By a suitable arrangement only the difference frequency is extracted.

$$f_0 - f_s = 455 \text{ kHz} = f_i$$

**IF Amplifier.** The input to the IF amplifier is the 455 kHz AM signal, a replica of the original AM carrier signal except that the frequency has been lowered to 455 kHz. The IF amplifier significantly increases the level of this signal. The output of this stage is fed into the detector.

**Detector.** Here the A.F. modulating signal is separated from the carrier. The output of the detector consists of only the audio signal.

**Audio and power amplifiers.** The detected AF signal is amplified by the audio and power amplifiers and then applied to a loudspeaker. The speaker converts audio electric signal into sound waves.

**Automatic gain control (AGC).** Due to various reasons, the strength of the incoming signal may change. Consequently, the output of the loud speaker also changes and this leads to unpleasant hearing. To avoid this AGC is used, which maintains the output at constant value. Hence AGC is a system by means of which, the overall gain of a radio receiver is varied automatically with the changing strength of the received signal, to keep the output substantially constant.

The receiver can be tuned to accept any frequency in the AM band. The RF amplifier, mixer, and local oscillator are tuned simultaneously so that the LO frequency is always 455 kHz above the incoming RF signal frequency. This is called *gang tuning*.

#### Characteristics of a Radio-receiver

The performance of a radio-receiver is judged by its following characteristics:

- (i) **Sensitivity.** It determines the minimum strength of an input signal necessary for producing a standard output signal.
- (ii) **Selectivity.** It determines the extent to which the radio receiver can differentiate between the desired signal and the unwanted signals of other frequencies.
- (iii) **Fidelity.** It is a measure of the receiver's ability to reproduce faithfully at its output all the essential characteristics of the input signal.

### 40.8 AN INTEGRATED CIRCUIT FM TRANSMITTER

Fig. 40.9 shows a single-chip FM transmitter using the MC2833D FM modulator.

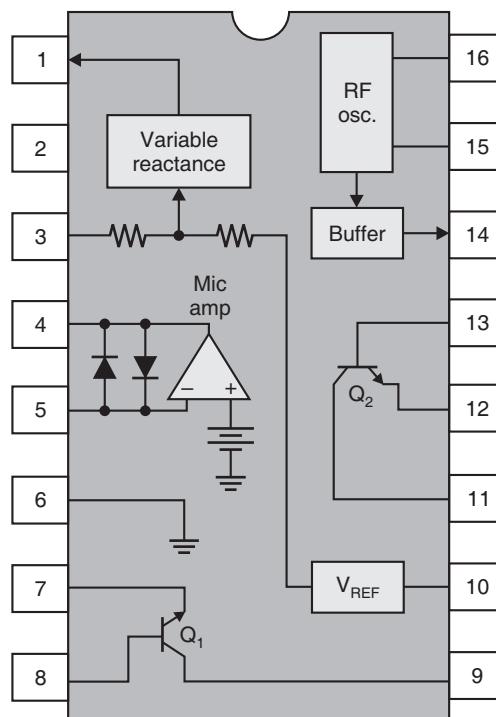


Fig. 40.9

- The 16-pin package configuration shows the basic functional blocks.
- It is designed for cordless telephone and other FM communications equipment.

### Working of the major functional blocks

- The microphone amplifier (mic amp) amplifies the low-level input from a microphone and feeds its output to the variable-reactance circuit which uses a varactor diode tuning circuit to control the RF oscillator. The reference voltage circuit ( $V_{REF}$ ) provides stable bias to the reactance circuit.
- The two individual transistors can be connected as tuned amplifiers to boost the power output.

#### 40.8.1. FM Transmitter

- The carrier frequencies of the radio FM transmitters (that emit the program for “broad audience”) are placed in the waveband from 88 MHz to 108 MHz.
- The maximum frequency shift of the transmitter (during the modulation) is  $\pm 75$  kHz.

**Block diagram.** Fig. 40.10 shows the block diagram of an Armstrong FM transmitter. It is the most frequently used one.

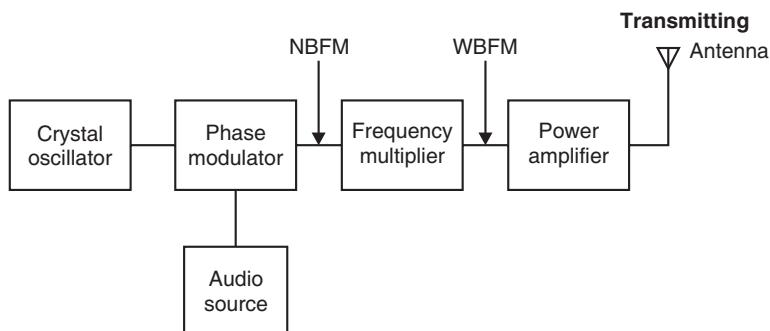


Fig. 40.10

### Working of the different stages

**Crystal oscillator.** The crystal oscillator generates the stable carrier signal.

**Phase modulator.** The modulating signal and the carrier signal are applied to the phase modulator operating in the low power level to generate a narrowband FM wave (Fig. 40.11).

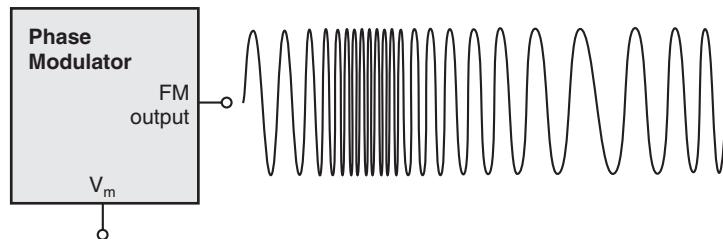


Fig. 40.11

Information being transferred, *i.e.*, the modulating signal, is signal from some LF source.

**Frequency multiplier.** The narrowband FM wave is then passed through several stages of frequency multipliers to increase the frequency deviation and also carrier signal frequency to the required level. The several stages of frequency multiplication helps in choosing a suitable combination for achieving the required level of multiplication factors needed for deviation and carrier signal frequency.

**Power amplifier.** The output of the frequency multipliers stage will be a wideband FM, but at the low power level. The WBFM is then passed through one or more stages of power amplifiers to add required power levels. The WBFM with high power is then finally transmitted via the antenna towards the receiver.

## 40.9 DETECTION OF FM WAVE

**Principle.** We detect the modulated signal on to the carrier wave via frequency variations. It changes the frequency deviations of modulated wave into amplitude variation corresponding to the modulating signal. The transformation of frequency variation into amplitude variation should be linear and efficient.

**Methods.** There are several methods for demodulating an FM signal. These include slope detection, phase-shift discrimination, ratio detection, quadrature detection, and phase-locked loop demodulation. We discuss PLL demodulation.

### Phase-locked loop (PLL) demodulation

**Basic concept.** The *phase-locked loop* (PLL) is a feedback circuit consisting of a phase detector, a low pass filter, and voltage controlled oscillator (VCO).

The PLL is capable of locking onto or synchronizing with an incoming signal. When the phase of the incoming signal changes, indicating a change in frequency, the phase detector's output increases or decreases just enough to keep the VCO frequency the same as the frequency of the incoming signal.

**PLL block diagram.** Fig. 40.12 shows PLL block diagram.

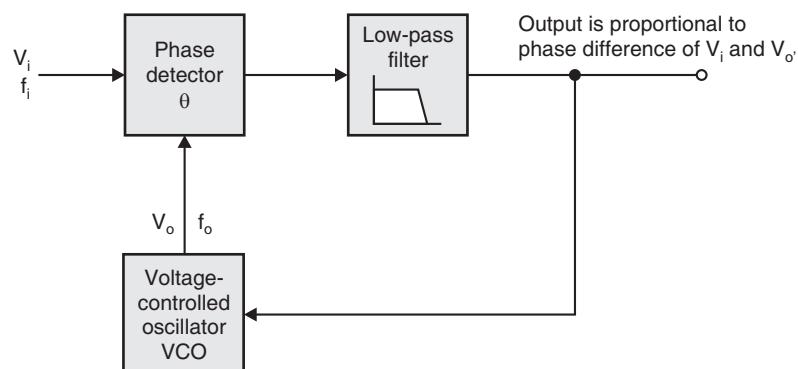


Fig. 40.12

**Operation.** The phase detector compares the phase difference between the incoming signal,  $V_i$ , and the VCO signal,  $V_o$ . When the frequency of the incoming signal,  $f_i$ , is different from that of the VCO frequency,  $f_o$ , the phase angle between the two signals is also different. The output of the phase detector and the filter is proportional to the phase difference of the two signals. This proportional voltage is fed to the VCO, forcing its frequency to move toward the frequency of the incoming signal until the two frequencies are equal. At this point, the PLL is locked onto the incoming frequency. If  $f_i$  changes, the phase difference also changes, forcing the VCO to track the incoming frequency.

## 40.10 THE SUPERHETERODYNE FM RECEIVER

**Principle.** It uses superhetrodyne principle.

- The FM broadcast signals lie in the frequency range between 88 MHz and 108 MHz.
- The IF (intermediate frequency) of an FM receiver is 10.7 MHz.

**Block diagram.** Fig. 40.13 shows the block diagram of a superheterodyne FM receiver.

Receiving 88 MHz–108 MHz 10.7 MHz FM

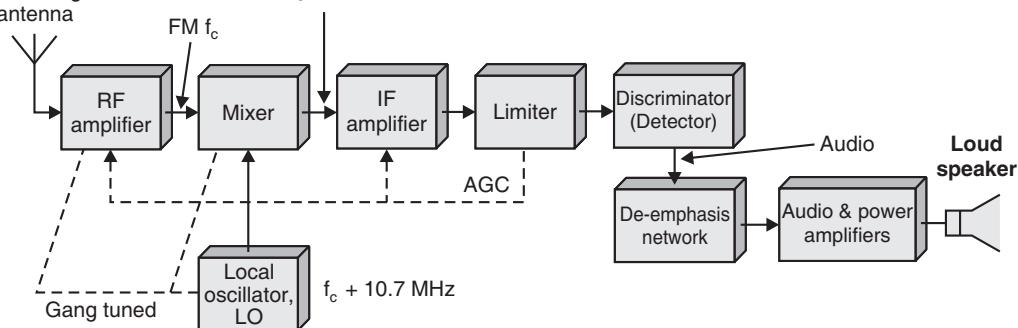


Fig. 40.13

The functions of various stages are as follows:

**Receiving antenna.** It is used to receive a portion of the transmitted waves.

**RF amplifier.** This circuit must be capable of amplifying any frequency between 88 MHz and 108 MHz. It is highly selective so that it passes only the selected carrier frequency and significant side-band frequencies that contain the **audio**.

**Local oscillator.** This circuit produces a sine wave at a frequency 10.7 MHz above the selected RF frequency.

**Mixer.** The output from the RF amplifier is fed to the mixer stage where it is combined with the output signal from a local oscillator. The two frequencies beat together and produce an intermediate frequency (IF). The intermediate frequency (IF) is equal to the difference between oscillator frequency and the RF frequency.

- The output of the mixer is a 10.7 MHz FM signal regardless of the RF carrier frequency.

**IF Amplifier.** This circuit amplifies the 10.7 MHz FM signal.

**Limiter.** Limiter keeps the amplitude variations of signal under control. The limiter produces a constant amplitude FM output at the 10.7 MHz intermediate frequency.

**Discriminator.** The **discriminator** recovers the audio from the FM signal.

- It extracts the intelligence modulated on to the carrier via frequency variations.

**De-emphasis network.** The higher modulating frequencies are amplified more than the lower frequencies at the transmitting end of an FM system by a process called *preemphasis*. The de-emphasis circuit in the FM receiver brings the high frequency audio signals back to the proper amplitude relationship with the lower frequencies.

**Audio and power amplifiers.** This circuit amplifies the detected audio signal and drives the loud speaker to produce sound.

## TELEVISION

### 40.11 TELEVISION CAMERA — THE IMAGE ORTHICON

A schematic diagram of the image orthicon camera is shown in Fig. 40.14.

It works on the principle of photoelectric emission.

Image orthicon has an image section at its one end and an electron gun and electron multiplier at its other end. Deflecting and focusing coils are fixed at the middle.

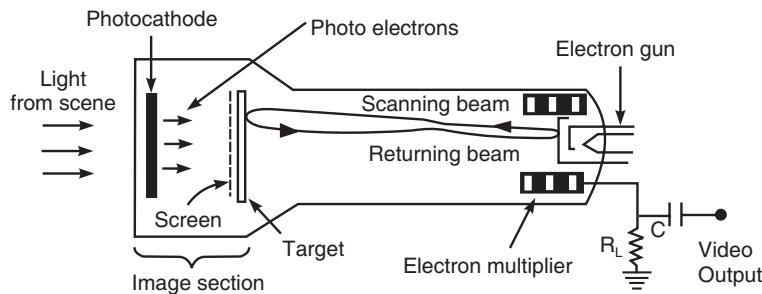


Fig. 40.14

**(i) The Image Section.** The image section has a photosensitive plate called *photocathode*. It is kept at a negative voltage of the order of  $-400$  volt. It is so thin that electrons are emitted from its back side in proportion to the light intensity falling on its front side. A *screen* kept at a slightly positive voltage ( $+1$  volt), having about 400 meshes per cm and a target mounted close to it are kept at the other end of the image section.

When an optical image of a scene is focussed on the photocathode, photoelectrons are emitted from its back side. These electrons are attracted towards the screen, but most of them pass through the meshes and strike the target. As a result of this, secondary electrons are emitted from the target which are drawn to the screen. Thus, the target is left with a positive charge distribution similar to the optical image focussed on the photocathode.

**(ii) Scanning Section.** The back of the target plate is then scanned by an electron beam emitted from the cathode of an electron gun. The number of electrons in the beam is maintained constant. The scanning beam deposits electrons at each point on the back of the target just enough to neutralise the positive charge on the image side. The rest of the electrons in the scanning beam are turned back towards the electron gun. Therefore, the electron density of the returning beam varies in accordance with the positive charge distribution on the target.

**(iii) Electron Multiplier Section.** The electrons in the returning beam are made to strike the aperture of the electron gun which emits secondary electrons in proportion to the number of electrons in the incident beam. These secondary electrons are then deflected into an electron multiplier system whose output is the amplified reproduction of the returning beam. Thus, a video signal corresponding to the scene projected on the photocathode is available across the load  $R_L$ . This video signal is amplified, modulated and transmitted.

#### Scanning Process in TV

Scanning is the *process of exploring the scene to be reproduced*. An actual scene, in general, is characterised by a distribution of light and shade over it. The television system at the transmitter breaks a scene into many small elements whose size corresponds to the smallest details of the scene. These are known as the *picture elements*. In the process of scanning, the variations of light intensity are converted into electrical impulses by using the phenomenon of photoelectricity. The electrical impulses of different picture elements are picked up one after another and then transmitted rapidly in sequence through a single channel. At the receiving end, they are reassembled in their proper sequence to create a replica of the original image. Due to rapid scanning and persistence of vision, all the picture elements appear simultaneously before our eyes.

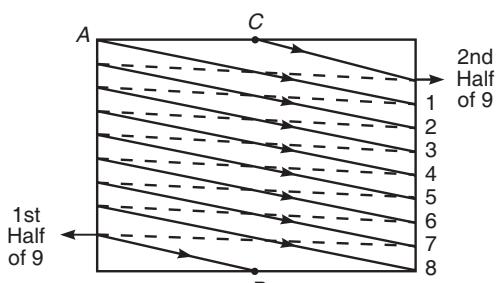


Fig. 40.15

The *resolution or definition* of television is determined by the number of elements into which a scene is broken. Larger the number of elements, better is the definition and finer is the image.

Now a days, the method of *interlaced scanning* is used widely.

**Interlaced Scanning.** Schematic illustration of linear interlaced scanning is given in Fig. 40.15. Each picture (frame) is divided into two parts called “fields” and 50 views of the scene are presented to the eye per second. One field consists of odd lines and the other field consists of even lines. Fig. 40.15 shows the interlaced technique of 9-line system. All the odd lines (1st field) in the frame are scanned first from top to bottom skipping over the even lines. After completing all the odd lines, the electron beam retraces back to the top of the frame to start scanning the even lines. Since 25 frames are scanned per second, and each frame is scanned in two fields, the field scanning frequency is  $25 \times 2 = 50$  Hz. By dividing the picture into two fields, interlaced scanning eliminates *flicker*.

## 40.12 COLOUR TELEVISION

In colour TV system, additional circuits are used for colour transmission and reception with black and white TV system. All the colours can be reproduced by properly combining the three primary colours : *red*, *green* and *blue*. For transmission of colour pictures, therefore, the object at the transmitting station is televised with three TV cameras, fitted with optical filters, to allow transmission of these three colours respectively. By combining the colour outputs of these cameras, two basic signals are produced for transmission over the 6 MHz channel. These two signals are respectively known as the *luminance signal* and the *chrominance signal*.

**The luminance signal.** Luminance signal contains only the brightness changes of the picture, as does the ordinary video signal. A conventional black-and-white receiver will receive only this luminance signal in a colour telecast. Therefore, the reproduction will be in black-and-white.

**The chrominance signal.** The colour information is contained in the chrominance signal. The chrominance signal is a 3.58 MHz signal modulated by colour information.

A colour TV receiver combines the luminance and the chrominance signals and recovers the red, green, and blue video signals in their proper proportion. These signals are amplified and fed to a tricolour picture tube for the display of the coloured image on the viewing screen.

**The colour picture tube.** A tricolour picture tube, known as *shadow-mask colour tube* is shown in Fig. 40.16. It has three electron guns (red, green and blue) in order to produce three electron beams for primary colours. The amplified chrominance signal which contains red, green and blue colours is applied to corresponding electron guns. The electron guns are so positioned that their beams cross at some distance from the fluorescent screen. A thin metal sheet or *shadow-mask* containing

a large number of very fine holes is placed at the plane where the beams cross over. The viewing screen is a glass plate deposited with closely placed groups of phosphor-dot trios. Each trio consists of three phosphor dots which, when hit by electrons, emit red, green and blue light, respectively. A particular hole in the shadow mask is coupled with a particular trio on the screen. The electrons from

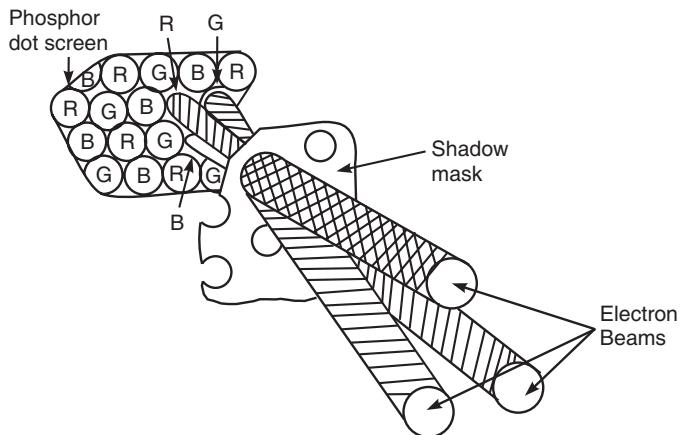


Fig. 40.16

the “red” gun can strike only red emitting dots. Similarly, the electrons from the “green” and “blue” guns can hit only green emitting dots and blue emitting dots, respectively. If the number of electrons supplied by each gun is separately controlled, the proportions of red, green, and blue in any trio can be varied. Viewing from a large distance from the screen, the eye cannot resolve the individual dots but can only sense a colour given by the proportions of red, green and blue light emitted from the trio. The complete colour picture is produced by scanning the screen with the three beams deflected together.

#### 40.12.1. Colour Television Camera

Fig. 40.17 shows the block diagram of colour TV camera.

Light from the scene, after passing through the lens  $L$ , is incident on a dichroic mirror system.  $M_1$  and  $M_2$  are total reflecting and  $M_B$  and  $M_R$  are beam splitting (Partially reflective and partially transmissive).

(i) The dichroic mirror  $M_B$  reflects blue light and permits red and green lights to pass through it. The blue light after reflection at the mirror  $M_1$  enters the blue camera tube via the blue filter. The blue filter allows only blue light, filtering other colours, if any.

(ii) The red and green lights from  $M_B$  fall on dichroic mirror  $M_R$ , which reflects red light and transmits green colour through it. The green light after passing through the green filter, enters the green camera tube.

(iii) The red light from  $M_R$  gets reflected at the mirror  $M_2$  and passes into the red camera via red filter.

In this way the red, green and blue colour content of a picture is separated into three different light signals.

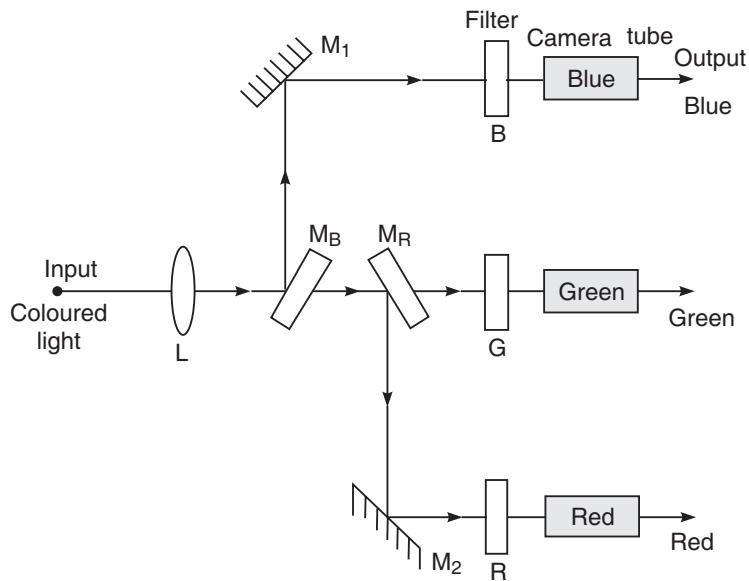


Fig. 40.17

#### 40.12.2 Transmission of Colour TV Signal

Fig. 40.18 shows the block diagram of colour TV transmitter.

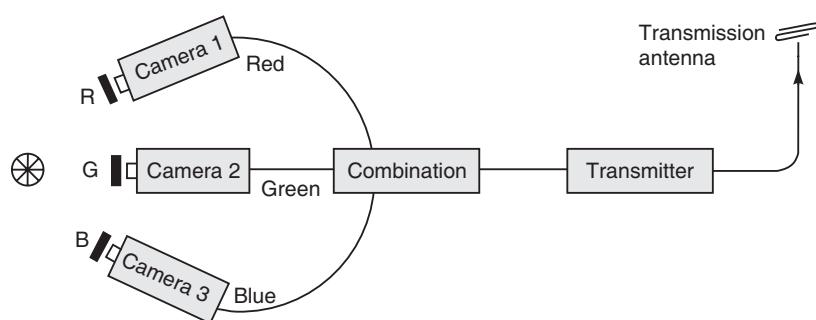


Fig. 40.18

The camera system consists of three ordinary television cameras, all focussed on the object to be televised. The red, green and blue filters are placed in front of the cameras 1, 2 and 3. The signals from the three cameras are combined in a special way and telecast together.

#### 40.12.3. Reception of colour TV Signal

Fig. 40.19 shows the block diagram of colour TV Receiver.

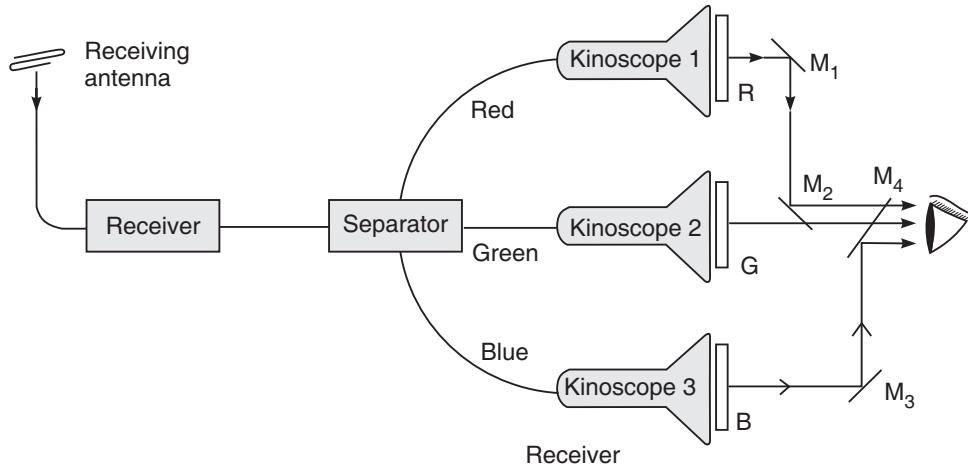


Fig. 40.19

At the receiver, the three signals are received together and then separated. Each signal is used to operate a separate kinescope in front of whose screen is a second filter of the same colour. The camera behind each colour filter responds to parts of the object having the same colour and therefore the corresponding kinescope of the receiver shows only that colour. A system of mirrors is used to combine the three coloured beams into a single beam, so that the final image is also coloured like the original object.

#### Synchronization in TV

The scanning operations at the transmitter and the receiver must be synchronised or kept in step to reproduce at the receiver the image of the object being televised. This is accomplished by synchronizing signals which are produced at the transmitting station and radiated along with the picture information. These synchronising signals have the form of rectangular pulses and are known as *sync pulses*. They are produced periodically at the line and field scanning frequencies.

For synchronization of horizontal line scanning, horizontal sync pulse is transmitted for each horizontal line. There are 625 horizontal lines in each frame (in Indian system). Each frame is scanned in  $1/25$  sec.

Total number of horizontal lines scanned per second is  $625 \times 25 = 15625$ .

Thus the horizontal synchronizing frequency,  $f_h = 15625$  Hz.

Similarly a vertical synchronizing pulse is transmitted for each field to keep the vertical scanning in step. The *vertical synchronizing frequency* is

$$f_v = 2 \times \text{picture repetition rate} = 2 \times 25 = 50 \text{ Hz.}$$

Thus the vertical synchronizing frequency is equal to the field scanning frequency i.e., 50 Hz.

To avoid interference with the picture on the TV screen, the synchronizing pulses are transmitted during the blanking periods i.e., when no scene is being pictured or scanned. Actually, the scanning spot returns quickly to start the next line at the end of each horizontal line. Similarly, the spot flies back to the top of the frame after reaching its bottom. During these retrace or return intervals, the picture tube is made inoperative by means of *blanking pulses*. These *blanking pulses* are negative pulses of large amplitude, produced at the transmitting station and are added to the video signal. These pulses are imposed on the transmitting signals and modulate the transmitter carrier.

At the receiving station, after the detector stage, the blanking pulses are separated from the composite video signal and used to blank the output of the picture tube.

### Television Channel Width

The band of frequencies required to transmit a TV signal is called a “*TV channel*”. Television broadcasting stations in India are assigned channel 7 MHz wide, arranged as shown in Fig. 40.20. This includes both picture and sound. VHF and UHF bands are used for transmission of different TV programmes, each having a channel width of 7 MHz. From the transmitter to the receiver, TV signals travel by means of tropospheric or space waves.

In the entire band width of 7 MHz, the video carrier (picture carrier) is placed at 1.25 MHz. The audio carrier (sound carrier) is placed at 6.75 MHz with a maximum deviation of  $\pm 50$  kHz. The complete lower side band for the picture signal is not transmitted; this is referred to as *vestigial side band transmission*. The width of vestigial side band is 0.75 MHz. The width of the main side band for video is 5 MHz. The upper side band of the amplitude modulated video carrier extends at full amplitude from 1.25 to 6.25 MHz, a range of 5 MHz, and at reduced amplitude upto 6.75 MHz. The sound accompanying the TV picture signal is obtained by a separate FM transmitter having a maximum frequency deviation of 50 kHz and a carrier frequency 5.5 MHz greater than the video carrier. This frequency difference (5.5 MHz) is called the *inter carrier frequency* and is used in extracting the sound information from the video detector.

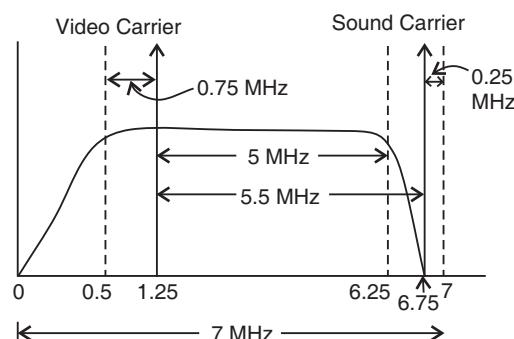


Fig. 40.20

### 40.13 BLOCK DIAGRAM OF TV TRANSMITTING SYSTEM

Fig. 40.21 shows the block diagram of a TV transmitter.

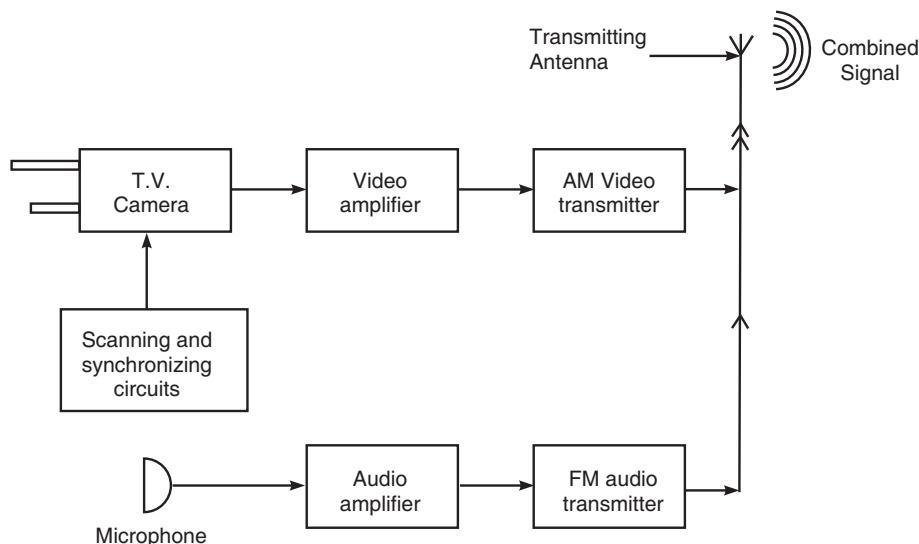


Fig. 40.21

- The picture (video) and sound (audio) signals are transmitted at different frequencies (to avoid interference) by a single antenna. These signals are processed separately at the TV station.

### Working of the major functional blocks

**TV camera.** The video section consists of a TV camera tube which produces a video signal of the picture to be televised.

The synchronising and scanning circuits provide the necessary sweep voltages to the vertical and horizontal deflector plates. Moreover, these synchronising pulses are needed for the reassembly of the image at the receiver in step with the scanning at the transmitter. So these signals are transmitted along with the video signals.

**Video amplifier.** The video signals are amplified in a video amplifier.

**AM video transmitter.** The video signals are then used to amplitude modulate (AM) a radio frequency (RF) carrier in an AM video transmitter.

**Antenna.** The modulated wave is fed to the transmitting antenna. Both video and audio signals are transmitted into space, by the same antenna.

**Microphone.** In the audio section, the sound of the scene is picked up by a microphone. The microphone converts the audio signals into electrical signals.

**Audio amplifier.** The electrical signals are amplified in the audio amplifier.

**FM audio transmitter.** Then the audio signals are used to frequency modulate a RF carrier whose frequency is 5.5 MHz greater than the video carrier.

This frequency modulated (FM) sound (or audio) signal is also carried to the same transmitting antenna. The combined (AM video and FM audio) signal is radiated by the antenna in the form of electromagnetic waves.

### 40.14 BLOCK DIAGRAM OF TV RECEIVING SYSTEM

Fig. 40.22 shows a block diagram of TV receiver.

Receiving Antenna

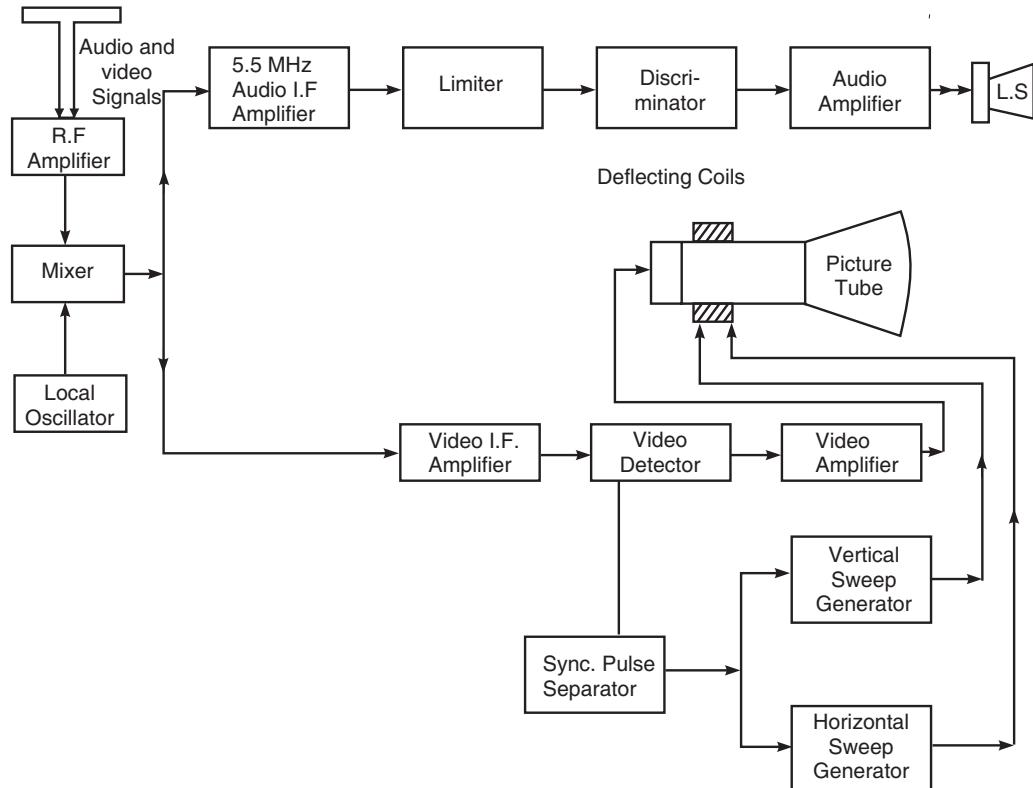


Fig. 40.22

**Principle.** It uses the superheterodyne principle.

**Working of different stages**

**Receiving antenna.** The incoming signal, consisting of audio and video RF signals, is picked up by the receiving antenna.

**RF amplifier.** The selected signals are amplified by the RF amplifier.

**Mixer.** The output of the RF amplifier is fed to the frequency mixer. The mixer heterodynes the amplified RF signal with the output of the local oscillator. It produces two intermediate frequencies, one for the picture and the other for the sound signal. Usually picture IF is 38.9 MHz and sound IF is 33.4 MHz. These frequencies are separated by the help of tuned circuits in the output of mixer stage. The output of the mixer is fed to two sets of tuned IF amplifier sections—one tuned to sound IF and the other to picture IF.

**5.5 MHz audio I.F. amplifier.** The sound IF signal thus obtained is applied to separate sound portion of the receiver. The amplified audio signal from IF amplifier is delivered to the *limiter stage*.

**Limiter.** The limiter removes any interference riding on the top of the signals in the form of amplitude modulation.

**Discriminator.** The sound signal then passes through a discriminator (or a FM detector). The FM detector extracts audio signal from FM carrier.

**Audio amplifier.** The audio signal is fed to an audio amplifier which amplifies the signal.

**Loud speaker.** The amplified signal is given to the loudspeaker which produces the original sound.

**Video I.F. amplifier.** It amplifies the video I.F. signals and feeds on to the video detector.

**Video detector and video amplifier.** After detection and amplification, the video signals obtained are identical to those produced by the orthicon and the pulse generator at the transmitter.

**Picture tube.** The output from the video amplifier is applied to the *control grid of the picture tube or kinescope*. When electrons strike the fluorescent screen of the kinescope, high energy luminous radiations come out. The brightness of these radiations varies with the variation of the amplitude of the picture signal and hence with the brightness of the transmitted image. Dark and bright spots thus appear on the screen in step with the dark and bright spots of the picture televised. These spots merge to give the exact replica of the picture televised. Thus the picture is reproduced.

### CATHODE RAY OSCILLOSCOPE

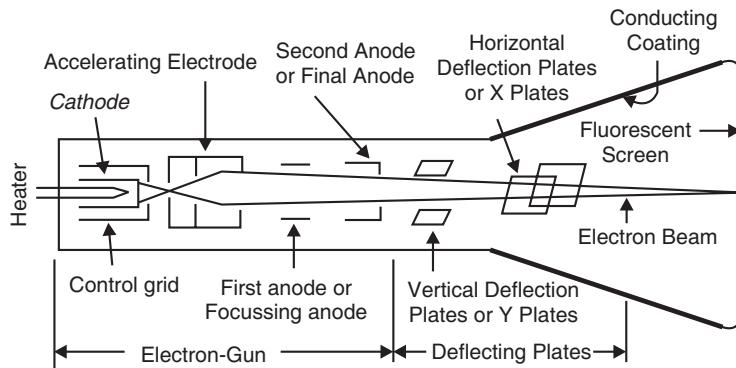
## 40.15 CONSTITUENTS OF CATHODE RAY OSCILLOSCOPE

Cathode ray oscilloscope is an electronic device that produces visual representation of a signal wave form. The main constituents of a cathode ray oscilloscope are

- (1) Cathode ray tube
- (2) Power supply system
- (3) Time base circuit
- (4) Deflection system

The essential parts of a cathode ray tube are shown in Fig. 40.23. It consists of a highly evacuated glass tube. It has three basic parts:

- (i) *Electron gun*, which is an arrangement for producing and focussing an electron beam.
- (ii) *Deflecting plate*, which is an arrangement for deflecting the beam either electrostatically or magnetically.
- (iii) *Fluorescent screen*, upon which the electron beam is focussed to create a well-defined spot.



40.23

Electron gun consists of the following parts:

- (i) *Thermionic cathode* for emission of electrons.
- (ii) A *control grid* for varying the electron current density.
- (iii) An *accelerating electrode*
- (iv) A *focusing system*.

The electrons are emitted from the cathode by thermionic emission. They are controlled by a control grid. Adjacent to the control grid, is the accelerating electrode, which is maintained at a high positive potential and it converges the beam in addition to accelerating the electrons. Due to the mutual repulsion between electrons, the electron beam has a tendency to spread. This is avoided by using the focusing system.

**Deflection system.** The electrostatic deflection system consists of one pair of plane parallel plates  $XX$  which are kept vertical while the other pair of plates  $YY$  are placed horizontal. The deflection voltage applied between  $XX$  plates and  $YY$  plates produce deflection along  $X$  and  $Y$  axis respectively.

Thus, the voltage applied simultaneously to  $X$  and  $Y$  plates, gives control of the beam and the spot of light moves in both, i.e.,  $X$  and  $Y$  directions.

**Fluorescent screen.** The inner face at the flared end of the funnel shaped tube in the screen is coated with some fluorescent material such as zinc orthosilicate, zinc oxide etc. When high velocity electron beam strikes the screen, a spot of light is produced at the point of impact. The colour of the spot depends on the nature of the fluorescent material.

The conical surface of the glass envelope is coated with a conducting coating of carbon particles, called aquadog.

**Action of cathode ray tube.** On heating the cathode, plenty of electrons are emitted from it. These electrons pass through the control grid which influences the amount of current flow. The intensity of spot on the screen can be varied by varying the negative potential on the control grid. As the electron beam leaves the control grid, it comes under the influence of accelerating and focusing electrodes. These electrodes produce a field which acts as an electrostatic lens to converge the electron beam to a spot on the screen.

The focused beam, as it emerges from the focusing and the accelerating system comes under the influence of vertical and horizontal deflection plates. With no voltage on the deflecting plates, the electron beam travels straight and will produce a spot of light at the centre of the screen. The spot can be moved on the screen either horizontally or vertically by applying voltages to the concerned set of plates.

### 40.16 DEFLECTION SENSITIVITY OF A CRO OF ELECTROSTATIC TYPE

It is defined as the amount of deflection (in mm) of the electron spot produced on the screen when a voltage of one volt d.c. is applied between the deflection plates. Its unit is mm/volt. It is different for  $X$  and  $Y$  plates. The geometry of the electrostatic deflection system is shown in Fig. 40.24. Let

Length of two similar plates	=	$l$
Spacing between the plates	=	$S$
Distance of the plates from the screen	=	$D$
Potential applied between two plates	=	$V_d$
Accelerating potential applied to final anode	=	$V_a$
Charge on electron	=	$e$
Mass of electron	=	$m$

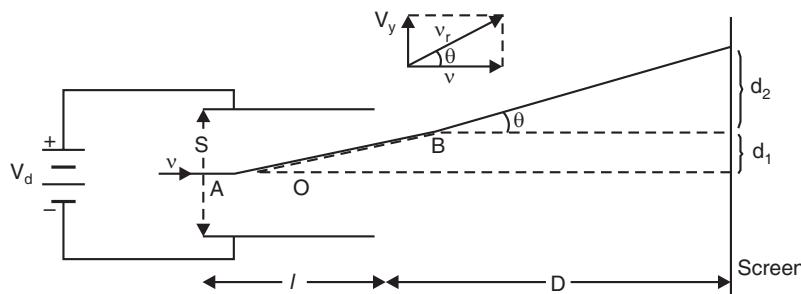


Fig. 40.24

Velocity  $v$  of an electron when entering the field of deflection plates is given by

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

or

$$v^2 = \frac{2eV_a}{m} \quad \dots (1)$$

As the beam passes through the electric field between the deflecting plates, the electrons are attracted towards the positive plate by a force

$$F = eE = \frac{eV_d}{S} \quad \dots (2)$$

This force produces an upward acceleration

$$a = \frac{F}{m} = \frac{eV_d}{mS}. \quad \dots (3)$$

However, the forward motion continues at velocity  $v$ . Time taken by the electron to traverse the electric field of the deflection plates,

$$t = \frac{l}{v} \quad \dots (4)$$

Let  $v_y$  be the upward velocity attained in this time interval  $\frac{l}{v}$ . Then

$$v_y = \frac{l}{v} \cdot a = \frac{l}{v} \cdot \frac{eV_d}{mS} \quad \dots (5)$$

The electron follows a curved path during its travel in the field (i.e., from *A* to *B*).

The vertical displacement during this period is given by

$$d_1 = \frac{1}{2} a \left( \frac{l}{v} \right)^2 = \frac{1}{2} \frac{eV_d}{mS} \cdot \frac{l^2}{v^2} \quad \dots (6)$$

Let the electron beam make an angle  $\theta$  with the axis, after emerging out from the field of deflection plates. Then

$$\tan \theta = \frac{v_y}{v} = \frac{d_2}{D}$$

$$\text{Hence, } d_2 = D \frac{v_y}{v} = \frac{l}{v^2} \cdot \frac{eV_d}{mS} \cdot D \quad \dots (7)$$

$$\begin{aligned} \text{Total deflection } d &= d_1 + d_2 = \frac{1}{2} \frac{eV_d}{mS} \cdot \frac{l^2}{v^2} + \frac{l}{v^2} \cdot \frac{eV_d}{mS} \cdot D \\ &= \frac{eV_d}{mS} \cdot \frac{l}{v^2} \left( \frac{l}{2} + D \right) = \frac{eV_d}{mS} \cdot \frac{l}{v^2} \cdot L \quad \left[ \text{where } L = \frac{l}{2} + D \right] \end{aligned}$$

Substituting for  $v^2$  from Eq. (1),

$$d = \frac{eV_d}{mS} \cdot l \cdot \frac{m}{2eV_a} \cdot L = \frac{IL}{2S} \cdot \frac{V_d}{V_a} \quad \dots (8)$$

$$\text{Deflection sensitivity} = \frac{d}{V_d} = \frac{IL}{2SV_a} \quad \dots (9)$$

Thus, the deflection sensitivity is

- (i) directly proportional to  $l$ , the length of the deflection plates.
- (ii) directly proportional to  $L$ , the distance between the screen and the centre of the deflection plates.
- (iii) Inversely proportional to  $S$ , the spacing between deflection plates.
- (iv) Inversely proportional to  $V_a$ , the final anode voltage.

#### 40.17 TIME BASE

A time base is a circuit which generates a *saw-tooth waveform*. It causes the spot to move in a horizontal or vertical direction *linearly with time*. The voltage so generated is often called “saw-tooth voltage” (Fig. 40.25). The voltage during each cycle increases linearly with time upto  $t_i$  and then falls to zero in a shorter time  $t_f$  called the flyback or *retrace time*.

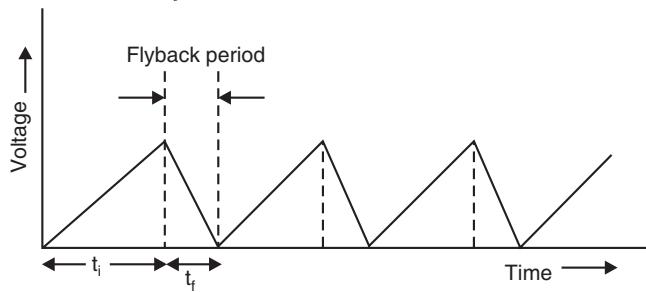


Fig. 40.25

**Necessity of time-base voltage.** If a d.c. voltage is applied across X-plates, the spot of light will move in a horizontal line on the screen. Similarly, if an a.c. voltage alone is applied to Y-deflection plates, the spot of light will move on the screen vertically up and down in a straight line. This line, however, does not reveal the nature of the applied voltage waveform. To get the actual waveform, it is necessary that it appears as plotted against time as an axis. Hence a time base voltage is applied to the X-plates of C.R.O. Under this voltage, the spot sweeps linearly across the screen from left to right and then flies back quickly to the starting position for the next sweep. This horizontal sweep, however, appears as a stationary line on the screen due to persistence of vision. When vertical motion of the spot produced by the a.c. voltage applied on Y-plates is superimposed on the horizontal sweep produced by X-plates, the actual waveform of the a.c. voltage is traced on the screen. The circuit connections for such an arrangement are shown in Fig. 40.26. An exact and stationary trace of the applied a.c. voltage waveform is obtained on the screen when the frequency of time base equals the frequency of the applied sinusoidal voltage. The process of matching the frequencies of the time base voltage and the applied voltage is called *synchronization*.

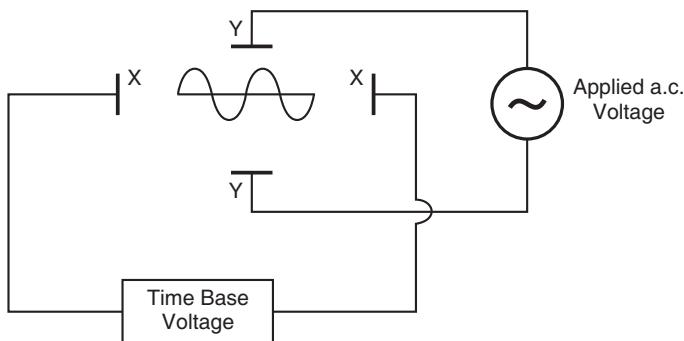


Fig. 40.26

## 40.18 APPLICATIONS OF THE CRO

(1) **Measurement of voltage.** The *deflection sensitivity* of a C.R.O is defined as the displacement of the spot on the screen for a potential of one volt applied to the deflecting plates. This deflection sensitivity is usually expressed as the ratio of the input voltage to the length of the trace on the screen. If the deflection on the screen is multiplied by the deflection sensitivity we get the magnitude of the applied voltage. This is in case of a d.c.

To measure the alternating voltage of sinusoidal wave form, it is applied across the Y plates. A straight line trace will be obtained on the screen. The length of this trace is measured. If this length is multiplied by the deflection sensitivity we get the peak to peak value of the ac voltage. Half this value gives the peak value of the dc. Dividing the peak value by  $\sqrt{2}$  we get the rms value of the applied ac voltage.

Thus, *C.R.O acts as an ideal voltmeter* due to its high input impedance.

(2) **Measurement of direct or alternating current.** The current to be measured is passed through a known resistance. Then the voltage developed across the resistance is applied across the vertical deflecting plates, the horizontal plates being kept short-circuited. Voltage is measured from the deflection on the screen. The corresponding current is calculated by the relation  $I = \frac{V}{R}$ .

(3) **Study of waveforms.** The ac voltage whose waveform is to be studied is applied to the Y plates. A linear time base is applied to the X plates. The frequency of the time base is then adjusted to some exact submultiple of the frequency of the applied a.c. voltage till a steady pattern of the actual

waveform is obtained on the screen. Then the ratio of frequency of a.c. voltage to the frequency of time base voltage will be equal to the number of complete a.c. voltage waves.

(4) **Measurement of frequency.** When two sinusoidal voltages are applied simultaneously to the two sets of deflecting plates, a Lissajous pattern is obtained on the screen.

Let  $v_x$  and  $v_y$  be the two instantaneous voltages applied to the X and Y plates.

$$v_x = V_x \sin \omega_x t \quad \dots (1)$$

$$\text{and} \quad v_y = V_y \sin (\omega_y t + \phi) \quad \dots (2)$$

Here,  $V_x$  and  $V_y$  are the voltage amplitudes,  $\omega_x$  and  $\omega_y$  are the angular frequencies of the voltages and  $\phi$  is the phase angle for voltage  $v_y$ .

By adjusting the values of  $\omega_x$ ,  $\omega_y$ ,  $V_x$ ,  $V_y$  and  $\phi$ , various Lissajous figure patterns are obtained on the screen.

**Case 1.** When  $\omega_x = \omega_y$  and  $\phi = 0$ , i.e., the signals have the same frequency and are in the same phase. From Eqs. (1) and (2),

$$v_y = \frac{V_y}{V_x} \cdot v_x \quad \dots (3)$$

This is the equation of a straight line passing through the origin and making an angle of  $\tan^{-1} \frac{V_y}{V_x}$ , with the horizontal [Fig. 40.27a].

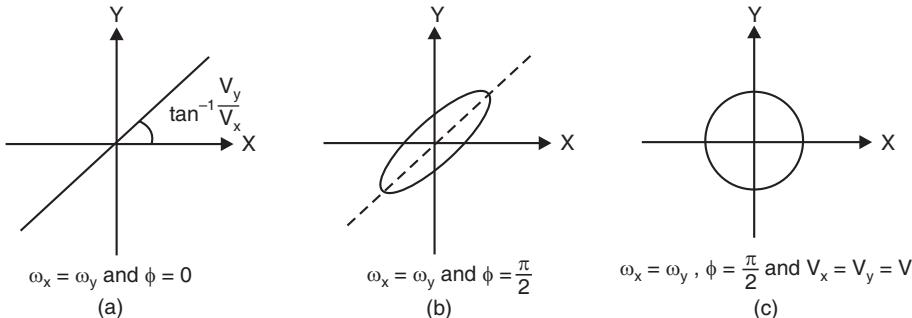


Fig. 40.27

**Case 2.** Let  $\omega_x = \omega_y$  and  $\phi = \frac{\pi}{2}$ . Then,

$$\frac{v_x^2}{V_x^2} + \frac{v_y^2}{V_y^2} = 1 \quad \dots (4)$$

This is the equation of an ellipse [Fig. 40.27b]

**Case 3.** Let  $\omega_x = \omega_y$ ,  $\phi = \frac{\pi}{2}$  and  $V_x = V_y = V$ . Then,

$$v_x^2 + v_y^2 = V^2 \quad \dots (5)$$

This is the equation of a circle [Fig. 40.27(c)].

To measure the unknown frequency of a sinusoidal ac voltage, it is applied to the Y-plates. To the X-plates is applied the sinusoidal voltage obtained from a standard variable frequency oscillator. The frequency of standard oscillator is now varied until a single loop stationary pattern is obtained.

The frequency of the signal is then the same as that of the oscillator, which may be noted from the calibrated dial of the oscillator.

(5) **Measurement of phase difference.** Lissajous figures can be used to measure the relative phase angle of two alternating voltages of equal amplitude and same frequency. They are applied to the  $X$  and  $Y$ -plates of a CRO. Usually an ellipse is obtained on the screen, the orientation of which with respect to coordinate axes depends upon the phase difference between the two voltages. The maximum vertical deflection  $Y$  and the intercept  $y$  on the  $Y$ -axis are measured (Fig. 40.28). Then if  $\phi$  is the phase difference between the two voltages, it can be shown that

$$\sin \phi = \frac{y}{Y} \quad \dots (1)$$

The phase difference is also given by

$$\sin \phi = \frac{x}{X} \quad \dots (2)$$

**EXAMPLE 1.** In a cathode ray oscilloscope, the spacing of the plates is 0.5 cm, the length of the deflection plates is 2.0 cm and the distance of the screen from the center of the plates is 12 cm. Find the value of the displacement produced by a deflecting voltage of 30 volts. Given that the final anode voltage = 2000 volts.

**SOL.** The electrostatic deflection sensitivity of a CRO =  $\frac{d}{V_d} = \frac{l}{2S} \frac{L}{V_a}$

The displacement of spot on the screen is

$$d = \frac{l}{2S} \frac{L}{V_a} V_d = \frac{2 \times 12 \times 30}{2 \times 0.5 \times 2000} \text{ cm} = 0.36 \text{ cm}$$

**EXAMPLE 2.** The writing speed of a cathode-ray oscilloscope is the speed with which an electron beam can trace a line on the screen. A certain manufacturer claims that the writing speed of his oscilloscope is  $6 \times 10^8$  m/s. Can his claim be true? Explain. (Bangalore, April 92)

**SOL.** According to special theory of relativity, no particle can move with the velocity of light. But it is given that the writing speed of the oscilloscope is  $6 \times 10^8$  m/s. His claim is incorrect because (1) no particle is moving on the screen (2) it is only the phase velocity, not the group velocity.

**Block diagram of CRO.** The block diagram of C.R.O is shown in Fig. 40.29. The connections to the  $X$  plates and  $Y$  plates are made through the amplifiers to obtain the full screen deflection.

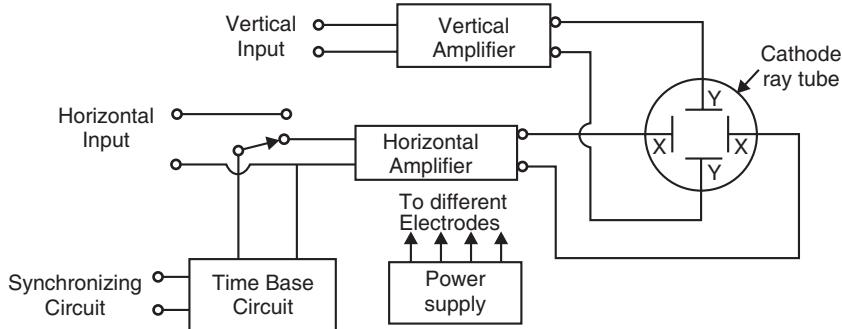


Fig. 40.29

The signal which is applied to vertical input terminal determines the vertical displacement of the beam. The horizontal displacement of the beam is obtained either by an external signal applied to the

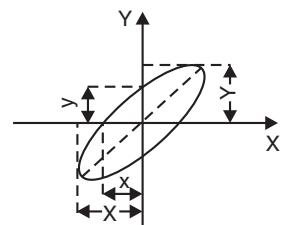


Fig. 40.28

horizontal input or by time base circuit which is internal to C.R.O. The time base circuit is connected to an external synchronizing circuit. The power supply is used to supply the required voltages to various blocks.

### NETWORK THEOREMS

#### 40.19 THEVENIN'S THEOREM

**Statement.** The current in a load impedance connected between two terminals of a network of generators and linear impedances is same as due to a single voltage generator of e.m.f. equal to open circuit voltage between the same terminals of the network and internal impedance equal to the impedance of the network between the same terminals of the network when all the generators in the network have been replaced by their internal impedances.

The schematic representation of Thevenin's theorem is shown in Fig. 40.30.

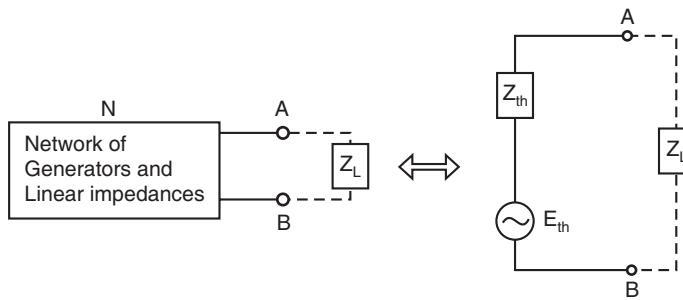


Fig. 40.30

$N$  is a network containing a number of generators and linear impedances with output terminals  $A$  and  $B$ . Let  $E_{th}$  be the open-circuit voltage across  $A$  and  $B$ . Let  $Z_{th}$  be the impedance measured between the terminals  $A$  and  $B$  when all the generators have been replaced by their respective internal impedances. The network  $N$  will produce the same current in an external load impedance  $Z_L$  connected across  $A$  and  $B$  as a single voltage generator of e.m.f.  $E_{th}$  and internal impedance  $Z_{th}$  would do.

**Proof.** Consider a network containing resistive impedances  $Z_1$ ,  $Z_2$  and  $Z_3$  and one generator of e.m.f.  $E$  and internal impedance zero (i.e., ideal generator) [Fig. 40.31 (a)].  $I_1$  is current supplied by the source.  $I_2$  is the current flowing through the load impedance  $Z_L$ .

Applying Kitchoff's II law to meshes I and II, we get

$$I_1 Z_1 + (I_1 - I_2) Z_3 = E \quad \dots (1)$$

$$\text{or} \quad I_1 (Z_1 + Z_3) - I_2 Z_3 = E \quad \dots (1)$$

$$\text{and} \quad I_2 Z_2 + I_2 Z_L - (I_1 - I_2) Z_3 = 0$$

$$I_2 (Z_2 + Z_L + Z_3) = I_1 Z_3 \quad \dots (2)$$

$$\text{or} \quad I_1 = \frac{I_2 (Z_2 + Z_3 + Z_L)}{Z_3}$$

Substituting this value of  $I_1$  in Eqn. (1), we get

$$I_2 \frac{(Z_2 + Z_3 + Z_L)}{Z_3} (Z_1 + Z_3) - I_2 Z_3 = E$$

$$\text{or} \quad I_2 [(Z_2 + Z_3 + Z_L) (Z_1 + Z_3) - Z_3^2] = EZ_3$$

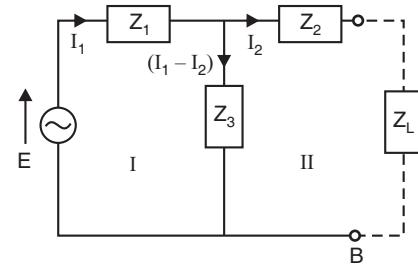


Fig. 40.31 (a)

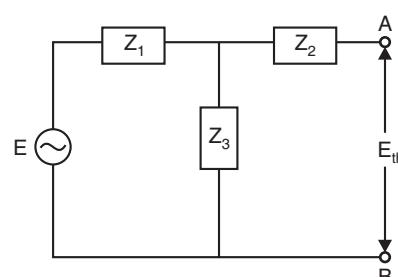


Fig. 40.31 (b)

or

$$I_2 = \frac{EZ_3}{Z_2(Z_1 + Z_3) + Z_1Z_3 + Z_L(Z_1 + Z_3)}$$

∴

$$I_2 = \frac{EZ_3 / (Z_1 + Z_3)}{Z_2 + \left( \frac{Z_1Z_3}{Z_1 + Z_3} \right) + Z_L} \quad \dots (3)$$

Again, from Fig. 40.31 (b), the open-circuit voltage at terminals A, B (i.e., when  $Z_L$  is removed) is

$$E_{th} = \frac{EZ_3}{Z_1 + Z_3} \quad \dots (4)$$

Further if E is replaced by its zero internal impedance [Fig. 40.31(c)], the impedance of network between A and B is

$$\begin{aligned} Z_{th} &= Z_2 + \frac{1}{\frac{1}{Z_1} + \frac{1}{Z_3}} \\ &= Z_2 + \frac{Z_1Z_3}{Z_1 + Z_3} \quad \dots (5) \end{aligned}$$

Thus Eq. (3) is equivalently written as

$$I_2 = \frac{E_{th}}{Z_{th} + Z_L} \quad \dots (6)$$

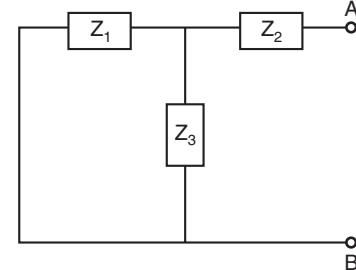


Fig. 40.31 (c)

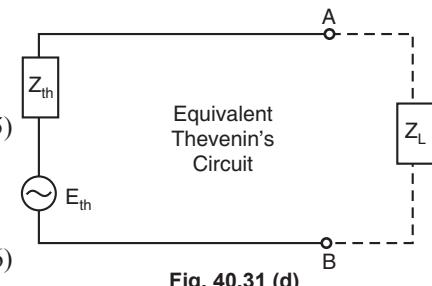
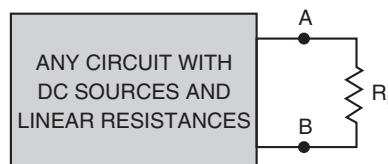


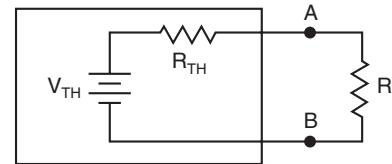
Fig. 40.31 (d)

This proves Thevenin's theorem. The equivalent circuit is shown in Fig. 40.31 (d).

**Explanation.** Consider a network containing dc sources and linear resistances [Fig. 40.32(a)].



(a) Black box has a linear circuit inside of it;



(b) Thevenin circuit

Fig. 40.32

- To zero a voltage source, replace it by a short.
- To zero a current source, replace it by an open.

● In Fig. 40.32 (a) the **Thevenin voltage**  $V_{TH}$  is defined as the voltage across the load terminals when the load resistor is open. Because of this, the Thevenin voltage is sometimes called the *open-circuit voltage*.

**Thevenin voltage:**  $V_{TH} = V_{OC}$  ... (1)

● The **Thevenin resistance** is defined as the resistance that an ohmmeter measures across the load terminals of Fig. 40.32 (a), when all sources are reduced to zero and the load resistor is open.

**Thevenin resistance:**  $R_{TH} = R_{OC}$  ... (2)

● Thevenin's theorem can be applied to ac circuits containing inductors, capacitors, and resistors.

**EXAMPLE 1.** Use Thevenin's theorem to find the current in load resistance of  $100\ \Omega$  in the given circuit [Fig. 40.33].

**SOL.** (i) First of all we disconnect load resistance of  $100\ \Omega$  [Fig. 40.34]

(ii) To determine Thevenin's e.m.f.  $E_{th}$ . If  $I$  is the current in  $10\ \Omega$  resistor, then current in  $5\ \Omega$  resistor is also  $I$  since terminals  $A$  and  $B$  are open circuited.

According to Kirchoff's II law, the equation for closed loop is

$$10I + 5I = 10$$

$$\therefore I = \frac{10}{15} = \frac{2}{3}\ A$$

Voltage across  $AB$  = Voltage across  $5\ \Omega$  resistor  
 $= (2/3) \times 5 = (10/3)\ V$

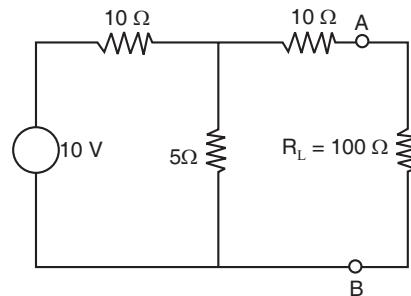


Fig. 40.33

(iii) To determine Thevenin's Resistance  $R_{th}$ . Now the source of e.m.f. is replaced by its internal resistance which is zero in this case [Fig. 40.34 (b)].

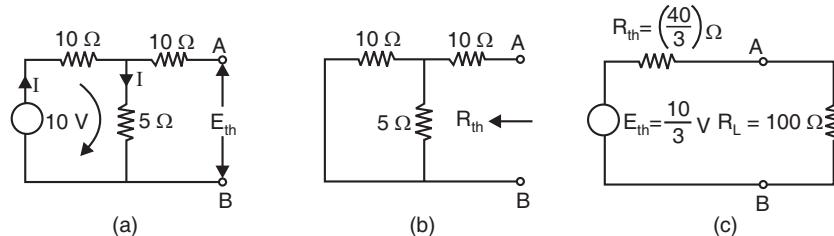


Fig. 40.34

As viewed across  $A$ ,  $B$ ,  $10\ \Omega$  and  $5\ \Omega$  resistances are in parallel and then  $10\ \Omega$  resistor is in series.

$$R_{th} = 10 + \frac{10 \times 5}{10 + 5} = 10 + \frac{50}{15} = 10 + \frac{10}{3} = \frac{40}{3}\ \Omega$$

The Thevenin's circuit is shown in Fig. 40.34 (c).

$\therefore$  The current through load  $R_L$  is

$$I_L = \frac{\text{Voltage}}{\text{Resistance}} = \frac{E_{th}}{R_{th} + R_L} = \frac{\left(\frac{10}{3}\right)}{\left(\frac{40}{3} + 100\right)} = \frac{1}{34}\ A$$

## 40.20 NORTON'S THEOREM

**Statement.** Any two terminal active network, consisting of linear impedances and generators may be replaced by a current generator in parallel with an impedance. The current is equal to that obtained through a short circuit across the two terminals and the impedance is that viewed across the two terminals when all the generators are replaced by their internal impedances.

The schematic representation of Norton's theorem is shown in Fig. 40.35.

This theorem is useful in reducing a complicated network into a simple parallel circuit consisting of an ideal current source and a parallel impedance. Hence the theorem finds its utility where current sources are of major interest than e.m.f.'s e.g., in transistor circuits.

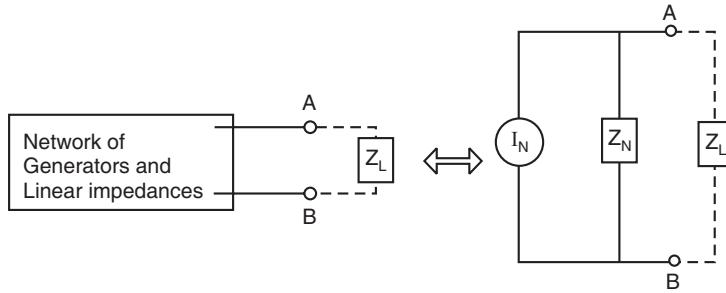


Fig. 40.35

**Proof.** Consider the Thevenin's and Norton's equivalent circuits of same network shown in Fig. 40.36 (a) and (b).

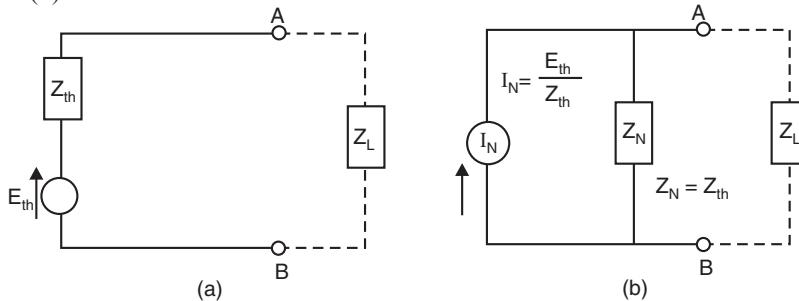


Fig. 40.36

The current  $I_2$  for Thevenin's equivalent circuit in load  $Z_L$  is

$$I_2 = \frac{E_{th}}{Z_{th} + Z_L} \quad \dots (1)$$

The current  $I_L$  for Norton's equivalent circuit in load  $Z_L$  is

$$I_L = \frac{I_N Z_N}{Z_N + Z_L}$$

But  $Z_N = Z_{th}$  and  $I_N = \frac{E_{th}}{Z_{th}}$

$$I_L = \frac{E_{th}}{Z_{th} + Z_L} \quad \dots (2)$$

∴

$$I_2 = I_L$$

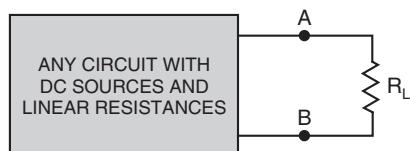
The current across the load impedance  $Z_L$  given by two equivalent circuits are equal.

Thus Norton's theorem is proved.

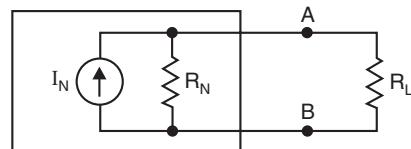
**Explanation.** In Fig. 40.37 (a), the Norton current  $I_N$  is defined as the load current when the load resistor is shorted. Because of this, the **Norton current** is called the *short-circuit current*.

**Norton current:**

$$I_N = I_{SC} \quad \dots (1)$$



(a) Black box has a linear circuit inside of it;



(b) Norton circuit

Fig. 40.37

The **Norton resistance** is the resistance that an ohmmeter measures across the load terminals when all sources are reduced to zero and the load resistor is open.

$$\text{Norton resistance: } R_N = R_{OC} \quad \dots (2)$$

- For ac circuits, the Norton current  $I_N$  is expressed as a complex number in polar form.

The Norton impedance  $Z_N$  is expressed as a complex number in rectangular form.

**EXAMPLE 1.** Draw Thevenin's and Norton's equivalent circuits for following network of resistances [Fig. 40.38]. Calculate the current in the load in each case.

**SOL.** (i) We first remove  $R_L$  and leave  $A$  and  $B$  in open circuit.

Applying Kirchoff's second law to mesh of circuit [Fig. 40.39(a)]

$$0.6I + 0.8I + 0.2I = 24 \text{ or } I = 15 \text{ A}$$

$$E_{th} = \text{voltage across } CD = 15 \times 0.8 = 12 \text{ V}$$

(ii) Thevenin's internal impedance  $Z_{th}$ . The battery is removed and its internal resistance 0.2  $\Omega$  is left behind [Fig. 40.39(b)]. The internal impedance as viewed from  $A$  and  $B$  is the Thevenin's impedance given by

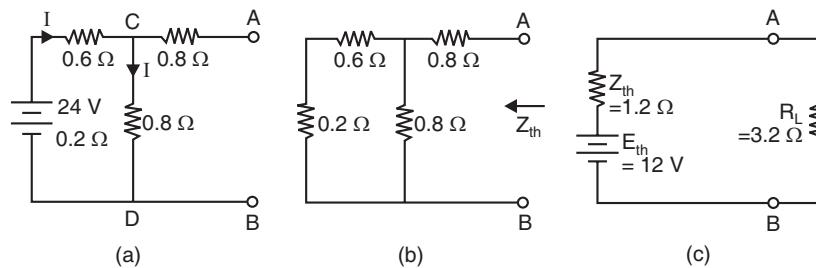


Fig. 40.39

$$Z_{th} = 0.8 + \frac{(0.8)(0.6+0.2)}{0.8+(0.6+0.2)} = 0.8 + 0.4 = 1.2 \Omega$$

The Thevenin's equivalent circuit is shown in Fig. 40.39 (c).

(iii) The current in load  $R_L$  when connected between  $AB$  is

$$I_L = \frac{12 \text{ Volt}}{1.2 \Omega + 3.2 \Omega} = \frac{12 \text{ Volt}}{4.4 \Omega} = \frac{30}{11} \text{ A}$$

**Norton's Equivalent Circuit.** The Norton's equivalent circuit may be found from Thevenin's equivalent circuit. The short-circuit current is found by shorting the terminal  $A$  and  $B$  together.

$$I_N = \frac{E_{th}}{Z_{th}} = \frac{12}{1.2} = 10 \text{ A}$$

The Norton's equivalent circuit is shown in Fig. 40.39 (d). When the load  $R_L = 3.2 \Omega$  is connected across  $AB$ , the current in the load is

$$I_L = \frac{Z_N}{Z_N + R_L} I_N = \frac{1.2}{1.2 + 3.2} \times 10 = \frac{12}{4.4} = \frac{30}{11} \text{ A}$$

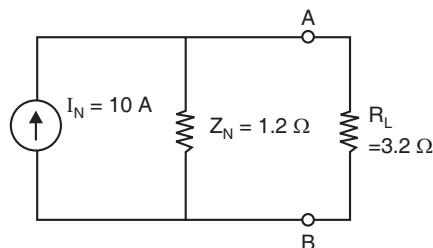


Fig. 40.39(d)

**EXERCISE**

1. Why is modulation employed in the transmission of information? Explain the salient features of amplitude modulation with the help of appropriate wave-forms. Explain how A.M. signals are detected using a diode.  
*(Kakatiya 93)*
2. Obtain an expression for varying voltage of an amplitude modulated wave and the total power of the AM wave. Discuss the waveform and sidebands in it.
3. Give the theory of frequency modulation.
4. Draw the block diagram of AM transmitter. Explain briefly the function of each block.
5. Explain the working of a diode detector in A.M.  
*(S.K.U. 95)*
6. Draw the block diagram of a typical superheterodyne receiver and explain the functional operation of each stage.
7. Draw a block diagram of FM transmitter. Explain briefly the function of each block.
8. Describe the various units in phase lock loop and explain its working. How is PLL used for FM demodulation?
9. Draw the block diagram of a superheterodyne FM receiver and explain the functional operation of each stage.
10. Explain, giving block diagram, the principle of transmission and reception of colour T.V. signal.  
*(Madras 1988)*
11. Describe the operation of the image orthicon tube.  
*(Madras 1991)*
12. What is meant by scanning in TV system? Write a note on interlaced scanning.  
*(Madras 1991)*
13. Draw the block diagram of monochrome TV transmitter. Explain briefly the function of each block.
14. Draw the block diagram of monochrome TV receiver. Explain the function of each block briefly.
15. With a neat diagram describe a cathode ray oscilloscope. Obtain an expression for the deflection sensitivity of the electrostatic type.  
*(Bangalore Nov. 92)*
16. What is a time-base circuit? Explain how a CRO is used to  
(i) determine the frequency and (ii) measure the voltage.
17. State and prove the Thevenin theorem.  
*(Meerut 2002)*
18. State and prove Norton's theorem.  
*(Meerut 2003)*

## NUMBER SYSTEMS AND LOGIC CIRCUITS

**After reading this chapter, you should be able to**

- ◆ Express positive integers in different number systems (binary, octal, decimal, hexadecimal)
- ◆ Codify data elements or information (signal values) by binary variables (signals) using standard codes for positive integers (binary, BCD) and characters (ASCII code)
- ◆ Design and use: multiplexers (or data selectors), demultiplexers (or distributors)
- ◆ Explain and relate the following concepts for designing a logic circuit: truth table, boolean algebra and logic functions
- ◆ Explain the functionality of logic gates (AND, NAND, OR, NOR, XOR, NOT)
- ◆ Explain the working of half adder and full adder.

## NUMBER SYSTEMS

### 41.1 DECIMAL NUMBER SYSTEM

This is the frequently used number system in our daily life. It uses ten numerals 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9. The *base* or the *radix* of the decimal system is 10. A number in decimal system is expressed in terms of the position or place values.

For example, the number 10523 is represented as

$$10523 = 1 \times 10^4 + 0 \times 10^3 + 5 \times 10^2 + 2 \times 10^1 + 3 \times 10^0.$$

The digit (3) has the position value  $10^0$  and is the *least significant digit (LSD)*.

The digit (1) has the place value  $10^4$  and is the *most significant digit (MSD)*.

Similarly, the number 2564.397 can be expressed as

$$2564.397 = 2 \times 10^3 + 5 \times 10^2 + 6 \times 10^1 + 4 \times 10^0 + 3 \times 10^{-1} + 9 \times 10^{-2} + 7 \times 10^{-3}.$$

That is the powers to the base 10 are numbered to the left of the decimal point starting with 0 and to the right of the decimal point starting with -1.

### 41.2 BINARY NUMBER SYSTEM

In a *binary system* of representations, the *base* (or *radix*) is 2. It uses only two numerals 0 and 1. In a digital system there are only two possible states or conditions. For example, a situation may be True or False, a switch close or open, a voltage signal High or Low etc. These states or conditions are designated as 1 and 0 respectively. The binary digits 0 and 1 are termed as bits. Like the decimal system, the binary system also has a place or position value representation.

For example, the number 15 of decimal system is written in the binary system as 1111, since,

$$1111 = 1 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 15.$$

This can also be written as  $1111_2 = 15_{10}$ , the subscript indicating the number system. In the binary 1111, the bit 1 at the extreme left is the MSB (most significant bit) and the bit 1 at the extreme right is the LSB.

Similarly, the binary 1011.011 is written in the decimal system as 11.375, since

$$\begin{aligned} 1011.011 &= 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3} \\ &= 8 + 0 + 2 + 1 + 0 + \frac{1}{4} + \frac{1}{8} = 11.375 \end{aligned}$$

or

$$1011.011_2 = 11.375_{10}$$

Thus, the place values of the bits in a binary number are given by ascending powers of 2 to the left of binary point starting from 0 and to the right of binary point in the descending powers of 2 starting from -1.

#### 41.2.1. Conversion of Binary Number into Decimal Number

##### (i) Conversion of integral binary numbers

**EXAMPLE 1.** Convert  $(1010)_2$  into its decimal equivalent.

**SOL.**  $1010 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 = 8 + 0 + 2 + 0 = 10$

$$\Rightarrow (1010)_2 = (10)_{10}$$

##### (ii) Conversion of fractional binary numbers

As an example, let us find the decimal equivalent of 0.1101.

$$0.1101 = 1 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3} + 1 \times 2^{-4}$$

$$\begin{aligned}
 &= \frac{1}{2} + \frac{1}{4} + 0 + \frac{1}{16} = 0.8125 \\
 \therefore 0.1101_2 &= 0.8125_{10} \\
 \downarrow &\quad \downarrow \\
 \text{Binary} &\quad \text{Decimal} \\
 \text{Point} &\quad \text{Point}
 \end{aligned}$$

**(iii) Mixed Number**

For mixed numbers, that is, having integer and fractional parts, each part is handled separately according to the following rules.

(i) The integer numbers are multiplied by their corresponding powers of base 2, i.e.,  $2^0, 2^1, 2^2, \dots$  from left of the binary point and then added.

(ii) The fractional numbers are multiplied by their corresponding powers of base 2, i.e.,  $2^{-1}, 2^{-2}, 2^{-3}, \dots$  from right of the binary point and then added.

Consider the binary number 1101.101. Its decimal equivalent is

$$\begin{aligned}
 1101.101 &= 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} \\
 &= 8 + 4 + 0 + 1 + \frac{1}{2} + 0 + \frac{1}{8} = 13.625
 \end{aligned}$$

**EXAMPLE 2.** Convert  $(1011.0101)_2$  into its decimal equivalent.

$$\begin{aligned}
 \text{Sol. } 1011.0101 &= 1 \times 2^0 + 1 \times 2^1 + 0 \times 2^2 + 1 \times 2^3 + 0 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3} + 1 \times 2^{-4} \\
 &= 1 + 2 + 0 + 8 + 0 + \frac{1}{4} + 0 + \frac{1}{16} \\
 &= 11.3125
 \end{aligned}$$

$$(1011.0101)_2 = (11.3125)_{10}$$

**EXAMPLE 3.** Convert the binary number  $(10111.1101)_2$  into its decimal equivalent.

**Sol.** A binary number is converted into a decimal number by multiplying each binary digit by its weight and adding the resulting products.

$$\begin{array}{r}
 1 \ 0 \ 1 \ 1 \ 1 \ 1 \\
 | \quad | \quad | \quad | \quad | \quad | \\
 1 \times 2^0 = 1 \\
 1 \times 2^1 = 2 \\
 1 \times 2^2 = 4 \\
 1 \times 2^3 = 8 \\
 0 \times 2^4 = 0 \\
 1 \times 2^5 = \underline{\underline{32}} \\
 \hline
 47
 \end{array}$$

Therefore,  $(10111)_2$  is written as  $(47)_{10}$ .

Conversion of  $(0.1101)_2$  is done in a similar manner.

$$\begin{array}{r}
 0 \cdot 1 \ 1 \ 0 \ 1 \\
 | \quad | \quad | \quad | \quad | \\
 1 \times 2^{-4} = 0 \cdot 0625 \\
 0 \times 2^{-3} = 0 \cdot 0000 \\
 1 \times 2^{-2} = 0 \cdot 2500 \\
 1 \times 2^{-1} = \underline{\underline{0 \cdot 5000}} \\
 \hline
 0 \cdot 8125
 \end{array}$$

Thus,  $(0.1101)_2$  is equal to  $(0.8125)_{10}$ .

Therefore,  $(101111.1101)_2$  is equal to  $(47.8125)_{10}$ .

#### 41.2.2. Conversion of Decimal Number into Binary Number

##### (i) Conversion of integral decimal numbers

The given decimal number is divided progressively by 2, until we get zero. The remainders, taken in the reverse order, give the binary number.

As an example let us convert the decimal 19 into its binary equivalent.

Successive divisions	Remainders
2   19	1
2   9	1
2   4	1
2   2	0
2   1	0
2   0	1

Top  
Bottom

Reading the remainders from the bottom to the top, the binary equivalent of 19 is found to be 10011, or

$$19_{10} = 10011_2.$$

**EXAMPLE 3.** Convert  $(25)_{10}$  into binary number.

Remainders		
2   25	1	
2   12	0	
2   6	1	
2   3	0	
2   1	1	
2   0	1	

$(25)_{10} = (11001)_2$

##### (ii) Conversion of fractional-decimal numbers

The given decimal number is multiplied by 2 progressively.

1. For each step that results in a 1 in the units place, transfer the 1 to the binary record and repeat the process with the fractional number.
2. For each multiplication by 2 that results in a product less than unity, record a 0 in the binary number and carry on the process.
3. The last step is reached if the fractional part is zero or it is terminated when the desired accuracy is attained.
4. The carries are taken in the forward (top to bottom) direction to give the equivalent binary.

**EXAMPLE.** Convert 0.9125 into its binary equivalent.

$$\begin{aligned}
 0.9125 \times 2 &= 1.8250 = 0.8250 \text{ with a carry 1} \\
 0.8250 \times 2 &= 1.6500 = 0.6500 \text{ with a carry 1} \\
 0.6500 \times 2 &= 1.3000 = 0.3000 \text{ with a carry 1} \\
 0.3000 \times 2 &= 0.6000 = 0.6000 \text{ with a carry 0} \\
 0.6000 \times 2 &= 1.2000 = 0.2000 \text{ with a carry 1} \\
 0.2000 \times 2 &= 0.4000 = 0.4000 \text{ with a carry 0}
 \end{aligned}$$

The process is terminated here to get an *approximate* result, namely, representation of 0.9125 by six binary digits.

$$\therefore 0.9125_{10} = 0.111010_2$$

The point in front of the binary is referred to as the *binary point*.

### (iii) Conversion of Mixed Decimal Number

To find the binary equivalent of a decimal number like 35.625, we split the number into an integer of 35 and a fraction of 0.625. Then the binary equivalent of each part is obtained separately by methods discussed above.

First let us find out the binary equivalent of the integer part 35 by divide-by-two method.

35 ÷ 2 = 17 + 1	remainder 1	Top
17 ÷ 2 = 8 + 1	remainder 1	↑
8 ÷ 2 = 4 + 0	remainder 0	
4 ÷ 2 = 2 + 0	remainder 0	
2 ÷ 2 = 1 + 0	remainder 0	
1 ÷ 2 = 0 + 1	remainder 1	
∴ 35 <sub>10</sub> = 100011 <sub>2</sub>		

The binary equivalent of 0.625 is found by the multiply-by-two method,

0.625 × 2	= 1.25	= 1 + 0.25	carry 1	Top
0.25 × 2	= 0.50	= 0 + 0.50	carry 0	↓
0.50 × 2	= 1.0	= 0 + 1	carry 1	Bottom

Taking the carrys from top to bottom

$$\begin{aligned} 0.625_{10} &= 101 \\ \therefore 35.625_{10} &= 100011.101_2 \end{aligned}$$

**EXAMPLE 1.** Convert (21.6)<sub>10</sub> into binary number.

Remainders			
2   21			
2   10	→	1	↑
2   5	→	0	
2   2	→	1	
2   1	→	0	
2   0	→	1	
<hr/>			
Integer			
0 . 6 × 2			
(1) . 2 × 2	→	1	↓
(0) . 4 × 2	→	0	
(0) . 8 × 2	→	0	
(1) . 6	→	1	
<hr/>			
.1001			

$$\therefore (21.6)_{10} = (10101.1001)_2$$

**EXAMPLE 2.** Convert the following decimal numbers to binary:

$$(a) 225.225 \quad (b) 10^4 \quad (c) 1998$$

**Sol.** (a) The integer is 225. The fraction is 0.225.

**Integer conversion:**

Successive Divisions		Remainder
2	225	
2	112	1 ↑
2	56	0
2	28	0
2	14	0
2	7	0
2	3	1
2	1	1
	0	1

**Fraction conversion:**

Binary
$0.225 \times 2 = 0.450 \rightarrow 0$
$0.450 \times 2 = 0.900 \rightarrow 0$
$0.900 \times 2 = 1.800 \rightarrow 1$
$0.800 \times 2 = 1.600 \rightarrow 1$
$0.600 \times 2 = 1.200 \rightarrow 1$
$0.200 \times 2 = 0.400 \rightarrow 0$
$0.400 \times 2 = 0.800 \rightarrow 0$
$0.800 \times 2 = 1.600 \rightarrow 1$
$0.600 \times 2 = 1.200 \rightarrow 1$
$0.200 \times 2 = 0.400 \rightarrow 0$ ↓

Reading the remainder from bottom to top,

$$225_{10} = 11100001_2$$

$$0.225_{10} = 0011100110_2.$$

Combining the above two results,

$$225.225_{10} = 11100001.0011100110_2$$

$$(b) 10000_{10} = 10011100010000_2$$

$$(c) 1998_{10} = 11111001110_2$$

**41.2.3 Binary Addition**

The sum of two binary numbers is calculated by the same rules as in decimal, except that the digits of the sum in any significant position can be only 0 or 1. Any “carry” obtained in a given significant position is used by the pair of digits one significant position higher.

The *four rules* for binary addition are given below:

$$1. \quad 0 + 0 = 0$$

$$2. \quad 0 + 1 = 1$$

$$3. \quad 1 + 0 = 1$$

$$4. \quad 1 + 1 = 10$$

(one-zero, not ten)

The last rule is often written as  $1 + 1 = 0$  with a carry of 1.

Example of addition of two binary numbers is shown below :

$$\begin{array}{r} \text{augend :} & 101101 \\ \text{addend :} & + 100111 \\ \text{sum :} & \hline 1010100 \end{array}$$

Consider another example

$$\begin{array}{r} 1 & 0 & 1 & 1 \\ + 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 & 0 \end{array} \quad \begin{array}{l} (11) \\ (9) \\ (20) \end{array}$$

**Explanation.** Addition is made column-wise.

**41.2.4 Binary Subtraction**

The *four rules* for binary subtraction are given below :

$$1. \quad 0 - 0 = 0$$

$$2. \quad 1 - 0 = 1$$

$$3. \quad 1 - 1 = 0$$

$$4. \quad 10 - 1 = 1.$$

The last rule indicates that when 1 is subtracted from  $10_2$  (= decimal 2), we get 1.

**EXAMPLE 1.** Subtract  $100111$  from  $101101$ .

**SOL.**

minuend :  $101101$

subtrahend :  $-100111$

difference :  $000110$

**EXAMPLE 2.** Subtract  $1001$  from  $1110$ .

1110	First column : (after borrow)	10 - 1 = 1      14
(-) 1001	2nd column : .....	0 - 0 = 0      (-) 9
<hr/>	<hr/>	<hr/>
0101	3rd column : .....	1 - 0 = 1      5
	4th column : .....	1 - 1 = 0

#### 41.2.5 Complement Method of Subtraction

**(a) 1's Complement.** The 1's complement of a binary number is the number which is obtained by changing its each 0 into a 1 and each 1 into a 0. For example, 1's complement of the binary number  $1011$  is  $0100$ .

**Subtraction by 1's complement :**

- (i) The 1's complement of the subtrahend is computed.
- (ii) The 1's complement is added to the minuend.
- (iii) If there is no end-around carry, then the answer is recomplemented and a negative sign is attached to it.
- (iv) If there is an end-around carry of 1, remove the carry 1 from the addition and add it to the remainder. In this case no recomplementing is necessary.

**EXAMPLE 1.** Let us subtract  $1101_2$  from  $1010_2$ .

##### 1's Complement Method

$$\begin{array}{r}
 1010 \\
 + 0010 \\
 \hline
 \square 1100
 \end{array}$$

↑

No carry

##### Conventional Method

$$\begin{array}{r}
 1010 \\
 - 1101 \\
 \hline
 - 0011
 \end{array}$$

Since there is no end-around carry (EAC) in this case, we write 1's complement of  $1101$  and attach a minus sign to it. Thus

Final answer  $\rightarrow -0011$ .

**EXAMPLE 2.** Consider the subtraction

1010	Minuend
- 1001	Subtrahend

1's complement of the subtrahend is  $0110$

$$\begin{array}{r}
 \begin{array}{r} 1010 \\ -1001 \\ \hline \end{array} & \begin{array}{r} 1010 \\ +0110 \\ \hline 10000 \end{array} & \begin{array}{r} 0000 \\ +1 \\ \hline 0001 \end{array} \\
 & \begin{array}{c} \uparrow \\ \text{carry} \end{array} & 
 \end{array}$$

**(b) 2's complement.** The 2's complement of a binary number is obtained by adding 1 to its 1's complement.

## **Subtraction by 2's complement**

- (i) Find the 2's complement of the subtrahend.
  - (ii) Add this complement to the minuend.
  - (iii) Drop the carry in the last position.
  - (iv) If the carry in the last position is 1, the result is positive.
  - (v) If there is no 1 carry in the last position, determine the 2's complement of the result of subtraction and attach a minus sign to it. Thus, the answer of subtraction is negative.

The following examples illustrate the procedure.

- (i) Let us subtract  $10001_2$  from  $10011_2$ .

## 2's Complement Method

$$\begin{array}{r}
 10011 \\
 + 01111 \\
 \hline
 [1] 00010
 \end{array}
 \quad
 \text{(2's complement of 10001)}
 \quad
 \begin{array}{r}
 10011 \\
 - 10001 \\
 \hline
 00010
 \end{array}$$

Drop the carry.

Therefore, final answer is : 00010.

- (ii) Let us subtract  $1101_2$  from  $1010_2$ .

$$\begin{array}{r}
 1010 \\
 -1101 \\
 \hline
 \end{array}$$

$$2\text{'s complement of subtraction} = 1\text{'s complement} + 1 = 0010 + 1 = 0011$$

$$\begin{array}{r}
 1010 \\
 + 0011 \\
 \hline
 \boxed{\phantom{0}} \quad 1101
 \end{array}
 \qquad (2\text{'s complement of } 1101)$$

↑

No carry

Thus, there is no 1 carry in the last position. Hence, we determine 2's complement of 1101 which is

$$0010 + 1 = 0011$$

After attaching the minus sign, the final answer is - 0011.

#### 41.2.6 Binary Multiplication

The following are the rules for binary multiplication.

- (i)  $0 \times 0 = 0$
- (ii)  $0 \times 1 = 0$
- (iii)  $1 \times 0 = 0$
- (iv)  $1 \times 1 = 1$

Binary multiplication is carried out as in decimal system.

**EXAMPLE 1.** Multiply 10110 by 110.

10110 × 110	<b>Verification</b>
00000	$22 \times 6$
10110	132
10110	
10000100	

∴ The result is 10000100.

#### 41.2.7 Binary Division

Binary division is done as in decimal system.

**EXAMPLE 2.** Divide 1111 by 110.

$\begin{array}{r} 10 \\ 110 \overline{)1111} \\ 110 \\ \hline 11 \end{array}$	<p>Verification  <math>15 \div 6</math> gives</p> $\begin{array}{r} 2 \\ 6 \overline{)15} \\ 12 \\ \hline 3 \end{array}$ <p>Quotient : 2      Remainder : 3</p>
$\therefore$ Quotient : 10 Remainder : 11	

### 41.3 OCTAL NUMBER SYSTEM

The *octal number system* has a base of 8. It uses the eight numerals 0, 1, 2, 3, 4, 5, 6, 7.

- For counting beyond 7, 2-digit combinations are formed taking the second digit followed by the first, then the second digit followed by the second and so on.

For example, after 7, the next number in the octal system is 10 (second digit followed by the first), then 11 (second digit followed by the second) and so on. Therefore, different octal numbers are:

$$\begin{array}{cccccccc} 0, & 1, & 2, & 3, & 4, & 5, & 6, & 7, \\ 10, & 11, & 12, & 13, & 14, & 15, & 16, & 17, \\ 20, & 21, & \dots & \dots & \dots & \dots & \dots & \dots \end{array}$$

The following Table gives a short list of a few octal numbers and their decimal equivalents.

Octal system	Decimal system	Octal system	Decimal system
0	0	15	13
1	1	16	14
2	2	17	15
3	3	20	16
4	4	21	17
5	5	22	18
6	6	23	19
7	7	24	20
10	8	25	21
11	9	26	22
12	10	27	23
13	11	30	24
14	12	31	25

- The position value of each digit is in ascending powers of 8 for integers and descending powers of 8 for fractions as shown below:

$$\leftarrow 8^3 \ 8^2 \ 8^1 \ 8^0 \quad \bullet \quad 8^{-1} \ 8^{-2} \ 8^{-3} \rightarrow$$

↑  
octal point

#### (i) Octal to Decimal Conversion.

The conversion from an octal to decimal number is done by multiplying each significant digit of the octal number by its respective weight and adding the products.

For example, 425.350 in the octal system can be expressed in decimal system in the following way.

$$\begin{aligned} 425.350 &= 4 \times 8^2 + 2 \times 8^1 + 5 \times 8^0 + 3 \times 8^{-1} + 5 \times 8^{-2} + 0 \times 8^{-3} \\ &= 4 \times 64 + 2 \times 8 + 5 \times 1 + 3 \times \frac{1}{8} + 5 \times \frac{1}{64} \\ &= 256 + 16 + 5 + 0.375 + 0.078 = 277.453_{10} \end{aligned}$$

**EXAMPLE 1.** Convert the octal numbers (a)  $(237)_8$  and (b)  $(120)_8$  to decimals.

**SOL.** (a)

$$\begin{aligned} (237)_8 &= 2 \times 8^2 + 3 \times 8^1 + 7 \times 8^0 \\ &= 2 \times 64 + 3 \times 8 + 7 \times 1 \\ &= 128 + 24 + 7 \\ &= (159)_{10} \end{aligned}$$

(b)

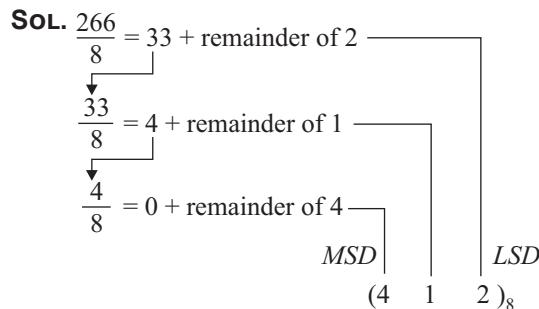
$$\begin{aligned} (120)_8 &= 1 \times 8^2 + 2 \times 8^1 + 0 \times 8^0 \\ &= 1 \times 64 + 2 \times 8 + 0 \times 1 \\ &= 64 + 16 + 0 \\ &= (80)_{10} \end{aligned}$$

#### (ii) Decimal to octal conversion:

**(a) Integers.** The given decimal number is divided progressively by 8 until we get a quotient of zero.

The remainders, taken in the reverse order, give the octal number.

**EXAMPLE.** Convert  $(266)_{10}$  to its octal equivalent.



The first remainder becomes the LSD. The last remainder becomes the MSD.

$$\therefore (266)_{10} = (412)_8$$

- If a calculator is used to perform the divisions in the above process, the result will include a decimal fraction instead of a remainder.

The remainder is obtained by multiplying the decimal fraction by 8.

For example,  $266/8$  produces 33.25. The remainder is  $0.25 \times 8 = 2$ .

**(b) Fractions.** We multiply each digit by 8 and record the carry in integer position. These carries taken in **forward i.e.**, top to bottom, direction give the required octal fraction.

An example of this is to convert decimal 0.23 into an octal fraction.

$$\begin{array}{lll} 0.23 \times 8 = 1.84 = 0.84 & \text{with a carry of } 1 & | \\ 0.84 \times 8 = 6.72 = 0.72 & \text{with a carry of } 6 & \text{Read down} \\ 0.72 \times 8 = 5.76 = 0.76 & \text{with a carry of } 5 & \downarrow \\ & \text{etc.} & \end{array}$$

The carries read downward give the octal fraction 0.165. We terminated after three places.

- If more accuracy is required, we would continue multiplying to obtain more octal digits.

**EXAMPLE 1.** Convert  $(444 \cdot 456)_{10}$  to an octal number.

**SOL.** **Integer conversion:**

Division	Generated remainder
8) 444	
8) 55 →	4 ↑
8) 6 →	7 ↑
8) 0 →	6

Reading the remainders from bottom to top,  
the decimal number  $(444)_{10}$  is equivalent  
to octal  $(674)_8$ .

The octal equivalent of  $(444 \cdot 456)_{10}$  is  $(674.35136)_8$ .

**Fraction conversion:**

Multiplication	Generated integer
$0.456 \times 8 = 3.648 \rightarrow$	3
$0.648 \times 8 = 5.184 \rightarrow$	5
$0.184 \times 8 = 1.472 \rightarrow$	1
$0.472 \times 8 = 3.776 \rightarrow$	3
$0.776 \times 8 = 6.208 \rightarrow$	6

$(0.456)_{10} = (0.35136)$   
Here we have terminated after five place.

#### 41.4 HEXADECIMAL NUMBER SYSTEM

- The hexadecimal number system has a base of 16.

- It uses the sixteen numerals

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F (Table 1).

**Table 1. Hexadecimal Digits**

Decimal	Hexadecimal
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	A
11	B
12	C
13	D
14	E
15	F

- After reaching 9 in the hexadecimal system, you continue counting as follows:

A, B, C, D, E, F.

Thus, A represents 10, B represents 11, ... F represents 15.

- After reaching F, two digit combinations are formed taking the second digit followed by the first, then the second followed by the second and so on.

After reaching F, two digit combinations are formed such as 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 1A, 1B, 1C, 1D, 1E, 1F, 20 (Table 2).

**Table 2**

Hexadecimal system	Decimal system
10	16
11	17
12	18
13	19
14	20
15	21
16	22
17	23
18	24
19	25
1A	26
1B	27
1C	28

1D	29
1E	30
1F	31
20	32

### (i) Hexadecimal-to-Decimal Conversion

In the hexadecimal number system, each digit position corresponds to a power of 16. The weights of the digit positions in a hexadecimal number are as follows:

$$\begin{array}{ccccccccc} 16^3 & 16^2 & 16^1 & 16^0 & . & 16^{-1} & 16^{-2} & 16^{-3} \\ & & & & \uparrow & & & \\ & & & & \text{Hexadecimal point} & & & \end{array}$$

Therefore, to convert from hexadecimal to decimal, multiply each hexadecimal digit by its weight and add the resulting products.

#### Examples of Hexadecimal to Decimal Conversion

$$(i) \quad 3\text{C}8_{16} = 3 \times 16^2 + 12 \times 16^1 + 8 \times 16^0 \\ = 768 + 192 + 8 = 968_{10}$$

$$(ii) \quad \text{E5F8}_{16} = 14 \times 16^3 + 5 \times 16^2 + 15 \times 16^1 + 8 \times 16^0 \\ = 57344 + 1280 + 240 + 8 = 58,872_{10}$$

**EXAMPLE 1.** Convert hexadecimal F8E6.39 to the equivalent decimal.

**SOL.** Hexadecimal F8E6.39 converts to decimal as follows:

$$\begin{aligned} \text{F8E6.39} &= \text{F}(16^3) + 8(16^2) + \text{E}(16^1) + 6(16^0) + 3(16^{-1}) + 9(16^{-2}) \\ &= 15(16^3) + 8(16^2) + 14(16^1) + 6(16^0) + 3(16^{-1}) + 9(16^{-2}) \\ &= 61,440 + 2048 + 224 + 6 + 0.1875 + 0.0352 \\ \text{F8E6.39} &= 63,718.2227 \end{aligned}$$

### (ii) Decimal-to-Hexadecimal Conversion

To convert from decimal to hexadecimal, **hex dabble method** is used.

- The decimal number is divided successively by 16, until we get a quotient of zero.
- Write down the **remainder** after each division.
- The remainders are converted to hex notations and read in the **reverse order i.e., from bottom to top**.

**EXAMPLE 1.** Convert decimal 2479 to hexadecimal.

**SOL.** To convert decimal 2479 to hexadecimal, the first division is

$$\begin{array}{r} 154 \qquad 15 \rightarrow \text{F} \\ 16 \overline{)2479} \end{array}$$

In this first division, we get a quotient of 154 with a remainder of 15 (equivalent to F).

Notice that remainders greater than 9 have to be changed to hexadecimal digits (10 becomes A, 15 becomes F, etc.). See Table 1.

The next step is

$$\begin{array}{r} 9 \qquad 10 \rightarrow \text{A} \\ 154 \qquad 15 \rightarrow \text{F} \\ 16 \overline{)2479} \end{array}$$

Here we obtain a quotient of 9 with a remainder of 10 (same as A).

The final step is

$$\begin{array}{r}
 0 & 9 \rightarrow 9 \\
 9 & 10 \rightarrow A \\
 154 & 15 \rightarrow F \\
 \hline
 16) 2479
 \end{array}
 \quad \text{Read down}$$

Therefore, hexadecimal 9AF is equivalent to decimal 2479.

### II Method

16	2479	remainder	hexadecimal notation
16	154	15	F ↑
16	9	10	A ↑
16	0	9	9

- Therefore, hexadecimal 9AF is equivalent to decimal 2479.

**EXAMPLE 2.** Convert decimal 72905 to hexadecimal system.

SOL.	16	72905	remainder	hexadecimal notation
	16	4556	9	9 ↑
	16	284	12	C
	16	17	12	C
	16	1	01	1
		0	01	1

Reading the remainders from bottom to top, the result is 11CC9.

Decimal 72905 = 11CC9 hexadecimal.

### 41.5 BINARY CODED DECIMAL CODE (BCD CODE)

- In this code, each digit of a decimal number (0 through 9) is represented by its four bit binary equivalent.
- The binary code for 9 is 1001. Since a decimal digit can be as large as 9, four bits are required to code each digit. Digital systems understand binary numbers only.

Binary Coded Decimal (BCD) refers to representation of digits 0–9 in decimal system by 4-bit unsigned binary numbers. The usual method is to follow 8421 encoding which employs conventional route of weight placements like 8 representing the weight of the 4<sup>th</sup> place (as  $2^{4-1} = 8$ ), 4, i.e.,  $2^{3-1}$  of the 3<sup>rd</sup> place, 2, i.e.,  $2^{2-1}$  of the 2<sup>nd</sup> place and 1, i.e.,  $2^{1-1}$  of the 1<sup>st</sup> place.

The decimal numbers 0–9 in 8421 code can be represented as shown in Table.

Table BCD-8421 Code

Decimal	BCD-8421
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101

6	0110
7	0111
8	1000
9	1001

**EXAMPLE.** Decimal number 29 in BCD-8421 is written as 00101001 (0010 representing 2 and 1001 representing 9).

- BCD equivalent and binary equivalent differ for decimal numbers 10 and above.

Decimal number	Binary equivalent	BCD Code
10	1010	0001 0000
11	1011	0001 0001
12	1100	0001 0010
:	:	:
27	11011	0010 0111
:	:	:
45	010001	0100 0101

- The binary weights of the four bits are  $2^3, 2^2, 2^1, 2^0$ . So it is called 8421 code.
- Group of four binary digits represents a single decimal digit.
- The lowest 4-bit group is 0000. The largest 4-bit group is 1001. Thus, the BCD digits are from 0000 to 1001.

#### Decimal number –to–BCD code Conversion.

**Illustrations.** (i) Consider a decimal number such as 874. Each digit is changed to its binary equivalent as follows:

8	7	4	(decimal)
↓	↓	↓	
1000	0111	0100	(BCD)

(ii) Decimal number 396 is changed to its binary equivalent as

3	9	6	
↓	↓	↓	
0011	1001	0110	

Thus decimal  $396 = 0011 \ 1001 \ 0110$  in BCD code.

**EXAMPLE 1.** Convert the decimal number 943 to its BCD code.

**SOL.**      9      4      3      (decimal)

1001	0100	0011	(BCD)
------	------	------	-------

Clearly, only 4 bit binary numbers from 0000 through 1001 are used. In other words, only 10 of the 16 possible 4 bit binary code groups are used. If any of the ‘forbidden’ 4 bit numbers ever occurs in a machine using BCD code, it is usually an indication that an error has occurred.

**EXAMPLE 2.** Represent  $(137)_{10}$  in straight binary code and BCD code.

**SOL.**       $(137)_{10} = (10001001)_2$       (Straight binary code)

$(137)_{10} = 0001 \ 0011 \ 0111$  (BCD code)

**BCD code –to Decimal number Conversion**

- The BCD code can be converted into decimal number as follows:

BCD	0101	0111	1000
	↓	↓	↓
decimal	5	7	8

578 is the decimal equivalent of the BCD 0101 0111 1000

**Advantages.** (i) BCD code is very convenient and useful code for input and output operations in digital circuits. It is used to represent decimal digits in systems like digital calculators, voltmeter etc.

(ii) We need to remember the binary numbers for 0 to 9 as we encode only one decimal digit at a time.

**Disadvantage.** Rules for binary addition do not apply to the whole BCD number but only to the four bit group.

**EXAMPLE.** If we add 11 and 9, then

Decimal	Binary	BCD code
11	1011	0001 0001
+ 9	+ 1001	+ 1001
20	10100	0001 1010

↑  
Error

We get an unacceptable answer by BCD code because in BCD code 1001 is the largest four bit group and four bit groups such as 1010, 1011, 1100, 1101, 1110 and 1111 are forbidden. If these groups appear in a machine using BCD code, an error is indicated. Thus rules for binary addition do not apply to the entire BCD number.

## 41.6 ASCII CODE

To get information into and out of a computer, we need to use some kind of *alphanumeric* code (one for letters, numbers, and other symbols). At one time, manufacturers used their own alphanumeric codes. This led to all kinds of confusion. Finally, industry settled on an input-output code. It is called ASCII (pronounced ask'-ee) code.

- ASCII is an abbreviation for *American Standard Code for Information Interchange*.
- This code allows manufacturers to standardize computer hardware such as keyboards, printers, and video displays.
- The ASCII code is a 7-bit code proposed by ANSI (American National Standard Institute). The format of ASCII code is  $X_6X_5X_4X_3X_2X_1X_0$ .

Here, each  $X$  is a 0 or a 1.

Table 41.1 lists the characters and their equivalent ASCII code.

**Table 41.1. ASCII Code**

$X_3X_2X_1X_0$	$X_6X_5X_4$					
	010	011	100	101	110	111
0000	SP	0	@	P		p
0001	!	1	A	Q	a	q
0010	"	2	B	R	b	r
0011	#	3	C	S	c	s

0100	\$	4	D	T	d	t
0101	%	5	E	U	e	u
0110	&	6	F	V	f	v
0111	,	7	G	W	g	w
1000	(	8	H	X	h	x
1001	)	9	I	Y	i	y
1010	*	:	J	Z	j	z
1011	+	;	K		k	
1100	,	<	L		l	
1101	-	=	M		m	
1110	•	>	N		n	
1111	/	?	O		o	

### Using the ASCII Code

We can use Table 41.1 to find the ASCII code for the uppercase and lowercase letters of the alphabet and some of the most commonly used symbols.

- (i) The Table shows that the capital letter *A* has an  $X_6X_5X_4$  of 100 and an  $X_3X_2X_1X_0$  of 0001. The ASCII code for *A* is, therefore,

1000001

For easier reading, we can leave a space as follows:

100 0001 (A)

- (ii) The letter *a* is coded as

110 0001 (a)

- (iii) The codes for *b*, *c* and *d* are:

110 0010 (b)

110 0011 (c)

110 0100 (d)

- (iv) The ASCII code of the mathematical symbols + and = are:

010 1011 (+)

011 1101 (=)

- (v) In Table 41.1, SP stands for space (blank). Hitting the space bar of an ASCII keyboard sends this into a microcomputer:

010 0000 (space)

### Parity Generators/Checkers

In a digital system while codes are being transferred from one point to another, errors can occur. There will be undesired changes in the coded information like a 0 may change to 1 and a 1 may change to 0. Errors cause serious concern in digital systems.

Error detection can be done using parity bit. A parity bit should be attached to a group of information bits to make the total number of 1s either even or odd depending on whether even parity or odd parity system is used.

### Parity Bit

The ASCII code is used for sending digital data over telephone lines.

### Meaning of Parity

*Even parity* means an  $n$ -bit input has an even number of 1s. For instance, 110011 has even parity because it contains four 1s. *Odd parity* means an  $n$ -bit input has an odd number of 1s. For example, 110001 has odd parity because it contains three 1s.

Because of transients, noise, and other disturbances, 1-bit errors sometimes occur when binary data is transmitted over telephone lines. To catch these errors, a parity bit is usually transmitted along with the original bits.

Then a parity checker at the receiving end can test for even or odd parity, whichever parity has been prearranged between the sender and the receiver.

- Since ASCII code uses 7 bits, the addition of a parity bit to the transmitted data produces an 8-bit number in this format:

$$\begin{array}{c} X_7 X_6 X_5 X_4 \quad X_3 X_2 X_1 X_0 \\ \uparrow \\ \text{Parity bit} \end{array}$$

This is an ideal length because most digital equipment is set up to handle bytes of data.

## 41.7 DATA SELECTION CIRCUITS

Circuits that are used for Data Selection are called Data Selection circuits. They are of two types:

- (i) **Multiplexer (MUX)**
- (ii) **De-multiplexer (DEMUX)**

Multiplexing and Demultiplexing are used when data from several sources are to be transmitted over one line to a distant location and redistributed to several destinations.

### (i) Multiplexer (MUX)

*Multiplex* means *many into one*. A *multiplexer* is a circuit with many inputs but only one output. By applying control signals, we can steer any input to the output. Thus it is also called a *data selector* and control inputs are termed select inputs.

#### **$n$ -to-1 Multiplexer**

Fig. 41.1 shows  $n$ -to-1 multiplexer block diagram.

- The circuit has  $n$  input signals,  $m$  control signals and 1 output signal.
- $m$  control signals can select at the most  $2^m$  input signals.  
 $\therefore n \leq 2^m$ .

#### **4-to-1 Multiplexer**

Fig. 41.2 shows the logic symbol for 4-to-1 multiplexer.

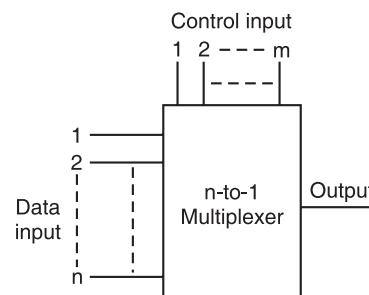


Fig. 41.1

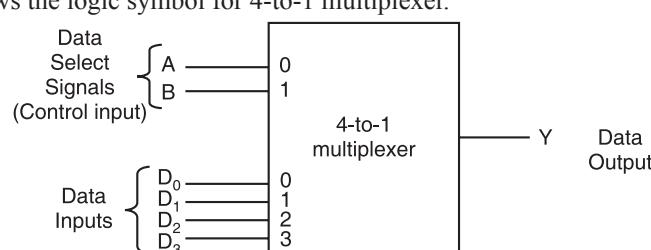
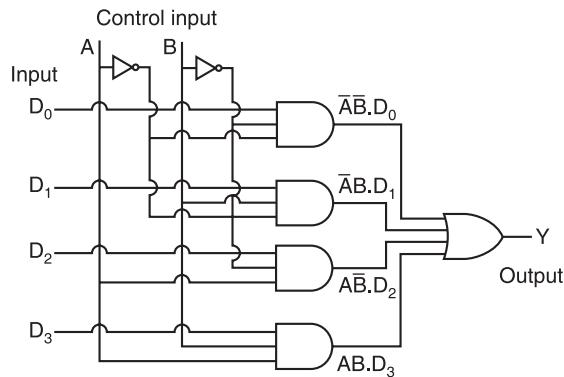


Fig. 41.2

**Logic circuit:**

Fig. 41.3 shows the circuit diagram of a 4-to-1 multiplexer.

Fig. 41.4 shows the truth table.



**Fig. 41.3**  
4-to-1 multiplexer truth table

A	B	Y
0	0	$D_0$
0	1	$D_1$
1	0	$D_2$
1	1	$D_3$

**Fig. 41.4**

- The four input signal lines are marked  $D_0$ ,  $D_1$ ,  $D_2$  and  $D_3$ .
- The control lines are marked  $A$  and  $B$ .
- Depending on control inputs  $A$ ,  $B$  one of the four inputs  $D_0$  to  $D_3$  is steered to output  $Y$ .
- For  $AB = 00$ , the first AND gate to which  $D_0$  is connected remains active and equal to  $D_0$  and all other AND gates are inactive with output held at logic 0. Thus, multiplexer output  $Y$  is same as  $D_0$ . If  $D_0 = 0$ ,  $Y = 0$  and if  $D_0 = 1$ ,  $Y = 1$ . Data on input  $D_0$  appears on the output line.

Similarly, for  $AB = 01$ , second AND gate will be active and all other AND gates remain inactive. Thus, output  $Y = D_1$ . Data on  $D_1$  appears on the output line.

If  $AB = 10$ , then data on  $D_2$  will appear on the output.

If  $AB = 11$ , then data on  $D_3$  will appear on the output.

**Logic Equation (For optional reading)**

Let us write the logic equation of this circuit. Clearly, it will give a SOP representation, each AND gate generating a product term, which finally are summed by OR gate. Thus,

$$Y = A'B'.D_0 + A'B.D_1 + AB'.D_2 + AB.D_3$$

$$\text{If } A = 0, B = 0, \quad Y = 0'0'.D_0 + 0'.0.D_1 + 0.0'.D_2 + 0.0D_3$$

$$\text{or,} \quad Y = 1.1.D_0 + 1.0.D_1 + 0.1.D_2 + 0.0.D_3$$

$$\text{or,} \quad Y = D_0$$

**Applications:**

- Multiplexers are used for data transmission or communication.

2. A multiplexer is used for connecting two or more sources to a single destination among computer units.
3. It is also useful for constructing a common-bus system in a microprocessor.

### (ii) Demultiplexer

**What is a Demultiplexer?** Demultiplex means *one into many*.

A demultiplexer is a logic circuit with one input and many outputs. By applying control signals, we can steer the input signal to one of the output lines.

**Demultiplexer block diagram.** Figure 41.5 shows the demultiplexer block diagram. The circuit has 1 input signal,  $m$  control or select signals and  $n$  output signals where  $n \leq 2^m$ .

Demultiplexer is a logic circuit that switches digital data from one input line to several output lines in a specified time sequence. It reverses the multiplexing function. It has one input line,  $m$  select lines and  $2^m$  output lines.

### Logic circuit of 1-to-2 demultiplexer,

Figure 41.6 shows the circuit diagram of a 1-to-2 demultiplexer.

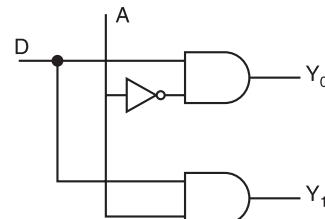


Fig. 41.6

**Commercially available** demultiplexer ICs.

IC No.	Demux Type	Decoder Type
74154	1-to-16	4-to-16
74138	1-to-8	3-to-8
74155	1-to-4	2-to-4

Note the similarity of multiplexer and demultiplexer circuits in generating different combinations of control variables through a bank of AND gates.

### 1-to-4 Demultiplexer

Figure 41.7 shows logic symbol for 1-to-4 DEMUX.

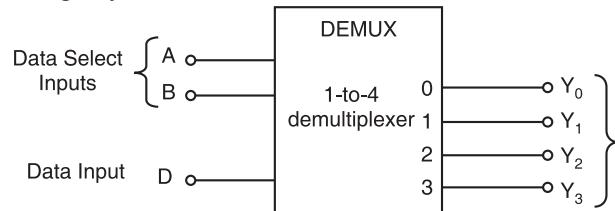


Fig. 41.7

### Logic circuit

Figure 41.8 shows a 1-to-4 demultiplexer.

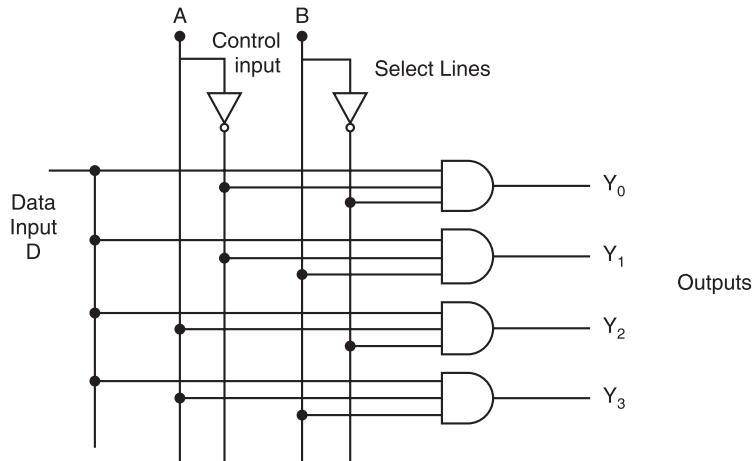


Fig. 41.8

The input bit is labelled  $D$ . This data bit ( $D$ ) is transmitted to the data bit of the output lines. But which one? This depends on the value of  $AB$ , the control input.

- When  $AB = 00$ , the upper AND gate is enabled while all other AND gates are disabled. Therefore, data bit  $D$  is transmitted only to the  $Y_0$  output, giving  $Y_0 = D$ . If  $D$  is low,  $Y_0$  is low. If  $D$  is high,  $Y_0$  is high.  
The value of  $Y_0$  depends on the value of  $D$ . All other outputs are in the low state.
- If the control nibble is changed to  $AB = 11$ , all gates are disabled except the bottom AND gate. Then,  $D$  is transmitted only to the  $Y_3$  output, and  $Y_3 = D$ .
- The two data select lines enable only one gate at a time and data appearing on the data-input line will pass through the selected gate to the output line.

## LOGIC CIRCUITS

### 41.8 BOOLEAN ALGEBRA

The algebra of logic prominently used in the operation of computer devices is the algebra developed by George Boole. It is a binary or two-valued logic, *i.e.*, it permits only two values or states for its variables. These two states are ‘true’ and ‘false’ in logic but are represented by ‘on’ and ‘off’ states of electronic circuits.

The two variables of the Boolean algebra are usually represented by 0 and 1. Hence, every variable is either a 0 or a 1. There are no negative or fractional numbers. Logically, we may write :

If  $X = 0$  then  $X \neq 1$

And if  $X = 1$  then  $X \neq 0$ .

Boolean algebra uses only three operations on its variables. These operations are:

- The OR addition represented by a + (plus) sign.
- The AND multiplication represented by a  $\times$  (cross), or a . (dot) sign.

We will normally write the adjacent letters without the dot, such as  $AB$ .

- The NOT operation represented by a bar over a variable.

#### 41.8.1 Boolean Addition and Multiplication

Addition in Boolean algebra involves variables having values of either a binary 1 or a binary 0. The basic rules for Boolean addition are as follows :

$$\begin{aligned} 0 + 0 &= 0 \\ 0 + 1 &= 1 \\ 1 + 0 &= 1 \\ 1 + 1 &= 1 \end{aligned}$$

In the application of Boolean algebra to logic circuits, *Boolean addition is the same as the OR*.

The basic rules for Boolean multiplication are :

$$\begin{aligned} 0 \cdot 0 &= 0 \\ 0 \cdot 1 &= 0 \\ 1 \cdot 0 &= 0 \\ 1 \cdot 1 &= 1 \end{aligned}$$

*Boolean multiplication is the same as the AND.*

#### 41.8.2 Logic Expressions

##### NOT

The operation of an inverter (NOT circuit) can be expressed with symbols as follows : If the input variable is called  $A$  and the output variable is called  $X$ , then  $X = \bar{A}$ . This expression states that the output is the complement of the input, so that if  $A = 0$ , then  $X = 1$ , and if  $A = 1$ , then  $X = 0$  (Fig. 41.9).

$$(i) \bar{0} = 1 \quad (ii) \bar{1} = 0$$

##### AND

The operation of a two-input AND gate can be expressed in equation form as follows : If one input variable is  $A$ , the other input variable is  $B$ , and the output variable is  $X$ , then the Boolean expression for this basic gate function is  $X = AB$  (Fig. 41.10).

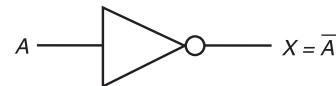


Fig. 41.9

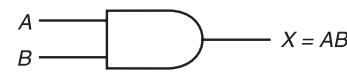


Fig. 41.10

##### OR

The operation of a two-input OR gate can be expressed in equation form as follows : If one input is  $A$ , the other input is  $B$ , and the output is  $X$ , then the Boolean expression is  $X = A + B$  (Fig. 41.11).

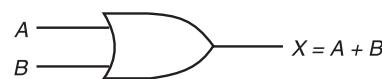


Fig. 41.11

It is read as  $A$  (OR)  $B$  and not as  $A$  (plus)  $B$ .

In Boolean algebra, there are three operations only. They are

1. ‘OR’ operation, denoted by  $(+)$
2. ‘AND’ operation, denoted by  $(\cdot)$
3. ‘NOT’ operation, denoted by  $(-)$

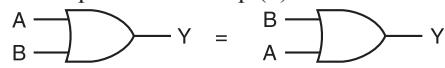
#### 41.8.3 Commutative Laws

- (i) The commutative law of addition for two variables is written algebraically as

$$A + B = B + A \quad \dots(1)$$

This states that the order in which the variables are ORed makes no difference.

Fig. 41.12 is a graphical equivalent for Eq. (1).



**Fig. 41.12**

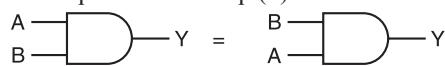
- The inputs to an OR gate can be transposed without changing the output.

(ii) The commutative law of multiplication of two variables is

$$AB = BA \quad \dots(2)$$

This states that the order in which the variables are ANDed makes no difference.

Fig. 41.13 is a graphical equivalent for Eq. (2).



**Fig. 41.13**

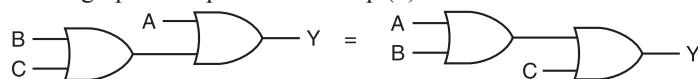
#### 41.8.4 Associative Laws

(i) The associative law of addition is stated as follows for three variables :

$$A + (B + C) = (A + B) + C \quad \dots(3)$$

This law states that in the ORing of several variables, the result is the same regardless of the grouping of the variables.

Fig. 41.14 is a graphical equivalent for Eq. (3).



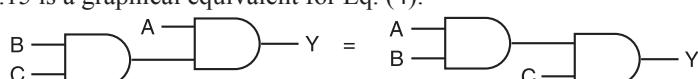
**Fig. 41.14**

(ii) The associative law of multiplication is stated as follows for three variables :

$$A(BC) = (AB)C \quad \dots(4)$$

This law tells us that it makes no difference in what order the variables are grouped when ANDing several variables.

Fig. 41.15 is a graphical equivalent for Eq. (4).



**Fig. 41.15**

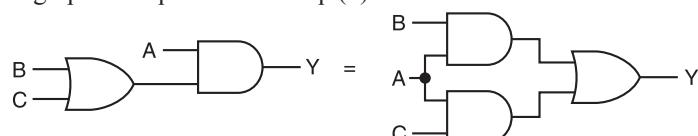
- The associative laws show that the order of combining variables has no effect on the final answer.

#### 41.8.5 Distributive Law

The distributive law is written for three variables as follows :

$$A(B+C) = AB + AC \quad \dots(5)$$

Fig. 41.16 is a graphical equivalent for Eq. (5).



**Fig. 41.16**

This law states that ORing several variables and ANDing the result with a single variable is equivalent to ANDing the single variable with each of the several variables and then ORing the products.

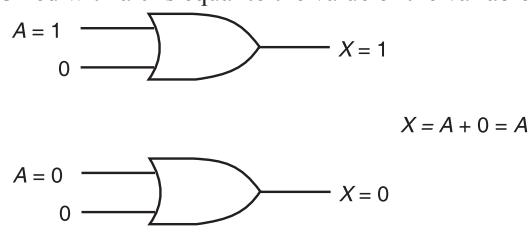
#### 41.8.6 Rules for Boolean Algebra

Table 41.2 lists *Basic rules of Boolean algebra*.

**Table 41.2**

1.  $A + 0 = A$
2.  $A + 1 = 1$
3.  $A \cdot 0 = 0$
4.  $A \cdot 1 = A$
5.  $A + A = A$
6.  $A + \bar{A} = 1$
7.  $A \cdot A = A$
8.  $A \cdot \bar{A} = 0$
9.  $\bar{\bar{A}} = A$
10.  $A + AB = A$
11.  $A + \bar{A}B = A + B$
12.  $(A + B)(A + C) = A + BC$

**Rule 1.** A variable ORed with a 0 is equal to the value of the variable ( $A + 0 = A$ ) (Fig. 41.17).



**Fig. 41.17**

Similarly, the other rules are proved.

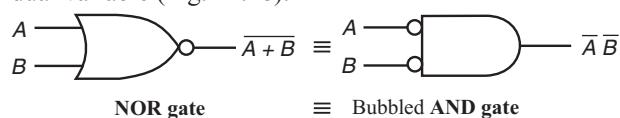
#### 41.9 DE MORGAN'S THEOREMS

##### Theorem 1.

**Statement:** The complement of the sum of two variables is equal to the product of complements of the variables, i.e.,

$$\overline{A + B} = \overline{A} \overline{B}$$

The complement of two variables ORed is the same as the AND of the complements of each individual variable (Fig. 41.18).



**Fig. 41.18**

$A$  and  $B$  are two variables.

- The first theorem says a NOR gate and a bubbled AND gate are equivalent.

**Truth Table**

$A$	$B$	$\overline{A} + \overline{B}$	$\overline{A} \overline{B}$
0	0	1	1
0	1	1	1
1	0	1	1
1	1	0	0

**Proof:** This theorem can be proved by substituting the two permitted values of the variables, i.e., 0 and 1, on both sides of the identity.

In each case, the left hand side will equal the right hand side, thus proving the theorem.

**Case 1.** When

$$\begin{aligned} A &= 0, \quad B = 0 \\ \text{L.H.S.} &= \overline{A+B} = \overline{0+0} = \overline{0} = 1. \\ \text{R.H.S.} &= \overline{A} \cdot \overline{B} = \overline{0} \cdot \overline{0} = 1 \cdot 1 = 1. \end{aligned}$$

**Case 2.** When

$$\begin{aligned} A &= 0, \quad B = 1 \\ \text{L.H.S.} &= \overline{A+B} = \overline{0+1} = \overline{1} = 0. \\ \text{R.H.S.} &= \overline{A} \cdot \overline{B} = \overline{0} \cdot \overline{1} = 1 \cdot 0 = 0. \end{aligned}$$

**Case 3.** When

$$\begin{aligned} A &= 1, \quad B = 0 \\ \text{L.H.S.} &= \overline{A+B} = \overline{1+0} = \overline{1} = 0. \\ \text{R.H.S.} &= \overline{A} \cdot \overline{B} = \overline{1} \cdot \overline{0} = 0 \cdot 1 = 0. \end{aligned}$$

**Case 4.** When

$$\begin{aligned} A &= 1, \quad B = 1 \\ \text{L.H.S.} &= \overline{A+B} = \overline{1+1} = \overline{1} = 0. \\ \text{R.H.S.} &= \overline{A} \cdot \overline{B} = \overline{1} \cdot \overline{1} = 0 \cdot 0 = 0. \end{aligned}$$

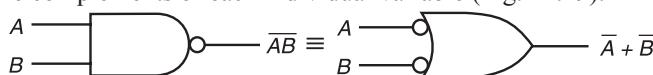
In every case, left hand side (L.H.S.) is equal to right hand side (R.H.S.). Therefore, the theorem is proved.

### Theorem 2.

**Statement:** The complement of the product of two variables is equal to the sum of complements of the variables, i.e.,

$$\overline{AB} = \overline{A} + \overline{B}$$

The complement of two variables ANDed is the same as the OR of the complements of each individual variable (Fig. 41.19).



NAND Gate  $\equiv$  Bubbled OR Gate

Truth Table			
A	B	$\overline{AB}$	$\overline{A} + \overline{B}$
0	0	1	1
0	1	1	1
1	0	1	1
1	1	0	0

Fig. 41.19

$A$  and  $B$  are two variables.

- The second theorem says a NAND gate and a bubbled OR gate are equivalent.

**Proof:** This theorem can be proved by substituting the two permitted values of the variables, i.e., 0 and 1, on both sides of the identity.

In each case, the left hand side will equal the right hand side, thus proving this theorem.

**Case 1.** When

$$\begin{aligned} A &= 0, \quad B = 0 \\ \text{L.H.S.} &= \overline{A \cdot B} = \overline{0 \cdot 0} = \overline{0} = 1. \\ \text{R.H.S.} &= \overline{A} + \overline{B} = \overline{0} + \overline{0} = 1 + 1 = 1. \end{aligned}$$

**Case 2.** When

$$\begin{aligned} A &= 0, \quad B = 1 \\ \text{L.H.S.} &= \overline{A \cdot B} = \overline{0 \cdot 1} = \overline{0} = 1. \\ \text{R.H.S.} &= \overline{A} + \overline{B} = \overline{0} + \overline{1} = 1 + 0 = 1. \end{aligned}$$

**Case 3.** When

$$A = 1, \quad B = 0$$

$$\text{L.H.S.} = \overline{A \cdot B} = \overline{1 \cdot 0} = \overline{0} = 1.$$

$$\text{R.H.S.} = \overline{A} + \overline{B} = \overline{1} + \overline{0} = 0 + 1 = 1.$$

**Case 4.** When

$$A = 1, \quad B = 1$$

$$\text{L.H.S.} = \overline{A \cdot B} = \overline{1 \cdot 1} = \overline{1} = 0.$$

$$\text{R.H.S.} = \overline{A} + \overline{B} = \overline{1} + \overline{1} = 0 + 0 = 0.$$

In every case, left hand side of the expression is equal to right hand side of the expression. Therefore, the theorem is proved.

## LOGIC GATES

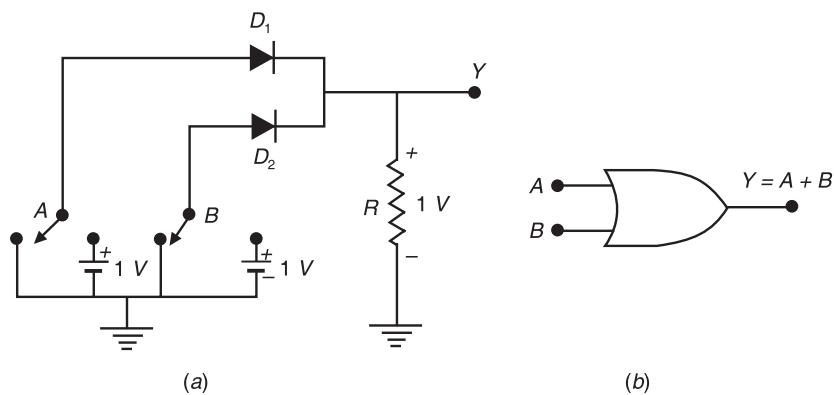
### 41.10 THE BASIC GATES—NOT, OR, AND

A logic gate is an electronic circuit which performs logic functions or takes a logic decision. It has one output and one or more inputs. Logic gates are the building blocks of the digital systems. They work on the logical algebra developed by George Boole. The Boolean operations namely ‘OR’ operation, ‘AND’ operation and ‘NOT’ operation are implemented by three logic gates called ‘OR’ gate, ‘AND’ gate and ‘NOT’ gate.

#### 41.10.1 The OR Gate

It implements Boolean ‘OR’ operation. An ‘OR’ gate has two (or) more inputs and one output. An ‘OR’ gate is a logic gate whose output is ‘1’ state if any or all the inputs are in ‘1’ state.

Fig. 41.20 (a) shows a two input OR gate using two ideal diodes  $D_1$  and  $D_2$ . Here  $A$  and  $B$  represent the two inputs and  $Y$  the output.  $R$  represents the output load resistor. Fig. 41.20 (b) gives the symbolic representation of the OR gate.



**Fig. 41.20**

Suppose that the two input voltages are either 0 or 1 V.

**Case (i).** When  $A = 0; B = 0$ , both the input voltages are zero. So both the diodes  $D_1$  and  $D_2$  do not conduct. Hence, current through  $R$  is zero and so output voltage is zero, i.e.,  $Y = 0$ .

**Case (ii).** When  $A = 1; B = 0$ , the diode  $D_1$  is forward biased and so it conducts.  $D_2$  does not conduct. Since  $D_1$  is conducting, current flows through  $R$ . Hence, there is an output voltage, i.e.,  $Y = 1$ .

**Case (iii).** When  $A = 0; B = 1$ , the diode  $D_2$  is forward biased and so it conducts.  $D_1$  does not conduct. Since  $D_2$  is conducting, current flows through  $R$ . Hence, there is an output voltage, i.e.,  $Y = 1$ .

**Case (iv).** When  $A = 1; B = 1$ , both the diodes  $D_1$  and  $D_2$  are forward biased. Hence, both are conducting. So current flows through  $R$  and there is an output voltage, i.e.,  $Y = 1$ .

The logic operation of the OR gate can be summarised in a tabular form known as Truth Table. A truth table may be defined as a table which gives the output state for all possible input combinations.

#### 41.10.2 The AND Gate

An "AND" gate implements Boolean 'AND' operation. It has two or more inputs and one output. An 'AND' gate is a logic gate whose output is '1', if and only if all the inputs are in '1' state.

Fig. 41.21 (a) shows a two-input AND gate using two ideal diodes and Fig. 41.21 (b) gives its symbolic representation. Here  $A$  and  $B$  represent the two inputs and  $Y$  the output.  $R$  represents the output load resistor. The two input voltages are assumed to be either 0 or 1 V.

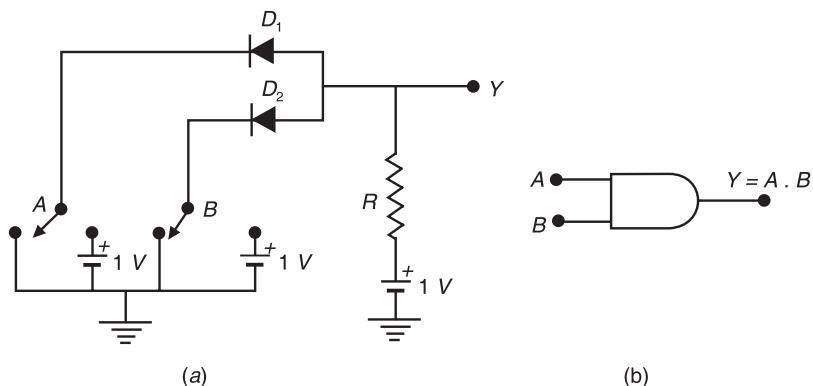


Fig. 41.21

(i) When both  $A = 0$  V and  $B = 0$  V, the inputs are short circuited to ground. The 1 V battery in the output side biases the diodes  $D_1$  and  $D_2$  in the forward direction. Hence, both diodes conduct. The output is also shorted to ground through the diodes. Thus, the output  $Y = 0$  V.

(ii) When  $A = 1$  V and  $B = 0$  V, diode  $D_2$  conducts and the output is short circuited to ground through this diode. Therefore, output  $Y = 0$  V.

(iii) When  $A = 0$  V and  $B = 1$  V, diode  $D_1$  conducts and the output is short circuited to ground through this diode. Therefore, the output  $Y = 0$  V.

(iv) When both  $A = 1$  V and  $B = 1$  V, neither  $D_1$  nor  $D_2$  conducts. No current, therefore, flows through  $R$  and the output  $Y = 1$  V.

The above results are tabulated in the Truth table.

Truth Table

$A$	$B$	$Y = A + B$
0	0	0
1	0	1
0	1	1
1	1	1

#### 41.10.3 The NOT Gate

The NOT circuit has only one input and one output. It inverts the polarity of a pulse applied to it. If the input is 1, the output is 0 and when the input is 0 the output is 1. That is, the NOT gate inverts the input. A 'NOT' gate using transistor is shown in Fig. 41.22 (a) and its symbol in Fig. 41.22 (b).

Truth Table

$A$	$B$	$Y = A \cdot B$
0	0	0
1	0	0
0	1	0
1	1	1

(i) When no signal is applied at the input, i.e.,  $A = 0$  V, the transistor is cut OFF, making the collector current zero. Thus, the potential drop across  $R$  is zero. The supply voltage of  $V_{cc}$  appears at the output terminal. Hence, the output  $Y = V_{cc}$ .

Thus, when input is low, output is high.

(ii) When a positive pulse is applied to  $A$ , i.e.,  $A = 1$ , the transistor conducts (fully ON) drawing maximum collector current. Hence, whole of  $V_{cc}$  drops across  $R$  and output  $Y = 0$  V. Thus, when input is high, the output is held at a low value.

The results are tabulated in the Truth table.

**Truth Table for NOT Gate**

Input	Output
$A$	$Y = \bar{A}$
0	1
1	0

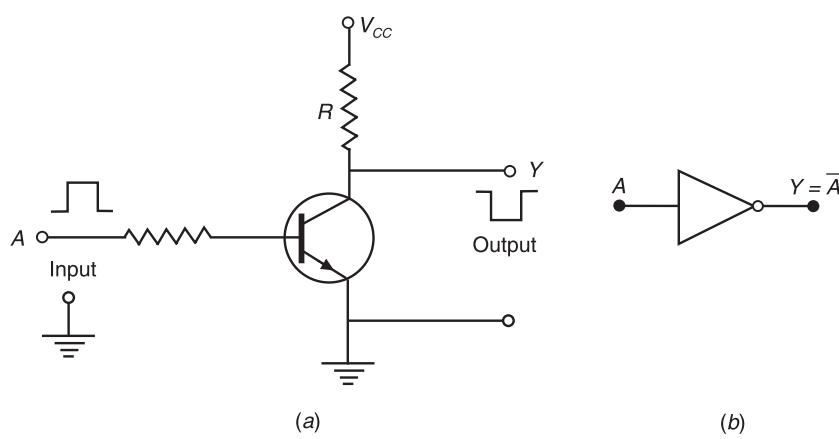


Fig. 41.22

## UNIVERSAL LOGIC GATES—NOR, NAND

### 41.11 THE NOR GATE

A negation following an OR is called a NOT – OR or a NOR gate.

It is the circuit combination in which a NOT circuit follows an OR gate.

A NOR gate can be implemented by placing a transistor ‘NOT’ circuit after the diode ‘OR’ circuit (Fig. 41.23).

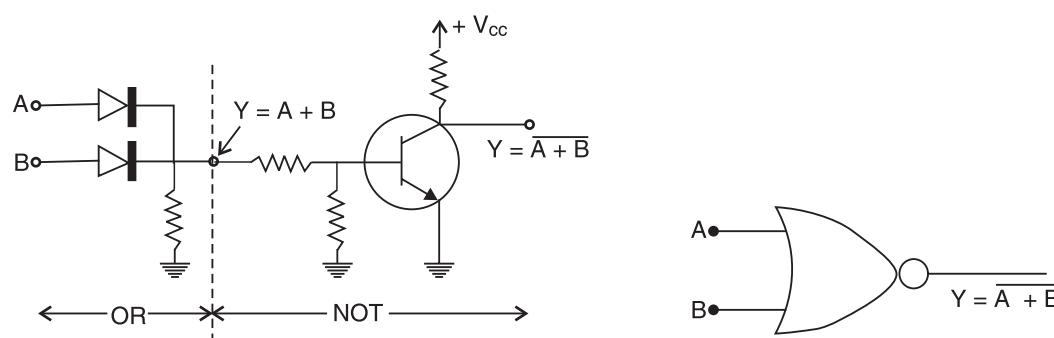


Fig. 41.23

The symbol of a NOR gate is shown in Fig 41.24

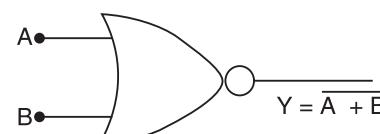


Fig. 41.24

The Boolean expression for the NOR gate is  $Y = \overline{A + B}$ .

The output of the OR gate is inverted by the NOT gate and thus yields NOR operation.

**Truth Table.** Consider the input values as follows:

**Case 1 :** When  $A = 0$  and  $B = 0$ , then  $A + B = 0$ .

$$\therefore Y = \overline{A + B} = 1.$$

**Case 2 :** When  $A = 0$  and  $B = 1$ , then  $A + B = 1$ .

$$\therefore Y = \overline{A + B} = 0.$$

**Case 3 :** When  $A = 1$  and  $B = 0$ , then  $A + B = 1$ .

$$\therefore Y = \overline{A + B} = 0.$$

**Case 4 :** When  $A = 1$  and  $B = 1$ , then  $A + B = 1$ .

$$\therefore Y = \overline{A + B} = 0.$$

**Truth Table – NOR Operation**

Input		Output
A	B	$Y = \overline{A + B}$
0	0	1
0	1	0
1	0	0
1	1	0

#### 41.11.1 NOR Gate is a Universal Gate

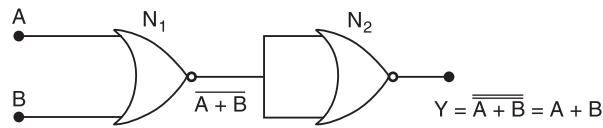
A NOR gate can be used to realize the basic logic functions: OR, AND and NOT. So NOR gate is also called **universal gate**. Thus, in digital circuits, NOR gate serves as a building block.

If all the three basic gates can be constructed using NOR gates alone, then NOR can be called a universal gate.

##### 1. As OR gate

An 'OR gate can be constructed using 'NOR' gates as follows.

Consider a two input 'NOR' gate ( $N_1$ ) and a single input 'NOR' gate  $N_2$  (Fig. 41.25).



**Fig. 41.25**

By connecting the output of two input 'NOR' gate ( $N_1$ ) as the input of the single input 'NOR' gate ( $N_2$ ), an 'OR' gate is constructed. The output of  $N_1$  is  $\overline{A + B}$ , and this is used as the input of  $N_2$ . Hence the output of  $N_2$  is

$$Y = \overline{\overline{A + B}} = A + B \quad (\because \overline{\overline{A}} = A).$$

Thus the combination of the 'NOR' gates  $N_1$  and  $N_2$  as shown in Fig. 41.25 acts as an 'OR' gate.

##### 2. As AND gate

An 'AND' gate can be constructed using 'NOR' gate as follows.

Consider two single input 'NOR' gates  $N_1$  &  $N_2$  and a two input 'NOR' gate  $N_3$ . The outputs of the NOR gates  $N_1$  &  $N_2$  are used as the inputs of  $N_3$  (Fig. 41.26).

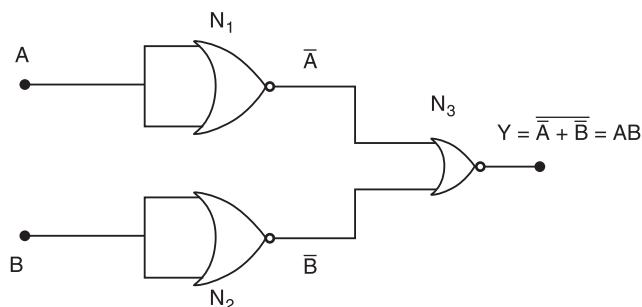


Fig. 41.26

The outputs of  $N_1$  &  $N_2$  are  $\bar{A}$  and  $\bar{B}$ . These are used as the inputs of  $N_3$ . Hence the output of  $N_3$  is

$$Y = \overline{\bar{A} + \bar{B}} = \overline{\overline{A} \cdot \overline{B}}$$
 (By De Morgan's Theorem)

$$= A \cdot B (\because \overline{\overline{A}} = A)$$

Thus the combination of the 'NOR' gates  $N_1$ ,  $N_2$  and  $N_3$  as shown in Fig. 41.26, acts as an 'AND' gate.

### 3. As NOT gate

(i) The two inputs of NOR gate are tied together, as shown in Fig. 41.27(a).

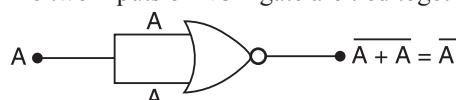


Fig. 41.27 (a)



Fig. 41.27 (b)

The output is  $\overline{A + A} = \overline{A}$

By De Morgan's Theorem,

$$\overline{A + A} = \overline{A}.$$

(ii) When only single input is used, the circuit symbol of Fig. 41.27 (b) is widely used.

## 41.12 THE NAND GATE

A negated AND is called a NOT-AND or NAND gate. It is obtained by connecting a NOT gate in the output of an AND gate.

Fig. 41.28 gives the logic symbol for two input NAND gate.

The output of the AND gate is inverted by the NOT gate.

Its output is given by the Boolean equation  $Y = \overline{AB}$ .

The NAND gate is implemented by placing a transistor NOT circuit after the diode AND circuit (Fig. 41.29).

**Truth Table.** Consider the input values as follows :

**Case 1:** When  $A = 0$  and  $B = 0$ , then  $AB = 0$ .

$$\therefore Y = \overline{AB} = 1.$$

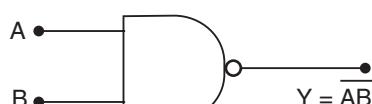


Fig. 41.28

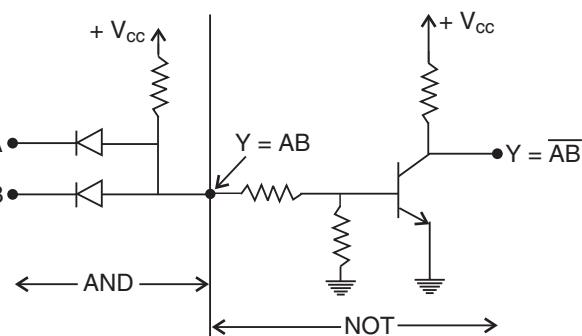


Fig. 41.29

**Case 2 :** When  $A = 0$  and  $B = 1$ , then  $AB = 0$ .

$$\therefore Y = \overline{AB} = 1.$$

**Case 3 :** When  $A = 1$  and  $B = 0$ , then  $AB = 0$ .

$$\therefore Y = \overline{AB} = 1.$$

**Case 4 :** When  $A = 1$  and  $B = 1$ , then  $AB = 1$ .

$$\therefore Y = \overline{AB} = 0.$$

Truth Table – NAND Operation

Input		Output
A	B	$Y = \overline{A \bullet B}$
0	0	1
0	1	1
1	0	1
1	1	0

Thus this gate gives output 1 if either A or B or both are 0.

#### 41.12.1 NAND Gate is a Universal Gate

A NAND gate is known as a universal gate because it can be used to realize all the three basic logic functions of an OR gate, AND gate and NOT gate. Thus, in digital circuits, NAND gate serves as a building block.

##### 1. As OR Gate

OR gate can be made out of the three NAND gates (Fig. 41.30).

However, the OR function may not be very clear from the figure because De Morgan's theorem is needed to prove that

$$\overline{\overline{A} \bullet \overline{B}} = A + B.$$

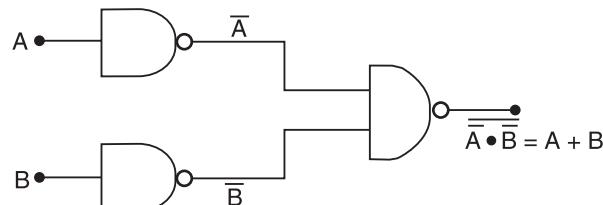


Fig. 41.30

##### 2. As AND Gate

An AND gate can be produced by using two NAND gates (Fig 41.31).

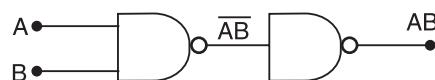


Fig. 41.31

##### 3. As NOT Gate

(i) If the two inputs of a NAND gate are connected together, then we get a NOT gate (Fig. 41.32a).

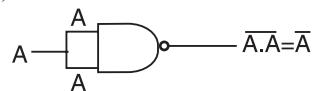


Fig. 41.32(a)



Fig. 41.32(b)

(ii) When only single input is used, the circuit symbol of Fig. 41.32b is used.

### 41.13 THE EXCLUSIVE-OR GATE

The exclusive-OR operation obeys the definition that a two input circuit provides an output when one input or the other is present but not when both inputs are present.

The exclusive-OR operation may also be stated as, "If  $A = 1$  and  $B = 0$ , or if  $B = 1$  and  $A = 0$ , then  $Y = 1$ ".

The Boolean expression is

$$Y = A\bar{B} + B\bar{A}.$$

The exclusive - OR operation is denoted by  $\oplus$ .

$$Y = A \oplus B = A\bar{B} + B\bar{A}.$$

The block diagram according to this logic is given in Fig. 41.33.

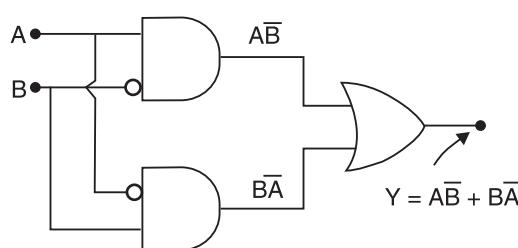


Fig. 41.33

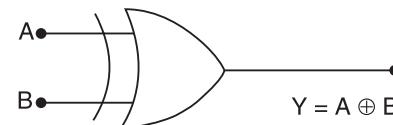


Fig. 41.34

It is constructed using two NOT gates, two AND gates and one OR gate.

The logic symbol for the exclusive-OR gate is shown in Fig. 41.34.

**Truth table.** We take the input values as follows:

**Case 1 :** When

$$A = 0, B = 0; \text{ so that } \bar{A} = 1 \text{ and } \bar{B} = 1.$$

Then

$$Y = 0 \cdot 1 + 1 \cdot 0 = 0 + 0 = 0.$$

**Case 2 :** When

$$A = 0, B = 1; \text{ so that } \bar{A} = 1 \text{ and } \bar{B} = 0.$$

Then

$$Y = 0 \cdot 0 + 1 \cdot 1 = 0 + 1 = 1.$$

**Case 3 :** When

$$A = 1, B = 0; \text{ so that } \bar{A} = 0 \text{ and } \bar{B} = 1.$$

Then

$$Y = 1 \cdot 1 + 0 \cdot 0 = 1 + 0 = 1.$$

**Case 4 :** When

$$A = 1, B = 1; \text{ so that } \bar{A} = 0 \text{ and } \bar{B} = 0.$$

Then

$$Y = 1 \cdot 0 + 0 \cdot 1 = 0 + 0 = 0.$$

Truth Table for exclusive-OR Gate

Input		Output
A	B	$Y = A\bar{B} + B\bar{A}$
0	0	0
0	1	1
1	0	1
1	1	0

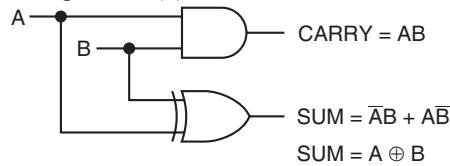
Thus output is 1 only when one of the two inputs is 1 but not both. Thus it *produces an output when the two inputs are different*.

## ARITHMETIC BUILDING BLOCKS

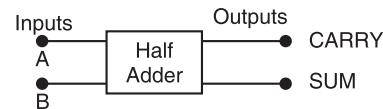
### 41.14 HALF-ADDER

- When we add two binary numbers, we start with the least-significant column. This means that we have to add two bits with the possibility of a carry. The circuit used for this is called a *half-adder*.
- A logic circuit that adds two bits producing a sum and a carry to be used in the next higher position is called a *half-adder*.

Fig. 41.35 (a) shows how to build a half-adder.



**Fig. 41.35(a)**



**Fig. 41.35(b)**

- The circuit consists of an exclusive-OR gate and an AND gate.
- The output of the exclusive OR gate is called the *SUM*.
- The output of the AND gate is called the *CARRY*.

Fig. 41.35 (b) shows the symbol of a half-adder.

The two inputs *A* and *B* represent the bits to be added.

*CARRY* and *SUM* are the two outputs.

*CARRY* bit is the output of AND gate.

**SUM bit** represents the output of **exclusive-OR gate**.

With two inputs *A* and *B*, there are four distinct cases.

(I) When *A* = 0 and *B* = 0,

$$\text{CARRY} = AB = 0$$

$$\text{SUM} = A \oplus B = 0 \oplus 0 = 0.$$

(II) When *A* = 0 and *B* = 1,

$$\text{CARRY} = AB = 0.1 = 0$$

$$\text{SUM} = A \oplus B = 0 \oplus 1 = 1.$$

(III) When *A* = 1 and *B* = 0,

$$\text{CARRY} = AB = 1.0 = 0$$

$$\text{SUM} = A \oplus B = 1 \oplus 0 = 1.$$

(IV) When *A* = 1 and *B* = 1,

$$\text{CARRY} = AB = 1.1 = 1$$

$$\text{SUM} = A \oplus B = 1 \oplus 1 = 0.$$

The truth Table for the Half-Adder operation is given in Table 1.

**Table 1. Half-adder Truth Table**

<i>Input</i>		<i>Output</i>	
<i>A</i>	<i>B</i>	<i>CARRY</i>	<i>SUM</i>
0	0	0	0

0	1	0	1
1	0	0	1
1	1	1	0

The **CARRY** column represents the AND operation.

The **SUM** column represents XOR operation.

- The AND gate produces a high output only when both inputs are high.
- The exclusive-OR gate produces a high output if either input, but not both, is high.

The Boolean expressions for half-adder operation is written as,

$$SUM = A \oplus B$$

$$CARRY = AB.$$

The logic circuit of Fig. 41.37 does electronically what we do mentally when we add 2 bits.

### 41.15 FULL ADDER

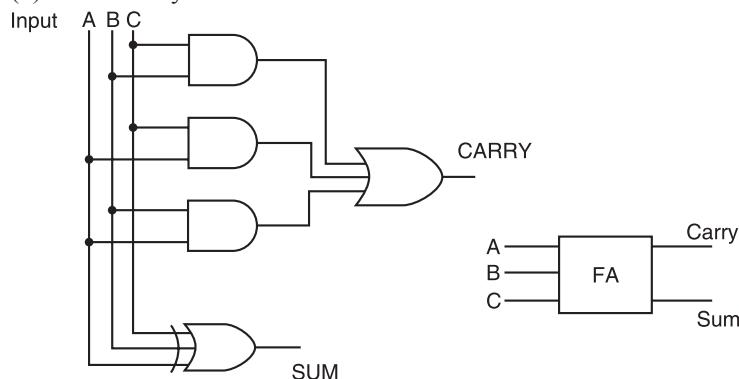
We define full-adder as a logic circuit that adds three bits-two bits to be added and a carry bit from previous addition, which results in a sum and a carry.

- A full Adder is a logic circuit that can add 3 bits at a time. The third bit is the carry from a lower column.
- The full-adder accepts two input bits and an input carry and generates a sum output and an output carry.

This implies that we need a logic circuit with three inputs and two outputs.

Fig. 41.36 (a) shows the simplest form of such a logic circuit.

Fig. 41.36 (b) shows the symbol of a full-adder.



(a) Logic circuit

(b) Equivalent symbol

Fig. 41.36

Table 2 shows the truth table of a full-adder.

Table 2. Full-Adder Truth Table

A	B	C	CARRY	SUM
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1

0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

You can easily check this truth table for its validity.

- CARRY is high in Fig. 41.36(a), when two or more of the  $ABC$  inputs are high. This agrees with the CARRY column in Table 2.
  - When an odd number of high  $ABC$  inputs drives the exclusive-OR gate, it produces a high output. This verifies the SUM column of the truth table.
  - When you examine each entry in Table 2, you can see that a full-adder performs binary addition on 3 bits.
  - The Boolean equations are

$$\text{CARRY} = AB + BC + AC \quad \text{and} \quad \text{SUM} = A \oplus B \oplus C.$$

A general representation of full-adder which adds  $i$ -th bit  $A_i$  and  $B_i$  of two numbers  $A$  and  $B$  and takes carry from  $(i-1)$ th bit is

$C_i = A_i B_i + B_i C_{i-1} + A_i C_{i-1}$  or  $C_i = A_i B_i + (A_i + B_i) C_{i-1}$  and  $S_i = A_i \oplus B_i \oplus C_{i-1}$

Here,  $C_i$  and  $S_i$  are carry and sum bits generated from the full adder.

- The second representation of  $C_i$  has an interesting meaning.

The first term gives, if both  $A_i$  and  $B_i$  are 1 then  $C_i = 1$ .

The second term gives if any of  $A_i$  or  $B_i$  is 1 and if there is carry from previous stage, i.e.,  $C_{i-1} = 1$  then also  $C_i = 1$ .

That this is the case, you can verify from full adder truth table.

- The circuit of Fig. 41.36 does electronically what we do mentally when we add 3 bits

Consider an example:

$$\begin{array}{r}
 \begin{array}{ccccc} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ + & 0 & 1 & 1 & 1 \\ \hline 1 & 0 & 1 & 0 & 1 & 1 \end{array} & \leftarrow CARRY \\
 & (28)_{10} + (15)_{10} = (43)_{10} \\
 & (11100)_2 + (01111)_2 = (101011)_2 \\
 & \leftarrow SUM
 \end{array}$$

## 41.16 SUBTRACTOR

- When we subtract one number from another, the number that is subtracted is called the *subtrahend*. The number from which it is subtracted is called the *minuend*.

### (i) Half Subtractor

**A logic circuit for the subtraction of B (subtrahend) from A (minuend) where A and B are one bit numbers** is called **half subtractor**. It can subtract only two binary digits at a time and produce an output of a **difference** and a **borrow**.

- A logic circuit which subtracts a one bit number from another one bit number is called a half-subtractor.
  - It can be realized using logic gates.

Fig. 41.37 shows the logic circuit for half subtractor.

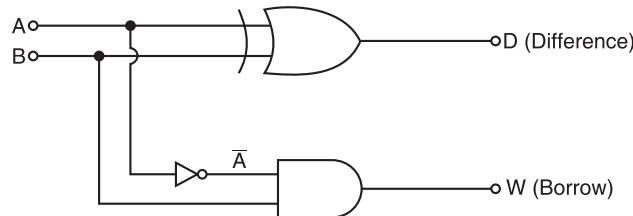


Fig. 41.37

- The input  $A$  is the minuend and the input  $B$  is the subtrahend.
- The output  $D$  of the XOR gate is the difference.

$$D = A \oplus B$$

$$D = \bar{A}B + A\bar{B}$$

- The output of the AND gate is the borrow.

$$W = \bar{A}B$$

- Fig. 41.38 shows the symbol for half subtractor.

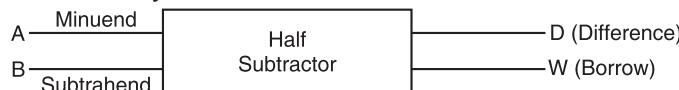


Fig. 41.38

Table 3 shows the truth table of the half-subtractor.

Table 3. Truth Table

Inputs		Outputs	
$A$	$B$	<i>Difference D</i>	<i>Borrow W</i>
0	0	0	0
0	1	1	1
1	0	1	0
1	1	0	0

**Illustration.** Let us say we subtract 1 from  $(10)_2$  i.e.,  $(2)_{10} - (1)_{10}$ .

$$\begin{array}{r} 1 \\ \text{---} \\ (-) \quad 0 \end{array} \quad \begin{array}{l} 0 \text{ (minuend)} \\ 1 \text{ (subtrahend)} \end{array}$$

$0 - 1$  is not possible. Hence it needs a borrow from the next (left) column. After borrowing 1 from the next column, the minuend and subtrahend can be written as

$$\begin{array}{r} \text{borrow} \rightarrow \quad 1 \\ \begin{array}{r} 0 \\ 0 \end{array} \quad (10)_2 \\ \hline \begin{array}{r} (-) \\ 0 \end{array} \quad \begin{array}{r} 1 \\ 1 \end{array} \end{array}$$

Hence  $(10)_2 - (01)_2 = (01)_2$  which is  $(2)_{10} - (1)_{10} = (1)_{10}$ .

### (ii) Full Subtractor

- A full subtractor is used for performing multibit subtraction wherein a borrow from the previous bit position may also be there.
- A digital circuit that accepts two one-bit signals and a borrow-in as input and produces their difference and borrow at output is called a full subtractor.

- The three inputs are  $A$  (minuend),  $B$  (subtrahend) and  $W_{in}$  (borrow from the previous stage).
- The two outputs are  $D$  (Difference) and  $W$  (borrow).

**Truth Table:** Table 4 gives the truth table of the device.

Table 4. Truth Table

Inputs			Output	
$A$	$B$	$W_{in}$	$D$	$W$
0	0	0	0	0
0	0	1	1	1
0	1	0	1	1
0	1	1	0	1
1	0	0	1	0
1	0	1	0	0
1	1	0	0	0
1	1	1	1	1

- Fig. 41.39 shows the symbol for Full subtractor.

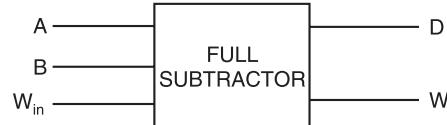


Fig. 41.39

- Fig. 41.40 shows the logic circuit for the entire Full subtractor.

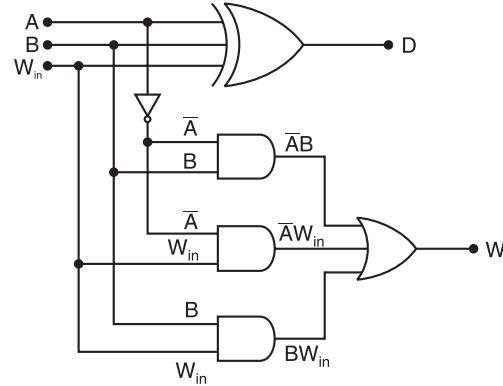


Fig. 41.40

#### Boolean Equation (Optional reading)

- (i) The difference ( $D$ ) of a full subtractor can be written in the form Boolean equation from the truth table as

$$D = \bar{A}\bar{B}W_{in} + \bar{A}B\bar{W}_{in} + A\bar{B}\bar{W}_{in} + ABW_{in}$$

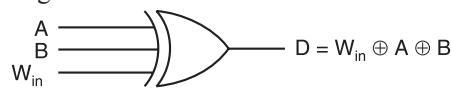
$$D = W_{in}(\bar{A}\bar{B} + AB) + \bar{W}_{in}(\bar{A}B + A\bar{B})$$

$$D = W_{in}\bar{X} + \bar{W}_{in}X \quad \text{where } X = A \oplus B$$

$$D = W_{in} \oplus X$$

$$D = W_{in} \oplus A \oplus B$$

Fig. 41.41 show the logic circuit for  $D$ .

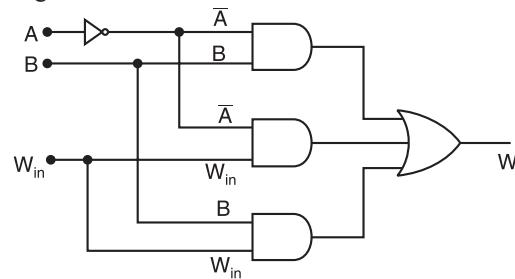


**Fig. 41.41**

(ii) The borrow ( $W$ ) can be written in the form of Boolean equation as

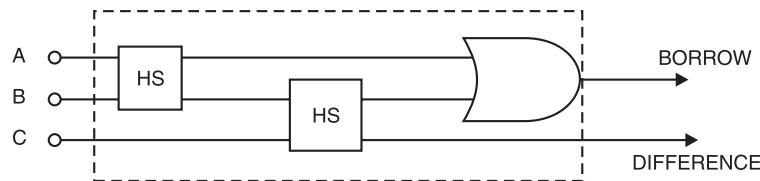
$$\begin{aligned}
 W &= \bar{A}\bar{B}W_{in} + \bar{A}B\bar{W}_{in} + \bar{A}BW_{in} + ABW_{in} \\
 &= \bar{A}\bar{B}W_{in} + \bar{A}B\bar{W}_{in} + BW_{in}(\bar{A} + A) \\
 &= \bar{A}\bar{B}W_{in} + \bar{A}B\bar{W}_{in} + BW_{in} \quad (\because \bar{A} + A = 1) \\
 &= \bar{A}\bar{B}W_{in} + \bar{A}B\bar{W}_{in} + BW_{in}(\bar{A} + 1) \quad (\because \bar{A} + 1 = 1) \\
 &= \bar{A}\bar{B}W_{in} + \bar{A}B\bar{W}_{in} + \bar{A}BW_{in} + BW_{in} \\
 &= \bar{A}\bar{B}W_{in} + \bar{A}B + BW_{in} \quad (\because \bar{W}_{in} + W_{in} = 1) \\
 &= \bar{A}\bar{B}W_{in} + \bar{A}B + BW_{in}(\bar{A} + 1) \quad (\because \bar{A} + 1 = 1) \\
 &= \bar{A}\bar{B}W_{in} + \bar{A}B + \bar{A}BW_{in} + BW_{in} \\
 &= \bar{A}W_{in}(\bar{B} + B) + \bar{A}B + BW_{in} \\
 W &= \bar{A}W_{in} + \bar{A}B + BW_{in} \quad (\because \bar{B} + B = 1)
 \end{aligned}$$

Fig. 41.42 shows the logic circuit for the borrow.



**Fig. 41.42**

- A full-subtractor can be constructed using two half-subtractors and an OR gate as shown in Fig. 41.43. The rectangular blocks marked HS represent half subtractors.



**Fig. 41.43**

### EXERCISE

- What is binary number system? How does it differ from decimal number system? Why is the binary number system used in computers?
- What are binary numbers? Explain with illustrations the methods of conversion from decimal to binary and binary to decimal numbers.
- Write four basic rules for adding binary digits. Give the truth table for binary addition.

4. State the rules for binary subtraction. Explain 1's complement and 2's complement method with examples.
5. Write short notes on : (i) Octal numbers (ii) Hexadecimal numbers and (iii) The ASCII code.  
(B.U. 2015)
6. Explain BCD code with examples. Mention one advantage and one disadvantage of coding a decimal number in BCD as compared to straight binary.  
(Bangalore, 2003)
7. What is meant by multiplexer? Explain a 4 to 1 multiplexer.  
(B.U. 2010)
8. What is a Demultiplexer? Draw the logic diagram of a 1 to 4 line demultiplexer and explain its operation.  
(Bangalore, 2003)
9. What is Boolean algebra? Mention its unique feature. Discuss the fundamental laws of Boolean algebra.  
[Karnatak University, Dharwad, Oct. 93]
10. Write and explain the Boolean expressions for OR, AND and NOT circuits.
11. State and prove De Morgan's theorems.  
(B.U. 2015)
12. (a) Draw logic symbols and truth tables of AND, OR, NOT logic gates. Explain its operation for two-input condition.  
(b) Draw the circuits of two-input AND, OR gates. Explain these two circuits.  
[April 95, KUD]
13. Explain the operation of a NOT circuit. Give its truth table and logic symbol.
14. Why is 'NAND' gate called 'universal building block'? Explain with example.  
(Madras, 1993)
15. Show how AND, OR and NOT gates can be realized from NAND gate.
16. Why is a NOR gate called building block? Explain with neat circuits.
17. Explain EX - OR gate and give its truth-table. Draw the logic diagrams of the EX - OR gate. Describe the operation of the gate.  
(Madras, 1992)
18. Draw the logic diagram of a half adder and write its truth table.  
(Osmania 93)
19. Explain the working of full adder.  
(Osmania 95)
20. Describe the construction and function of a half subtractor.  
(B.U. 2010)
21. Draw the logic circuit diagram of a full subtractor. Find out its outputs and draw the truth table.  
(B.U. 2015)

**PART - IX**

# **THEORETICAL PHYSICS**

**42. Classical Mechanics**

---

**43. Statistical Mechanics**

---

**44. Mathematical Physics**

---

## CLASSICAL MECHANICS

**After reading this chapter, you should be able to**

- ◆ Develop many of the mathematical techniques necessary for quantum mechanics
- ◆ Explain conservation theorems concerning with linear momentum, angular momentum and energy
- ◆ Derive Lagrange's equations from D'Alembert's principle
- ◆ Apply Lagrangian formalism to simple mechanical systems
- ◆ Develop Hamiltonian formulation of mechanics and discuss practical applications to mechanical problems

## 42.1 MECHANICS OF A PARTICLE

We shall study the conservation laws for a particle in motion using Newtonian mechanics.

### 1. Conservation of linear momentum:

Newton's second law of motion is

$$\frac{d}{dt}(mv) = \frac{d\mathbf{p}}{dt} = \mathbf{F} \quad \dots(1)$$

If the total force  $\mathbf{F}$  is zero, then  $\frac{d\mathbf{p}}{dt} = 0$  and the linear momentum is conserved.

If  $\mathbf{F}^{\text{ext}} = 0$ ,  $\frac{d\mathbf{P}}{dt} = 0$ . Integrating,  $\mathbf{P} = \text{constant}$ .

This gives the *theorem* for conservation of linear momentum of a particle.

### 2. Conservation of angular momentum:

Consider a particle of mass  $m$  and linear momentum  $\mathbf{p}$  at a position  $\mathbf{r}$  relative to origin  $O$  of an inertial reference frame (Fig. 42.1).

- The angular momentum  $\mathbf{L}$  of the particle with respect to the origin  $O$  is

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}, \quad \dots(1)$$

Let  $\mathbf{F}$  be the force acting on the particle. Then the torque  $\vec{\tau}$  acting on the particle with respect to the origin  $O$  is

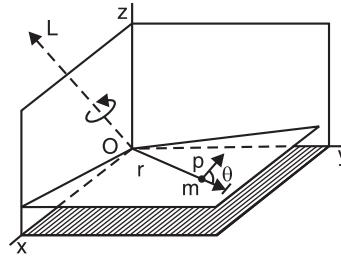


Fig. 42.1

$$\begin{aligned} \vec{\tau} &= \mathbf{r} \times \mathbf{F} \\ \vec{\tau} &= \mathbf{r} \times \mathbf{F} = \mathbf{r} \times \frac{d\mathbf{p}}{dt} \\ &= \frac{d}{dt}(\mathbf{r} \times \mathbf{p}) - \frac{d\mathbf{r}}{dt} \times \mathbf{p} \\ &= \frac{d}{dt}(\mathbf{r} \times \mathbf{p}) - \mathbf{v} \times m\mathbf{v}. \end{aligned} \quad \dots(2)$$

The second term is zero, as both vectors are parallel.

$$\begin{aligned} \therefore \vec{\tau} &= \frac{d}{dt}(\mathbf{r} \times \mathbf{p}) \\ \vec{\tau} &= \frac{d\mathbf{L}}{dt}. \end{aligned} \quad \dots(3)$$

Thus, time rate of change of the vector angular momentum of a particle is equal to the vector torque acting on it.

If  $\vec{\tau}_{\text{ext}} = 0$ , then  $\frac{d\mathbf{L}}{dt} = 0$ .  $\therefore \vec{\mathbf{L}} = \text{constant}$ .

Thus angular momentum is conserved in the absence of an external torque.

This is the principle of conservation of angular momentum.

### 3. Conservation of Energy:

If the forces acting on a particle are conservative, then the total energy of the particle, which is the sum of kinetic energy and potential energy, is constant or conserved.

Let the particle move from the point 1 to point 2 by the action of external force  $\vec{F}$  (Fig. 42.2). The total work done in displacing the particle from point 1 to point 2 is given by

$$\begin{aligned}
 W_{12} &= \int_1^2 dW = \int_1^2 \mathbf{F} \cdot d\mathbf{r} \\
 &= \int_1^2 \frac{d\mathbf{p}}{dt} \cdot d\mathbf{r} \\
 &= \int_1^2 \frac{d}{dt} (m\mathbf{v}) \cdot d\mathbf{r} \\
 &= m \int_1^2 \frac{d\mathbf{v}}{dt} \cdot \frac{d\mathbf{r}}{dt} dt \quad (\text{because } m \text{ is constant}) \\
 &= m \int_1^2 \frac{d\mathbf{v}}{dt} \cdot \mathbf{v} dt \\
 &= m \int_1^2 \frac{1}{2} \frac{d}{dt} (v^2) dt
 \end{aligned} \tag{1}$$

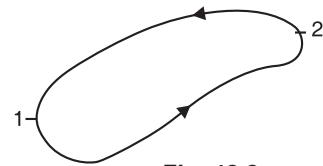


Fig. 42.2

or

$$\begin{aligned}
 W_{12} &= \frac{1}{2} m [v^2]_1^2 \\
 &= \frac{1}{2} m (v_2^2 - v_1^2) = \frac{1}{2} mv_2^2 - \frac{1}{2} mv_1^2
 \end{aligned}$$

Here,  $v_1$  and  $v_2$  are the velocities of the particle at points 1 and 2 respectively.

$$\therefore W_{12} = T_2 - T_1 \tag{2}$$

Here,  $T_1 = \frac{1}{2} m_1 v_1^2$  = kinetic energy of the particle at point 1,

and  $T_2 = \frac{1}{2} m_2 v_2^2$  = kinetic energy of the particle at point 2.

Thus, *the total work done by a force acting on a particle is equal to the change in the kinetic energy of the particle.*

This is called the *Work - Energy theorem*.

A conservative force  $\mathbf{F}$  can be expressed as the gradient of a scalar function called the potential function,

$$\mathbf{F} = -\nabla V \tag{3}$$

Here,  $V$  is called the potential or potential energy.

$$\begin{aligned}
 \text{We can write} \quad W_{12} &= \int_1^2 -\nabla V \cdot d\mathbf{r} \\
 &= \int_1^2 -\frac{dV}{dr} dr = - \int_1^2 dV \\
 \therefore W_{12} &= V_1 - V_2. \tag{4}
 \end{aligned}$$

From Eqs. (2) and (4) we get

$$\begin{aligned}
 T_2 - T_1 &= V_1 - V_2 \\
 \text{or} \quad T_2 + V_2 &= T_1 + V_1 = \text{constant}, \\
 \text{or in general} \quad T + V &= \text{constant},
 \end{aligned}$$

$$T + V = \text{constant},$$

which shows that the total energy of the particle is conserved.

This is the *energy conservation theorem*.

### Conservative Forces

**First Definition:** A force acting on a particle is conservative if the particle, after going through a complete round trip, returns to its initial position with the same kinetic energy as it had initially.

**Explanation:** Suppose we throw a ball upward against gravity. The ball reaches a certain height coming momentarily to rest so that its kinetic energy becomes zero. Then it returns to our hand under gravity with the same kinetic energy with which it was thrown. We assume the air-resistance is zero. Thus the force of gravity is conservative.

**Examples of Conservative Forces.** (i) Gravitational force (ii) Electrostatic force (iii) Elastic force.

*All central forces are conservative forces.*

**Second Definition:** A force acting on a particle is conservative if the net work done by the force in a complete round trip of the particle is zero.

**Explanation:** Suppose we throw a ball upward against gravity. When the ball is thrown up, the work done by the conservative force of gravity is negative. When the ball returns back, the work is positive. We assume that air-resistance is absent. So the negative and positive works are equal. Hence the net work done is zero.

If the force  $\mathbf{F}$  is conservative, then the work done by it around a closed path is zero, i.e.,

$$\oint \mathbf{F} \cdot d\mathbf{r} = 0 \quad \dots(1)$$

Physically it is clear that a system cannot be conservative if friction or other dissipative forces are present, for  $\mathbf{F} \cdot d\mathbf{r}$  due to friction is always positive and the integral cannot vanish.

According to Stokes theorem,

$$\oint \mathbf{F} \cdot d\mathbf{r} = \iint_S \text{curl } \mathbf{F} \cdot d\mathbf{S} = 0 \quad [\text{from Eq. (1)}]$$

or  $\iint_S \text{curl } \mathbf{F} \cdot d\mathbf{S} = 0$

or  $\text{curl } \mathbf{F} = 0$   
or  $\nabla \times \mathbf{F} = 0$

Therefore, for conservative forces  $\nabla \times \mathbf{F} = 0$

But curl of a gradient is always zero.

Therefore  $\mathbf{F}$  can be expressed as the gradient of a scalar function called the potential function,

$$\mathbf{F} = -\nabla V \quad \dots(2)$$

Here,  $V$  is called the potential or potential energy.

## 42.2 MECHANICS OF A SYSTEM OF PARTICLES

When the mechanical system consists of two or more particles, we must distinguish, between the *external forces* exerted upon the particles of the system by sources not belonging to the system, and the *internal forces* arising on account of the interactions between the particles of the system themselves.

The equation of motion in terms of Newton's second law for a general system of  $N$  particles is

$$m_i \mathbf{a}_i = \dot{\mathbf{p}}_i = \mathbf{F}_i^{(e)} + \sum_{j \neq i} \mathbf{F}_i^j, \quad i = 1, 2, \dots, N. \quad \dots(1)$$

From Eq. (1), we obtain equation of motion of each particle,  $i$ , corresponding to each value of  $i$ . They are  $N$  equations in all. Here,  $\mathbf{F}_i^{(e)}$  stands for the external force acting on  $i^{\text{th}}$  particle and  $\mathbf{F}_i^j$  is the *internal force on the  $i^{\text{th}}$  particle due to  $j^{\text{th}}$  particle*.

All the particles of the system exert forces on one another. Hence the internal force on  $i^{\text{th}}$  particle must be the sum of forces due to all other particles =  $\sum_{j=1}^N F_i^j$  excluding the term  $j = i$ , since by definition  $\mathbf{F}_i^j$  is obviously zero.

We shall modify Eqs. (1) by assuming that Newton's third law is valid for internal forces. That is, the force  $\mathbf{F}_i^j$  must be equal and opposite in direction to the force  $\mathbf{F}_j^i$  that the  $i^{\text{th}}$  particle exerts on the  $j^{\text{th}}$  particle. Vectorially

$$\mathbf{F}_i^j = -\mathbf{F}_j^i \quad \dots(2)$$

It automatically implies that internal forces occur in pairs and act along line joining the two particles. Any combination of mutual forces must be zero then.

Summing now over all particles of the system, we obtain the equation of motion of the system as a whole:

$$\begin{aligned} \sum \dot{\mathbf{p}}_i &= \frac{d^2}{dt^2} \sum_i m_i \mathbf{r}_i \\ &= \sum_i \mathbf{F}_i^{(e)} + \sum'_{i,j} \mathbf{F}_i^j \\ &= \sum_i \mathbf{F}_i^{(e)}, \end{aligned} \quad \dots(3)$$

since

$$\begin{aligned} \sum'_{i,j} \mathbf{F}_i^j &= -\sum'_{i,j} \mathbf{F}_j^i \\ &= \frac{1}{2} \sum'_{i,j} [\mathbf{F}_i^j + \mathbf{F}_j^i] = 0 \end{aligned} \quad [\text{from Eq. (2)}]$$

Here, a prime on the summation symbol  $\Sigma$  means that the term  $j = i$  is to be excluded from the sum.

We define the centre of mass  $\mathbf{R}$  of the system by

$$\mathbf{R} = \frac{\sum_i m_i \mathbf{r}_i}{\sum_i m_i} = \frac{\sum_i m_i \mathbf{r}_i}{M}.$$

Here,  $\sum_i m_i = M$  is the total mass of the system

(Fig. 42.3). Eq. (3) becomes

$$\begin{aligned} M \frac{d^2 \mathbf{R}}{dt^2} &= \sum_i \mathbf{F}_i^{(e)} \\ M \frac{d^2 \mathbf{R}}{dt^2} &= \mathbf{F}^{(e)} \end{aligned} \quad \dots(4)$$

The total linear momentum of the system is expressed as

$$\begin{aligned} \mathbf{P} &= \sum_i m_i \dot{\mathbf{r}}_i \\ &= \frac{d}{dt} \sum_i m_i \mathbf{r}_i \end{aligned}$$

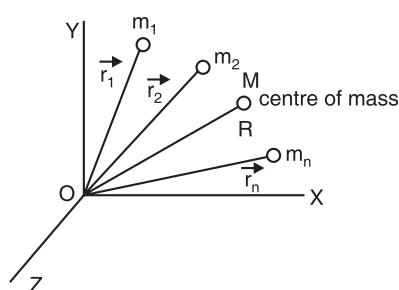


Fig. 42.3

$$\mathbf{P} = M \dot{\mathbf{R}} \quad \dots(5)$$

From Eq. (5), rate of change of total linear momentum is

$$\dot{\mathbf{P}} = M \ddot{\mathbf{R}} = \mathbf{F}^{(e)}. \quad \dots(6)$$

Eq. (6) defines two important characteristics of motion. They are:

(i) *Centre of mass moves as if the total external force  $\mathbf{F}^{(e)}$  acting on the entire mass of the system were concentrated at the centre of mass.*

(ii) *If the total external force vanishes, the total linear momentum is conserved.*

Property (ii) is the *theorem of conservation of linear momentum for a system of particles*. It also implies that since  $\dot{\mathbf{P}} = 0$ ,  $\mathbf{P} = \text{constant}$  or  $\dot{\mathbf{R}} = \text{constant}$ , i.e., the centre of mass moves with constant velocity in the absence of external forces.

Thus, we may state that *the velocity of the centre of mass of the system remains constant if there are no external forces acting on the system.*

**Example.** Consider the uniform motion of a radioactive nucleus undergoing disintegration (Fig. 42.4). The nucleus ejects different particles which move off in different directions in such a way that their centre of mass continues to move with constant velocity even after the disintegration.

**(a) Conservation theorem for linear momentum:**

The net linear momentum of a system of  $n$ -particles is

$$\mathbf{P} = \sum_{i=1}^n \mathbf{p}_i = \sum_{i=1}^n m_i \mathbf{v}_i$$

From Newton's second law,  $\mathbf{F}^{ext} = \frac{d\mathbf{P}}{dt}$

i.e., the rate of change of linear momentum of a system of particles is equal to the net external force acting on the system.

If  $\mathbf{F}^{ext} = 0$ ,  $\frac{d\mathbf{P}}{dt} = 0$ . Integrating,  $\mathbf{P} = \text{constant}$ .

This gives the *theorem for conservation of linear momentum of the system*.

**Statement:** “*If the sum of external forces acting on the system of particles is zero, the total linear momentum of the system is constant or conserved.*”

**(b) Conservation theorem for angular momentum.**

The angular momentum of  $i^{\text{th}}$  particle of the system about any point  $O$ , from definition is given by

$$\mathbf{L}_i = \mathbf{r}_i \times \mathbf{p}_i, \quad \dots(1)$$

Here,  $\mathbf{r}_i$  is the radius vector of  $i^{\text{th}}$  particle from the point  $O$  and  $\mathbf{p}_i$  its linear momentum (Fig. 42.5).

We obtain the total angular momentum of the system of particles by forming the cross product  $(\mathbf{r}_i \times \mathbf{p}_i)$  for the  $i^{\text{th}}$  particle and summing over all particles.

$$\mathbf{L} = \sum_i (\mathbf{r}_i \times \mathbf{p}_i) \quad \dots(2)$$

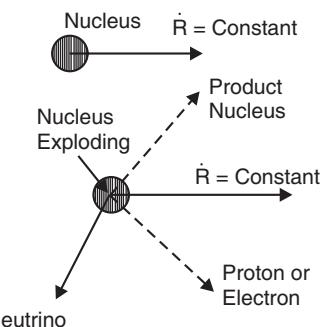


Fig. 42.4

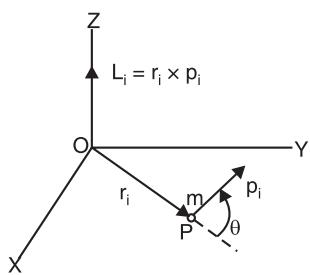


Fig. 42.5

$$\begin{aligned}
 \dot{\mathbf{L}} &= \frac{d}{dt} \sum_i (\mathbf{r}_i \times \mathbf{p}_i) = \sum_i [\mathbf{r}_i \times \dot{\mathbf{p}}_i + \dot{\mathbf{r}}_i \times \mathbf{p}_i] \\
 &= \sum_i (\mathbf{r}_i \times \dot{\mathbf{p}}_i) \quad (\text{because } \dot{\mathbf{r}}_i \times \mathbf{p}_i = 0) \\
 &= \sum_i \mathbf{r}_i \times \mathbf{F}_i^{(e)} + \sum_{i,j}' \mathbf{r}_i \times \mathbf{F}_i^j. \quad \dots(3)
 \end{aligned}$$

Second term in Eq. (3) denotes the sum of internal torques which vanishes if the interacting forces are Newtonian in character. We then have the important result

$$\frac{d\mathbf{L}}{dt} = \sum_i \mathbf{r}_i \times \mathbf{F}_i^{(e)} = \tau^{(e)} = \text{sum of external torques.} \quad \dots(4)$$

The time derivative of the total angular momentum is equal to the moment of the external forces about the given point. From this we have the conservation of angular momentum of a system of particles.

*If total external torque  $\tau^{(e)} = 0$ ,  $\mathbf{L}$  is constant in time or conserved.*

*Thus, if external torque acting on a system of particles is zero, the angular momentum of the system remains constant.*

This is the conservation theorem for angular momentum of a system of particles.

**EXAMPLE 1.** Express angular momentum of the system as the sum of angular momentum of motion of the centre of mass and angular momentum of the motion about the centre of mass.

**SOL.** Let  $\mathbf{r}_i$  be the position vector of the  $i^{\text{th}}$  particle relative to a point  $O$  fixed in an inertial frame (Fig. 42.6).

From Fig.,  $\mathbf{r}_i = \mathbf{r}'_i + \mathbf{R}$  and  $\mathbf{v}_i = \mathbf{v}'_i + \mathbf{v}$   $\dots(1)$

•  $\mathbf{r}'_i$  and  $\mathbf{v}'_i$  denote the radius vector and velocity of the  $i^{\text{th}}$  particle referred to centre of mass  $O'$  as the new origin and

$\mathbf{v} = \dot{\mathbf{R}}$  is the velocity of the centre of mass relative to  $O$ .

Then  $\mathbf{L} = \sum m_i (\mathbf{r}'_i + \mathbf{R}) \times (\mathbf{v}'_i + \mathbf{v})$

$$= \sum_i (\mathbf{R} \times m_i \mathbf{v}) + \sum_i \mathbf{r}'_i \times m_i \mathbf{v}'_i + \left( \sum_i m_i \mathbf{r}'_i \right) \times \mathbf{v} + \mathbf{R} \times \frac{d}{dt} \left( \sum_i m_i \mathbf{r}'_i \right). \quad \dots(2)$$

But  $\sum_i m_i \mathbf{r}'_i = 0$ , from the definition of centre of mass.

So the last two terms in Eq. (2) vanish.

Also  $\sum_i (\mathbf{R} \times m_i \mathbf{v}) = \mathbf{R} \times M\mathbf{v}$ .

$$\therefore \mathbf{L} = \mathbf{R} \times M\mathbf{v} + \sum_i \mathbf{r}'_i \times \mathbf{p}'_i \quad \dots(3)$$

Eq. (3) shows that we can decompose angular momentum of a system of particles, with reference to a given origin  $O$ , into two distinct parts:

(i) angular momentum of the system about the origin as if the total mass were concentrated at the centre of mass.

(ii) plus the angular momentum of the motion about the centre of mass.

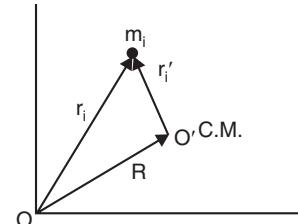


Fig. 42.6

**(c) Conservation of Energy for a System of Particles****(i) Work Energy Theorem**

Consider a system of particles. Let  $W_i$  be the work done by the external force  $\mathbf{F}_i$  acting on  $i^{\text{th}}$  particle in displacing it from position 1 to 2. Then,

$$W_i = \int_1^2 \mathbf{F}_i \cdot d\mathbf{r}_i = \left[ \frac{1}{2} m_i v_i^2 \right]_1^2 = [T_i]_1^2$$

Here,  $T_i$  = kinetic energy of  $i^{\text{th}}$  particle.

Now, summing over all the particles of the system, we have

$$\begin{aligned} \sum W_i &= \sum [T_i]_1^2 \\ W_{12} &= T_2 - T_1, \end{aligned} \quad \dots(1)$$

Here,  $W_{12}$  = total work done by the external force.

$T_1$  and  $T_2$  are the kinetic energies of the system at points 1 and 2 respectively.

Thus, *the work done is equal to the change in the kinetic energy*. This is the *work-energy theorem*.

**(ii) Conservation theorem for energy**

For a conservative system, the force  $\mathbf{F}_i$  acting on  $i^{\text{th}}$  particle is expressed as the gradient of some scalar function, i.e.,

$$\mathbf{F}_i = -\nabla V_i$$

Here,  $V_i$  is the potential or potential energy of  $i^{\text{th}}$  particle.

$$\text{Now, } \nabla V_i = \frac{\partial V_i}{\partial r_i}. \quad \therefore \quad \mathbf{F}_i = -\frac{\partial V_i}{\partial r_i}$$

Work done by the external force  $\mathbf{F}_i$  acting on the  $i^{\text{th}}$  particle in displacing it from position 1 to 2 is given by

$$\begin{aligned} W_i &= \int_1^2 \mathbf{F}_i \cdot d\mathbf{r}_i = - \int_1^2 \frac{\partial V_i}{\partial r_i} dr_i \\ &= -[V_i]_1^2 \end{aligned}$$

Now summing over all the particles of the system, we have

$$\begin{aligned} \sum W_i &= \sum_i [-V_i]_1^2 \\ \therefore \quad W_{12} &= V_1 - V_2 \end{aligned} \quad \dots(2)$$

Here,  $W_{12}$  = total work done by the external forces.

$V_1$  and  $V_2$  are the potential energies of the system at points 1 and 2 respectively.

Comparing Eqns. (1) and (2),  $T_2 - T_1 = V_1 - V_2$

or  $T_2 + V_2 = T_1 + V_1 = E$  = total energy of system.

This is the *energy conservation theorem*.

*If the forces acting on the system of particles are conservative, the total energy of the system of particles, which is the sum of the total kinetic energy and the total potential energy of the system is conserved.*

### 42.3 BASIC CONCEPTS

**Degrees of Freedom.** *The number of mutually independent variables required to define the state or position of a system is the number of degrees of freedom possessed by it.*

For example, the position of a simple ideal mass-point can be defined completely by the three cartesian coordinates  $x, y, z$ . So it has three degrees of freedom. Extending this idea, for a system of  $N$  particles moving independently of each other, the number of degrees of freedom is  $3N$ .

**Constraints.** *Constraints are restrictions imposed on the position or motion of a system, because of geometrical conditions.*

**Examples.** (1) The beads of an abacus are constrained to one-dimensional motion by the supporting wires.

(2) Gas molecules within a container are constrained by the walls of the vessel to move only *inside* the container.

(3) The motion of rigid bodies is always such that the distance between any two particles remains unchanged.

(4) A particle placed on the surface of a solid sphere is restricted by the constraint so that it can only move on the surface or in the region exterior to the sphere.

(5) The motion of point mass of a simple pendulum is restricted since the point mass always remains at a constant distance from the point of suspension.

For a particle constrained to move on a plane, only two variables  $x, y$  or  $r, \theta$  are sufficient to describe its motion and the particle is said to have two degrees of freedom. Thus, the constraint on the motion of the particle in a plane reduces the number of degrees of freedom by one.

Very often, we can express constraints in terms of certain equations. For example, the equation of constraint in the case of a particle moving on or outside the surface of a sphere of radius  $a$  is  $x^2 + y^2 + z^2 \geq a^2$  if the origin of the coordinate system coincides with the centre of the sphere.

#### Types of Constraints

##### (i) Holonomic and non-holonomic constraints.

If the constraints can be expressed as equations connecting the co-ordinates of the particles (and possibly time) in the form

$$f(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_n, t) = 0 \quad \dots(1)$$

then the constraints are said to be *holonomic*.

**Examples.** (1) *The constraints involved in the motion of rigid bodies* in which the distance between any two particular points is always fixed, are *holonomic* since the conditions of constraints are expressed as

$$(\mathbf{r}_i - \mathbf{r}_j)^2 - c_{ij}^2 = 0.$$

(2) *The constraints involved* when a particle is restricted to move along a curve or surface are holonomic. Here the equation defining the curve or surface is the equation of constraint.

(3) *The constraints involved* in the motion of the point mass of a simple pendulum are *holonomic*.

Let

$\mathbf{r}$  = position vector of the point mass

$\mathbf{a}$  = position vector of the point of suspension

$l$  = length of the string.

The equation of the constraint may be written as

$$(\mathbf{r} - \mathbf{a})^2 - l^2 = 0.$$

If the constraints cannot be expressed in the form of Eq. (1), they are called *non-holonomic constraints*.

**Examples.** (1) The constraints involved in the motion of the particle placed on the surface of a solid sphere are non-holonomic. The conditions of constraints in this case are expressed as

$$r^2 - a^2 \geq 0,$$

where  $a$  is the radius of sphere. This is an inequality and hence not in the form of Eq. (1).

(2) The walls of the gas vessel constitute a non-holonomic constraint.

(3) Another example of a non-holonomic constraint is the condition that the velocity of the point of contact of a sphere, rolling without slipping along a stationary rough surface, is equal to zero.

**(ii) Scleronomic and Rheonomic Constraints.** If the constraints are independent of time, they are called *scleronomic*.

The constraint in the case of rigid body motion is scleronomous. If the constraints are explicitly dependent on time, they are called *rheonomic*. A bead sliding on a moving wire is an example of rheonomic constraint.

**Definition of Constraint.** *A constraint is defined to be some geometrical restriction on the freedom of motion of the particles of the system, which might have to be satisfied by the co-ordinates or co-ordinate differences or sometimes by velocities rather than co-ordinates.*

The effect of having ' $k$ ' equations of constraints on the system, is that ' $k$ ' out of the original  $3N$  variables describing the system, become dependent rather than independent. Thus the number of independent variables is reduced to  $(3N-k)$  and the system is said to possess  $(3N-k)$  'degrees of freedom'.

In the solution of mechanical problems, the constraints introduce two types of difficulties:

(1) The co-ordinates  $r_i$  are connected by the equations of constraints. Therefore, they are not independent.

(2) The forces of constraint are not *a priori* known. In fact, they cannot be estimated till a complete solution of the problem is obtained.

The first problem can be solved by introducing *generalized co-ordinates*, whereas the second is practically an insurmountable problem. We therefore reformulate the problem such that the forces of constraint disappear.

### Generalised Co-ordinates

A system consisting of  $N$  particles, free from constraints, has  $3N$  independent co-ordinates or degrees of freedom. If the sum of the degrees of freedom of all the particles is  $k$ , then the system may be regarded as a collection of free particles subjected to  $(3N-k)$  independent constraints. So only  $k$  co-ordinates are needed to describe the motion of the system. These new co-ordinates  $q_1, q_2, q_3, \dots, q_k$  are called the *Generalised Co-ordinates of Lagrange*. Generalised co-ordinates may be lengths or angles or any other set of independent quantities which define the position of the system.

**Definition.** *The generalised co-ordinates of a material system are the independent parameters  $q_1, q_2, q_3, \dots, q_k$  which completely specify the configuration of the system, i.e., the position of all its particles with respect to the frame of reference.*

**Example 1.** Consider the simple pendulum of mass  $m_1$  with fixed length  $r_1$  (Fig. 42.7). The single co-ordinate  $\theta_1$  will determine uniquely the position of  $m_1$  since the simple pendulum is a system of one degree of freedom. Since the only variable involved is  $\theta_1$ , it can be chosen as the generalised co-ordinate. Thus  $q = \theta_1$ .

The two coordinates  $x_1$  and  $y_1$  could also be used to locate  $m_1$  but would require the inclusion of the *equation of the constraint*  $x_1^2 + y_1^2 = r_1^2$ . Since  $x_1$  and  $y_1$  are not independent, they are not generalised co-ordinates.

(2) When a particle moves in a plane, it may be described by cartesian co-ordinates  $x, y$  or the polar co-ordinates,  $r, \theta$ . We can write

$$q_1 = x \quad \text{or} \quad q_1 = r = \sqrt{(x^2 + y^2)},$$

$$q_2 = y \quad q_2 = \theta = \tan^{-1} \frac{y}{x}.$$

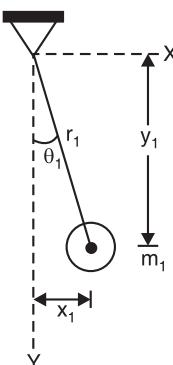


Fig. 42.7

(3) In considering the motion of a particle in a 'spherically symmetrical force field' or the motion of a particle constrained to move on a sphere of a fixed radius, the use of *spherical polar coordinates* ( $r, \theta, \phi$ ), is more convenient. Here,

$$\begin{aligned} q_1 &= r = (x^2 + y^2 + z^2)^{1/2}, \\ q_2 &= \theta = \cot^{-1} \frac{z}{(x^2 + y^2)^{1/2}} \\ q_3 &= \phi = \tan^{-1} \frac{y}{x}. \end{aligned}$$

(4) If it is preferred to accept a co-ordinate system moving uniformly with velocity  $v$  in  $x$ -direction, generalised co-ordinates are

$$\begin{aligned} q_1 &= x - \dot{x}t. \\ q_2 &= y \quad \dot{x} = v = \text{constant}. \\ q_3 &= z. \end{aligned}$$

#### Transformation Equations

The rectangular cartesian co-ordinates can be expressed as the functions of generalised co-ordinates. Let  $x_i, y_i$  and  $z_i$  be the cartesian co-ordinates of  $i^{\text{th}}$  particle of the system. Let  $t$  denote the time. Then, these cartesian co-ordinates can be expressed as functions of generalised co-ordinates  $q_1, q_2, q_3, \dots, q_k$  i.e.,

$$\left. \begin{aligned} x_i &= x_i(q_1, q_2, \dots, q_k, t) \\ y_i &= y_i(q_1, q_2, \dots, q_k, t) \\ z_i &= z_i(q_1, q_2, \dots, q_k, t) \end{aligned} \right\} \quad \dots(1)$$

Let  $\mathbf{r}_i$  be the position vector of  $i^{\text{th}}$  particle, i.e.,  $\mathbf{r}_i = \mathbf{i}x_i + \mathbf{j}y_i + \mathbf{k}z_i$  ... (2)

Then  $\mathbf{r}_i = \mathbf{r}_i(q_1, q_2, q_3, \dots, q_k, t)$ ,

Eq. (2) is the vector form of Eq. (1).

The equations like (1) and (2) are called *transformation equations*. The functions and their derivatives in the above two equations are supposed to be continuous. The equations also contain the constraints explicitly.

## 42.4 GENERALISED NOTATIONS

**(1) Generalised Displacement:** Consider a small displacement of an  $N$ -particle system defined by changes  $\delta\mathbf{r}_i$  in cartesian co-ordinates  $\mathbf{r}_i$  ( $i = 1, 2, \dots, N$ ) with time  $t$  held fixed.  $\mathbf{r}_i$  are functions of generalised co-ordinates defined by equation

$$\mathbf{r}_i = \mathbf{r}_i(q_1, q_2, \dots, q_{3N}, t).$$

An arbitrary virtual displacement  $\delta\mathbf{r}_i$  is written as

$$\delta\mathbf{r}_i = \sum_{j=1}^{3N} \frac{\partial \mathbf{r}_i}{\partial q_j} \delta q_j \quad (\text{as } \delta t = 0)$$

$\delta q_j$  are called the *generalised displacements* or virtual arbitrary displacements. If  $q_j$  is an angle co-ordinate,  $\delta q_j$  is an angular displacement.

**(2) Generalised Velocity:** The generalised velocities of a system are the total time derivatives of the generalised co-ordinates of the system.

$$\text{Thus} \quad \dot{q}_i = \frac{dq_i}{dt} \quad (i = 1, 2, 3, \dots, k).$$

**(3) Generalised Momentum:** The generalised momentum  $p_k$  associated with a generalised co-ordinate  $q_k$  is:

$$p_k = \frac{\partial T}{\partial \dot{q}_k}.$$

### Configuration Space

In the case of motion of a single particle we can represent its trajectory in the three dimensional space by specifying variables. For a system of  $N$  particles described by  $3N$  space coordinates with  $(3N - k)$  equations of constraint in the real space, it is difficult to visualise the motion of the entire system. It is convenient to describe the motion of a system having  $k$  coordinates in a hypothetical  $k$  dimensional space. The instantaneous configuration of the system is described by the values of the  $k$  generalised coordinates  $q_1, q_2, q_3, \dots, q_k$ , and corresponds to a particular point in a cartesian hyperspace where the  $q$ 's form the  $k$  coordinate axes. The point is called the system point and the  $k$  dimensional space is known as the *Configuration space*.

At some later instant, the state of the system changes and it will be represented by some other point in the configuration space. Thus, the system point moves in the configuration space tracing out a curve. This curve represents "the path of motion of the system". "The motion of the system", as used above, then refers to the motion of the system point along this path in *configuration space*. Time can be considered formally as a parameter of the curve since each point in the configuration space has one or more values of time associated with it.

*Configuration space has nothing in common with the three-dimensional space which we can visualise physically. It is a purely geometric structure by means of which the laws of the variation of the state of a system can be formulated in geometrical language.*

### Principle of Virtual Work

Consider a system described by  $n$  generalized coordinates  $q_j$  ( $j = 1, 2, 3, \dots, n$ ). Suppose the system undergoes a certain displacement in the configuration space in such a way that it does not take any time and that it is consistent with the constraints on the system. Such displacements are called *virtual* because they do not represent actual displacements of the system. Since there is no actual motion of the system, the work done by the forces of constraint in such a virtual displacement is zero.

Let the virtual displacement of the  $i^{\text{th}}$  particle of the given system be  $\delta \mathbf{r}_i$ . If the given system is in equilibrium, the resultant force acting on the  $i^{\text{th}}$  particle of the system must be zero, i.e.,  $\mathbf{F}_i = 0$ . It is, then, obvious that *virtual work*  $\mathbf{F}_i \cdot \delta \mathbf{r}_i = 0$  for the  $i^{\text{th}}$  particle and hence it is also zero for all the particles of the system.

$$\text{Thus } dW = \sum_i \mathbf{F}_i \cdot \delta \mathbf{r}_i = 0 \quad \dots(1)$$

The resultant force  $\mathbf{F}_i$  acting on the  $i^{\text{th}}$  particle is,

$$\mathbf{F}_i = \mathbf{F}_i^a + \mathbf{f}_i \quad \dots(2)$$

Here,  $\mathbf{F}_i^a$  is the applied force and  $\mathbf{f}_i$  is the force of constraint.

Eq. (1) then becomes

$$\sum_i \mathbf{F}_i^a \cdot \delta \mathbf{r}_i + \sum_i \mathbf{f}_i \cdot \delta \mathbf{r}_i = 0 \quad \dots(3)$$

We now consider systems for which the virtual work done by the forces of constraints is zero, i.e.,

$$\sum_i \mathbf{f}_i \cdot \delta \mathbf{r}_i = 0 \quad \dots(4)$$

Then Eq. (3) becomes

$$\sum_i \mathbf{F}_i^a \cdot \delta \mathbf{r}_i = 0 \quad \dots(5)$$

This equation is termed as *principle of virtual work*.

### D'Alembert's Principle

Most of the systems we come across in mechanics are not in static equilibrium. Hence the principle must be modified to include dynamic systems as well.

According to Newton's second law of motion,

$$\mathbf{F}_i = \dot{\mathbf{p}}_i \quad \text{or} \quad \mathbf{F}_i - \dot{\mathbf{p}}_i = 0 \quad \dots(6)$$

According to the above equation, a moving system of particles can be considered to be in equilibrium under the force  $(\mathbf{F}_i - \dot{\mathbf{p}}_i)$ , i.e., the actual applied force  $\mathbf{F}_i$  plus an additional force  $- \dot{\mathbf{p}}_i$  which is known as *reversed effective force* on  $i$ th particle. Let us again assume that *the forces of constraint do no work*. Then, we can generalize Eq. (5) by the use of Eq. (6) to the form

$$\sum_i (\mathbf{F}_i - \dot{\mathbf{p}}_i) \cdot \delta \mathbf{r}_i = 0 \quad \dots(7)$$

Eq. (7) is the mathematical statement of D'Alembert's principle.

## LAGRANGIAN FORMULATION OF MECHANICS

### The Lagrangian Function

The Lagrangian function  $L$  is the difference between the kinetic energy ( $T$ ) and potential energy ( $V$ ) of a system. It is a function of generalised coordinates, generalised velocities and time.

$$L(q, \dot{q}, t) = T(q, \dot{q}, t) - V(q, t)$$

The potential energy and Lagrangian function of a conservative system do not depend explicitly on time.

$$\therefore L(q, \dot{q}) = T(q, \dot{q}) - V(q) \text{ for a conservative system.}$$

## 42.5 DERIVATION OF LAGRANGE'S EQUATIONS OF MOTION

### (a) Lagrange's Equations from D'Alembert's Principle

Consider a system of particles whose position vectors are expressed as functions of generalized coordinates  $q_1, q_2, q_3, \dots, q_k, \dots, q_f$  and the time  $t$ .

Consider any particle of the system ( $i$ th particle) of mass  $m_i$  and acted upon by an external force  $\mathbf{F}_i$ .

According to D'Alembert's principle,

$$\sum_i (\mathbf{F}_i - \dot{\mathbf{p}}_i) \cdot \delta \mathbf{r}_i = 0 \quad \dots(1)$$

Here  $\dot{\mathbf{p}}_i$  is the inertial force for  $i$ th particle and  $\delta \mathbf{r}_i$  is the virtual displacement of  $i$ th particle due to action of force  $\mathbf{F}_i$ .

In general,

$$\mathbf{r}_i = \mathbf{r}_i(q_1, q_2, q_3, \dots, q_k, \dots, q_f, t) \quad \dots(2)$$

$$\begin{aligned} \dot{\mathbf{r}}_i &= \frac{\partial \mathbf{r}_i}{\partial t} = \frac{\partial \mathbf{r}_i}{\partial q_1} \dot{q}_1 + \frac{\partial \mathbf{r}_i}{\partial q_2} \dot{q}_2 + \dots + \frac{\partial \mathbf{r}_i}{\partial q_k} \dot{q}_k + \dots + \frac{\partial \mathbf{r}_i}{\partial q_f} \dot{q}_f + \frac{\partial \mathbf{r}_i}{\partial t} \\ &= \sum_k \frac{\partial \mathbf{r}_i}{\partial q_k} \dot{q}_k + \frac{\partial \mathbf{r}_i}{\partial t} \end{aligned} \quad \dots(3)$$

The virtual displacement  $\delta \mathbf{r}_i$  in terms of generalised co-ordinates is given by

$$\delta \mathbf{r}_i = \sum_k \frac{\partial \mathbf{r}_i}{\partial q_k} \delta q_k. \quad \dots(4)$$

Now,

$$\dot{\mathbf{p}}_i = m_i \ddot{\mathbf{r}}_i$$

Therefore Eq. (1) becomes  $\sum_i (\mathbf{F}_i - m_i \ddot{\mathbf{r}}_i) \cdot \delta \mathbf{r}_i = 0$   $\dots(5)$

or

$$\sum_i m_i \ddot{\mathbf{r}}_i \cdot \delta \mathbf{r}_i = \sum_i \mathbf{F}_i \cdot \delta \mathbf{r}_i \quad \dots(6)$$

Putting the value of  $\delta \mathbf{r}_i$  from Eq. (4) in (6), we get

$$\sum_i \sum_k m_i \ddot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \delta q_k = \sum_i \sum_k \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \delta q_k \quad \dots(7)$$

Now,

$$\frac{d}{dt} \left( \dot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \right) = \dot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} + \dot{\mathbf{r}}_i \cdot \frac{d}{dt} \left( \frac{\partial \mathbf{r}_i}{\partial q_k} \right),$$

$$\therefore \ddot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} = \frac{d}{dt} \left( \dot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \right) - \dot{\mathbf{r}}_i \cdot \frac{d}{dt} \left( \frac{\partial \mathbf{r}_i}{\partial q_k} \right) \quad \dots(8)$$

Putting the value of  $\ddot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k}$  from above equation in (7), we get

$$\sum_i \sum_k m_i \left[ \frac{d}{dt} \left( \dot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \right) - \dot{\mathbf{r}}_i \cdot \frac{d}{dt} \left( \frac{\partial \mathbf{r}_i}{\partial q_k} \right) \right] \delta q_k = \sum_i \sum_k \mathbf{F}_i \cdot \left( \frac{\partial \mathbf{r}_i}{\partial q_k} \right) \delta q_k \quad \dots(9)$$

Differentiating Eq. (3) partially with respect to  $\dot{q}_k$ ,

$$\frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_k} = \frac{\partial \mathbf{r}_i}{\partial q_k} \quad \dots(10)$$

Differentiating Eq. (3) partially with respect to  $q_k$ ,

$$\frac{\partial \dot{\mathbf{r}}_i}{\partial q_k} = \frac{\partial^2 \mathbf{r}_i}{\partial q_k \partial q_1} \dot{q}_1 + \frac{\partial^2 \mathbf{r}_i}{\partial q_k \partial q_2} \dot{q}_2 + \dots + \frac{\partial^2 \mathbf{r}_i}{\partial q_k \partial q_k} \dot{q}_k + \dots + \frac{\partial^2 \mathbf{r}_i}{\partial q_k \partial t} \quad \dots(11)$$

Also we have

$$\begin{aligned} \frac{d}{dt} \left( \frac{\partial \mathbf{r}_i}{\partial q_k} \right) &= \frac{\partial}{\partial q_1} \left( \frac{\partial \mathbf{r}_i}{\partial q_k} \right) \dot{q}_1 + \frac{\partial}{\partial q_2} \left( \frac{\partial \mathbf{r}_i}{\partial q_k} \right) \dot{q}_2 + \dots + \frac{\partial}{\partial q_k} \left( \frac{\partial \mathbf{r}_i}{\partial q_k} \right) \dot{q}_k \\ &\quad + \dots + \frac{\partial}{\partial q_f} \left( \frac{\partial \mathbf{r}_i}{\partial q_k} \right) \dot{q}_f + \frac{\partial}{\partial t} \left( \frac{\partial \mathbf{r}_i}{\partial q_k} \right) \\ &= \frac{\partial^2 \mathbf{r}_i}{\partial q_1 \partial q_k} \dot{q}_1 + \frac{\partial^2 \mathbf{r}_i}{\partial q_2 \partial q_k} \dot{q}_2 + \dots + \frac{\partial^2 \mathbf{r}_i}{\partial q_k \partial q_k} \dot{q}_k + \dots + \frac{\partial^2 \mathbf{r}_i}{\partial t \partial q_k} \quad \dots(12) \end{aligned}$$

Comparing Eqs. (11) and (12),

$$\frac{d}{dt} \left( \frac{\partial \mathbf{r}_i}{\partial q_k} \right) = \frac{\partial \dot{\mathbf{r}}_i}{\partial q_k} \quad \dots(13)$$

$$\text{From Eq. (10), } \frac{d}{dt} \left( \dot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \right) = \frac{d}{dt} \left( \dot{\mathbf{r}}_i \cdot \frac{\partial \dot{\mathbf{r}}_i}{\partial q_k} \right) = \frac{d}{dt} \frac{\partial}{\partial \dot{q}_k} \left( \frac{1}{2} \dot{\mathbf{r}}_i^2 \right) \quad \dots(14)$$

Substituting (13) and (14) in Eq. (9),

$$\begin{aligned} \sum_i \sum_k m_i \left[ \frac{d}{dt} \frac{\partial}{\partial \dot{q}_k} \left( \frac{1}{2} \dot{\mathbf{r}}_i^2 \right) - \dot{\mathbf{r}}_i \cdot \frac{\partial \dot{\mathbf{r}}_i}{\partial q_k} \right] \delta q_k &= \sum_i \sum_k \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \delta q_k \\ \text{or } \sum_i \sum_k \left[ \frac{d}{dt} \frac{\partial}{\partial \dot{q}_k} \left( \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 \right) - \frac{\partial}{\partial q_k} \left( \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 \right) \right] \delta q_k &= \sum_i \sum_k \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \delta q_k \\ \text{or } \sum_k \left[ \frac{d}{dt} \frac{\partial}{\partial \dot{q}_k} \left( \sum_i \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 \right) - \frac{\partial}{\partial q_k} \left( \sum_i \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 \right) \right] \delta q_k &= \sum_i \sum_k \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \delta q_k \end{aligned} \quad \dots(15)$$

$$\sum_i \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 = T = \text{Total kinetic energy of the system of particles} \quad \dots(16)$$

$$\text{and } \sum_i \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} = \mathbf{Q}_k \quad \dots(17)$$

Here  $\mathbf{Q}_k$ 's are components of *generalised force*.

$$\text{Eq. (15) becomes, } \sum_k \left\{ \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_k} - \frac{\partial T}{\partial q_k} \right\} \delta q_k = \sum_k \mathbf{Q}_k \delta q_k \quad \dots(18)$$

$$\therefore \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_k} - \frac{\partial T}{\partial q_k} = Q_k \quad \dots(19)$$

This is the general form of *Lagrange's equation*. There are  $f$  such equations corresponding to  $f$  generalised co-ordinates.

When the system is wholly conservative,

$$\mathbf{F}_i = -\nabla V_i = -\frac{\partial V_i}{\partial \mathbf{r}_i}, \quad \dots(20)$$

$$\mathbf{Q}_k = \sum_i \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} = -\sum_i \frac{\partial V_i}{\partial \mathbf{r}_i} \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} = -\sum_i \frac{\partial V_i}{\partial q_k} = -\frac{\partial}{\partial q_k} (\sum_i V_i) = -\frac{\partial V}{\partial q_k}$$

Here  $V = \sum_i V_i$  = total potential energy of the system.

Putting this value of  $\mathbf{Q}_k$  in Eq. (19), we get

$$\begin{aligned} \frac{d}{dt} \left( \frac{\partial T}{\partial \dot{q}_k} \right) - \frac{\partial T}{\partial q_k} &= -\frac{\partial V}{\partial q_k} \\ \text{or } \frac{d}{dt} \left( \frac{\partial T}{\partial \dot{q}_k} \right) - \frac{\partial T}{\partial q_k} + \frac{\partial V}{\partial q_k} &= 0 \\ \text{or } \frac{d}{dt} \left( \frac{\partial T}{\partial \dot{q}_k} \right) - \frac{\partial}{\partial q_k} (T - V) &= 0 \end{aligned} \quad \dots(21)$$

The potential energy  $V$  is the function of position co-ordinates  $q_k$  and not of the *generalised velocities*  $\dot{q}_k$ . Therefore, Eq. (21) may be written as

$$\frac{d}{dt} \frac{\partial}{\partial \dot{q}_k} (T - V) - \frac{\partial}{\partial q_k} (T - V) = 0 \quad \dots(22)$$

But  $L = T - V$ , where  $L$  is known as *Lagrangian function*.

$$\text{Eq. (22) becomes, } \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_k} \right) - \frac{\partial L}{\partial q_k} = 0 \quad \dots(23)$$

This is Lagrange's equation for a *conservative system*.

**(b) Lagrange's equations for systems containing dissipative forces.**

Consider a system of particles containing dissipative or frictional forces.

The frictional force is proportional to the velocity of the particle, *i.e.*,

$$\mathbf{F}_i^{(d)} = -\lambda_i \dot{\mathbf{r}}_i, \quad \dots(1)$$

Here  $\mathbf{F}_i^{(d)}$  denotes the dissipative force,  $\dot{\mathbf{r}}_i$  is the velocity of  $i^{\text{th}}$  particle and  $\lambda_i$  is the corresponding constant of proportionality.

Forces of this type are derivable from *Rayleigh's dissipation function*  $R$  defined by

$$R = \frac{1}{2} \sum_i \lambda_i \dot{\mathbf{r}}_i^2 \quad \dots(2)$$

Here  $i = 1, 2, \dots, n$  covers all the particles of the system.

$$\frac{\partial R}{\partial \dot{\mathbf{r}}_i} = \lambda_i \dot{\mathbf{r}}_i = -\mathbf{F}_i^{(d)} \quad \text{from Eq. (1)}$$

$$\text{or} \quad \mathbf{F}_i^{(d)} = -\frac{\partial R}{\partial \dot{\mathbf{r}}_i} \quad \dots(3)$$

Here term  $\frac{\partial R}{\partial \dot{\mathbf{r}}_i}$  represents the dissipative force.

The Lagrange's equation in terms of  $\mathbf{r}$  is given by

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \mathbf{r}} \right) - \frac{\partial L}{\partial \dot{\mathbf{r}}} = \mathbf{F}_i^{(d)} \quad \dots(4)$$

Lagrange's equation in generalised coordinates  $q_k$  is

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_k} \right) - \frac{\partial L}{\partial q_k} = \mathbf{Q}_k^{(d)}$$

where  $\mathbf{Q}_k^{(d)}$  is component of generalised force.

In order to find the Lagrange's equation in generalised co-ordinates, we have to find the components of generalised force resulting from the dissipative force.

If  $\mathbf{Q}_k^{(d)}$  is the component of generalised force along  $q_k$ , then

$$\begin{aligned} \mathbf{Q}_k^{(d)} &= \sum_i \mathbf{F}_i^{(d)} \frac{\partial \mathbf{r}_i}{\partial q_k} \\ &= -\sum_i \lambda_i \dot{\mathbf{r}}_i \frac{\partial \mathbf{r}_i}{\partial q_k} \quad [\text{from Eq. (1)}] \\ &= -\sum_i \lambda_i \dot{\mathbf{r}}_i \frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_k} \quad \left[ \because \frac{\partial \mathbf{r}_i}{\partial q_k} = \frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_k} \right] \end{aligned}$$

$$\text{or} \quad \mathbf{Q}_k^{(d)} = -\sum_i \lambda_i \frac{\partial}{\partial \dot{q}_k} \left( \frac{1}{2} \dot{\mathbf{r}}_i^2 \right)$$

$$= \frac{\partial}{\partial \dot{q}_k} \sum_i \left( -\frac{1}{2} \lambda_i \dot{\mathbf{r}}_i^2 \right)$$

$$\therefore \mathbf{Q}_k^{(d)} = -\frac{\partial R}{\partial \dot{q}_k} \quad \dots(5)$$

Therefore the Lagrange's equation for a system containing dissipative force is given by

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_k} \right) - \frac{\partial L}{\partial q_k} + \frac{\partial R}{\partial \dot{q}_k} = 0 \quad \dots(6)$$

Hence the term  $\frac{\partial R}{\partial \dot{q}_k}$  takes into account the dissipative forces.

Thus, if dissipative forces are acting on the system, we must specify two scalar functions—the Lagrangian  $L$  and Rayleigh's dissipation function  $R$ —to derive the equations of motion.

## 42.6 APPLICATIONS OF LAGRANGE'S EQUATIONS

In order to use Lagrange's equations for the solution of a physical problem, one must use the following steps:

- (i) Choose an appropriate coordinate system.
- (ii) Write down the expressions for potential and kinetic energies.
- (iii) Write down the equations of constraint, if any.
- (iv) Choose the generalized coordinates.
- (v) Set up the Lagrangian,  $L = T - V$
- (vi) Solve Lagrange's equations for each generalized coordinate using, if necessary, the equations of constraint.

### (a) The Atwood's machine:

Let two small heavy particles of masses  $M_1$  and  $M_2$  be connected by a light inextensible rope of length  $l$  passing over a frictionless light pulley (Fig. 42.8).

It is found that the heavier particle descends while the lighter ascends, the system moving with a constant acceleration  $\ddot{x}$ .

The Atwood's machine is a conservative system with a holonomic constraint. There is only one independent coordinate  $x$ . The position of the other particle is determined by the constraint that the length of the rope between them is  $l$ .

The P.E. of the system =  $V = -M_1gx - M_2g(l-x)$

The K.E. of the system =  $T = \frac{1}{2}(M_1 + M_2)\dot{x}^2$

Hence, the Lagrangian function is given by

$$L = T - V = \frac{1}{2}(M_1 + M_2)\dot{x}^2 + M_1gx + M_2g(l-x)$$

The Lagrange's equation for a conservative system is

$$\frac{d}{dt} \left[ \frac{\partial L}{\partial \dot{q}_i} \right] - \frac{\partial L}{\partial q_i} = 0$$

Since the system has only one degree of freedom, there is only one equation of motion, involving the derivatives.

$\therefore$  The equation of motion of the system is given by

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{x}} \right) - \frac{\partial L}{\partial x} = 0$$

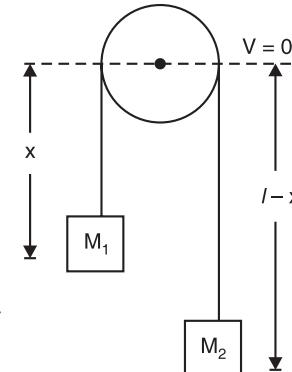


Fig. 42.8

From the expression for  $L$  we get,  $\frac{\partial L}{\partial x} = (M_1 - M_2)g$ ;

$$\frac{\partial L}{\partial \dot{x}} = (M_1 + M_2)\dot{x}; \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = \frac{d}{dt}[(M_1 + M_2)\dot{x}] = (M_1 + M_2)\ddot{x}.$$

$\therefore$  We have,  $(M_1 + M_2)\ddot{x} - (M_1 - M_2)g = 0$

or  $\ddot{x} = \frac{M_1 - M_2}{M_1 + M_2} g.$

**(b) Simple pendulum:** Let the mass of the pendulum bob be  $m$ . The angle  $\theta$  between rest position ( $OA$ ) and deflected position ( $OB$ ) is chosen as the generalized coordinate (Fig. 42.9). Let  $l$  be the length of the string. Then

$$\text{K.E. of the bob} = \frac{1}{2}mv^2$$

$$T = \frac{1}{2}ml^2\dot{\theta}^2$$

P.E. of the bob is

$$V = mg CA = mg(OA - OC) = mg(l - l \cos \theta) = mgl(1 - \cos \theta)$$

The Lagrangian function  $L$  is given by

$$L = T - V = \frac{1}{2}ml^2\dot{\theta}^2 - mgl(1 - \cos \theta) \quad \dots(1)$$

The Lagrangian equation in the generalized coordinate  $\theta$  is

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} = 0 \quad \dots(2)$$

From Eq. (1),  $\frac{\partial L}{\partial \dot{\theta}} = ml^2\dot{\theta}$  and  $\frac{\partial L}{\partial \theta} = -mgl \sin \theta$

Substituting these values in Eq. (2), we get

$$\frac{d}{dt}(ml^2\dot{\theta}) + mgl \sin \theta = 0$$

or  $ml^2\ddot{\theta} + mgl \sin \theta = 0$

or  $\ddot{\theta} + \frac{g}{l} \sin \theta = 0$

If the amplitude of motion is small,  $\sin \theta \approx \theta$ . Then

$$\ddot{\theta} + \frac{g}{l} \theta = 0 \quad \dots(3)$$

Eq. (3) represents a S.H.M. with period  $T = 2\pi\sqrt{l/g}$

### (c) Compound Pendulum

A rigid body capable of oscillating in a vertical plane about a fixed horizontal axis is called a compound pendulum. Let vertical plane of oscillation be  $xy$ ,  $O$  the point through which axis of rotation passes,  $C$  the centre of mass. Let mass of pendulum be  $m$ , moment of inertia about axis of rotation  $I$  and distance  $OC = h$  (Fig. 42.10). Let  $\theta$  be the angle through which the body is displaced.

K.E. of the system is

$$T = \frac{1}{2}I\dot{\theta}^2$$

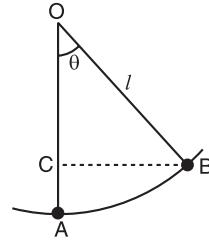


Fig. 42.9

Consider the horizontal plane passing through  $O$  as reference level.

P.E. of the system is

$$\begin{aligned} V &= -mg(OA) \\ &= -mgh \cos \theta \end{aligned}$$

The Lagrangian  $L$  is written as

$$\begin{aligned} L &= T - V = \frac{1}{2} I \dot{\theta}^2 + mgh \cos \theta \\ \therefore \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{\theta}} \right) - \frac{\partial L}{\partial \theta} &= I \ddot{\theta}; \quad -mgh \sin \theta \end{aligned}$$

The Lagrangian equation in the generalized coordinate  $\theta$  is

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{\theta}} \right) - \frac{\partial L}{\partial \theta} = 0$$

$$\text{or } \frac{d}{dt} (I \dot{\theta}) + mgh \sin \theta = 0$$

$$\text{or } \ddot{\theta} + \frac{mgh}{I} \sin \theta = 0$$

If amplitude of oscillation is small,  $\sin \theta \approx \theta$ . Then

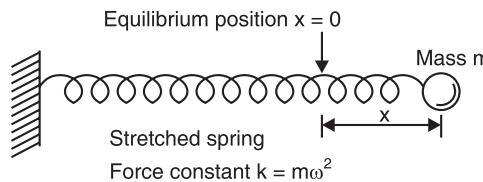
$$\ddot{\theta} + \frac{mgh}{I} \theta = 0$$

This is an equation for simple harmonic motion of time period

$$T = 2\pi \sqrt{\frac{I}{mgh}}$$

#### (d) Linear Harmonic Oscillator

The traditional ideal Harmonic oscillator is shown in Fig. 42.11. The displacement of the mass from its equilibrium position is  $x$ .



**Fig. 42.11**

Lagrange's equation of motion for one dimensional motion, say in  $x$  direction, is

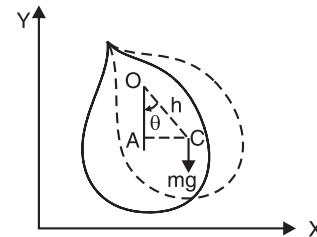
$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{x}} \right) - \frac{\partial L}{\partial x} = 0. \quad \dots(1)$$

The kinetic energy of this system is

$$T = \frac{1}{2} m \dot{x}^2$$

Potential energy is

$$\begin{aligned} V &= - \int \mathbf{F} \cdot d\mathbf{x} \\ &= - \int -kx \, dx = \frac{1}{2} kx^2 + c \end{aligned}$$



**Fig. 42.10**

Here,  $c$  is a constant of integration and  $k$  is spring constant.

If we choose the horizontal plane passing through the position of equilibrium as the reference level, then  $V = 0$  at  $x = 0$ .

$$\therefore c = 0$$

The Lagrangian of the system is

$$L = T - V = \frac{1}{2} m\dot{x}^2 - \frac{1}{2} kx^2 \quad \dots(2)$$

$$\frac{\partial L}{\partial \dot{x}} = m\dot{x} \quad \text{and} \quad \frac{\partial L}{\partial x} = -kx.$$

Eq. (1) takes the form

$$\frac{d}{dt}(m\dot{x}) - (-kx) = 0$$

or

$$m\ddot{x} + kx = 0 \quad \dots(3)$$

This relation shows that the motion is simple harmonic.

$$\therefore \text{The frequency} \quad \nu = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad \dots(4)$$

## HAMILTONIAN FORMULATION OF MECHANICS

### 42.7 PHASE SPACE

While using the Lagrangian formulation, there are  $3n$  equations of motion of the 2nd order for a system containing  $n$  particles in the absence of holonomic constraints. For general consideration it is more convenient to write  $6n$  partial differential equations of the 1st order in place of  $3n$  equations of the 2nd order. For the purpose there must be  $6n$  degrees of freedom or  $6n$  dimensional space known as *phase space*. In phase space, momenta are also regarded as independent variables like space co-ordinates. A single particle in phase space is specified by six co-ordinates, 3 position co-ordinates and 3 momentum co-ordinates. This six dimensional space is sometimes called the  $\mu$  space. *The phase space is a superposition of  $\mu$  spaces.*

In brief *phase space is a  $6n$  dimensional space formed of co-ordinates  $q_1, q_2, \dots, q_k, \dots, q_{3n}, p_1, p_2, \dots, p_k, \dots, p_{3n}$ .* A single point in this phase space will fix all the position coordinates and momenta. The point thus describes the state of motion of the system besides giving its position.

### 42.8 THE HAMILTONIAN FUNCTION H

The Lagrangian of a system is a function of

$q_1, q_2, \dots, q_k, \dots, \dot{q}_1, \dot{q}_2, \dots, \dot{q}_k, \dots$  and  $t$ .

$$L = L(q_k, \dot{q}_k, t)$$

$$\frac{dL}{dt} = \sum_k \frac{\partial L}{\partial q_k} \dot{q}_k + \sum_k \frac{\partial L}{\partial \dot{q}_k} \ddot{q}_k + \frac{\partial L}{\partial t} \quad \dots(1)$$

The Lagrangian equation is

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_k} \right) - \frac{\partial L}{\partial q_k} = 0$$

or

$$\frac{\partial L}{\partial q_k} = \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_k} \right) \quad \dots(2)$$

Substituting this value of  $\frac{\partial L}{\partial q_k}$  in Eq. (1),

$$\begin{aligned}\frac{dL}{dt} &= \sum_k \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_k} \right) \dot{q}_k + \sum_k \frac{\partial L}{\partial \dot{q}_k} \ddot{q}_k + \frac{\partial L}{\partial t} \\ &= \sum_k \frac{d}{dt} \left\{ \frac{\partial L}{\partial \dot{q}_k} \dot{q}_k \right\} + \frac{\partial L}{\partial t} \\ \text{or } \frac{d}{dt} \left\{ L - \sum_k \frac{\partial L}{\partial \dot{q}_k} \dot{q}_k \right\} &= \frac{\partial L}{\partial t} \quad \dots(3)\end{aligned}$$

But  $\frac{\partial L}{\partial \dot{q}_k} = p_k$  generalised momentum.

Eq. (3) becomes,  $\frac{d}{dt} \left\{ L - \sum_k p_k \dot{q}_k \right\} = \frac{\partial L}{\partial t}$

$$\text{or } \frac{d}{dt} \left\{ \sum_k p_k \dot{q}_k - L \right\} = -\frac{\partial L}{\partial t} \quad \dots(4)$$

$$\text{Hamiltonian function } H = \sum_k p_k \dot{q}_k - L(q, \dot{q}, t) \quad \dots(5)$$

The Hamiltonian  $H$  is related to the Lagrangian  $L$  through Eq. (5).

## 42.9 HAMILTON'S CANONICAL EQUATIONS OF MOTION

The Hamiltonian function  $H$  is function of  $p, q$  and  $t$ .

$$\begin{aligned}H &= H(p, q, t) \\ &= H(p_1, p_2, \dots, p_k, \dots, q_1, q_2, \dots, q_k, \dots, t) \\ dH &= \sum_k \frac{\partial H}{\partial p_k} dp_k + \sum_k \frac{\partial H}{\partial q_k} dq_k + \frac{\partial H}{\partial t} dt \quad \dots(1)\end{aligned}$$

Now,

$$H = \sum_k p_k \dot{q}_k - L(q, \dot{q}, t).$$

$$\therefore dH = \sum_k \dot{q}_k dp_k + \sum_k p_k d\dot{q}_k - dL, \quad \dots(2)$$

Now,

$$L = L(q_1, q_2, \dots, q_k, \dots, \dot{q}_1, \dot{q}_2, \dots, \dot{q}_k, \dots, t)$$

$$\begin{aligned}dL &= \sum_k \frac{\partial L}{\partial q_k} dq_k + \sum_k \frac{\partial L}{\partial \dot{q}_k} d\dot{q}_k + \frac{\partial L}{\partial t} dt \\ &= \sum_k \frac{\partial L}{\partial q_k} dq_k + \sum_k p_k d\dot{q}_k + \frac{\partial L}{\partial t} dt \quad \dots(3)\end{aligned}$$

Substituting this value of  $dL$  in Eq. (2),

$$\begin{aligned}dH &= \sum_k \dot{q}_k dp_k + \sum_k p_k d\dot{q}_k - \sum_k \frac{\partial L}{\partial q_k} dq_k - \sum_k p_k d\dot{q}_k - \frac{\partial L}{\partial t} dt \\ &= \sum_k \dot{q}_k dp_k - \sum_k \dot{p}_k dq_k - \frac{\partial L}{\partial t} dt \quad \dots(4)\end{aligned}$$

Comparing coefficients of  $dp_k$  and  $dq_k$  in Eqs. (1) and (4),

$$\left. \begin{aligned} \dot{q}_k &= \frac{\partial H}{\partial p_k}, \\ -\dot{p}_k &= \frac{\partial H}{\partial q_k}, \end{aligned} \right\} \quad \dots(5)$$

Eqs. (5) are called *Hamilton's equations or Hamilton's canonical equations of motion*.  $p$ 's and  $q$ 's are called canonical variables.

$$\text{Also, } -\frac{\partial L}{\partial t} = \frac{\partial H}{\partial t} \quad \dots(6)$$

## 42.10 PHYSICAL SIGNIFICANCE OF THE HAMILTONIAN FUNCTION

$$H = H(p_1, p_2, \dots, p_k, \dots, q_1, q_2, \dots, q_k, \dots, t)$$

The total time derivative of  $H$  is

$$\begin{aligned} \frac{dH}{dt} &= \sum_k \frac{\partial H}{\partial q_k} \dot{q}_k + \sum_k \frac{\partial H}{\partial p_k} \dot{p}_k + \frac{\partial H}{\partial t} \\ &= -\sum_k \dot{p}_k \dot{q}_k + \sum_k \dot{q}_k \dot{p}_k + \frac{\partial H}{\partial t} \\ &= \frac{\partial H}{\partial t}. \end{aligned} \quad \dots(1)$$

But

$$\frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t} \quad \dots(2)$$

$\therefore$

$$\frac{dH}{dt} = -\frac{\partial L}{\partial t} \quad \dots(3)$$

If  $L$  is not an explicit function of time,  $\frac{\partial L}{\partial t} = 0$ . Hence

$$\frac{dH}{dt} = 0.$$

or

$$H = \text{constant.} \quad \dots(4)$$

Therefore  $H$  is a constant of the system.

For conservative systems the potential energy  $V$  does not depend upon generalised velocity, i.e.,

$$\frac{\partial V}{\partial \dot{q}_k} = 0 \quad \dots(5)$$

Also we know

$$\begin{aligned} H &= \sum_k p_k \dot{q}_k - L \\ &= \sum_k \frac{\partial L}{\partial \dot{q}_k} \dot{q}_k - L \\ &= \sum_k \dot{q}_k \left\{ \frac{\partial}{\partial \dot{q}_k} (T - V) \right\} - L \\ &= \sum_k \dot{q}_k \left\{ \frac{\partial T}{\partial \dot{q}_k} - \frac{\partial V}{\partial \dot{q}_k} \right\} - L \\ &= \sum_k \dot{q}_k \frac{\partial T}{\partial \dot{q}_k} - L \end{aligned} \quad \text{from Eq. (5)}$$

$$\begin{aligned}
&= \sum \dot{q}_k \frac{\partial}{\partial \dot{q}_k} \left( \sum_i \frac{1}{2} m_i \dot{r}_i^2 \right) - L \\
&= \sum_k m \dot{q}_k^2 - L \\
&= 2T - L \\
&= 2T - (T - V) \\
&= T + V \\
&= \text{K.E.} + \text{P.E.} \\
&= E = \text{total energy of the system.}
\end{aligned}$$

Thus for conservative systems where the co-ordinate transformation is independent of time, the Hamiltonian function  $H$  represents the total energy of the system.

## 42.11 HAMILTON'S VARIATIONAL PRINCIPLE

**Statement:** The path actually traversed by a conservative, holonomic dynamical system from time  $t_1$  to  $t_2$  is one over which the integral of the Lagrangian between limits  $t_1$  and  $t_2$  is stationary i.e., the time integral of the Lagrangian is extremum.

**Explanation:** The motion of the system from time  $t_1$  to time  $t_2$  is such that the line integral

$$I = \int_{t_1}^{t_2} L dt \quad \dots(1)$$

where  $L = T - V$ , is an extremum for the path of motion.

$$\text{or } \delta \int_{t_1}^{t_2} L dt = 0 \quad \dots(2)$$

This principle helps us to distinguish the actual path from the neighbouring paths.

### 42.11.1 Derivation of Hamilton's Canonical Equations of Motion from Hamilton's Variational Principle

Hamilton's principle is

$$\delta I = \delta \int_{t_1}^{t_2} L dt = 0$$

The relation between Lagrangian and Hamiltonian is

$$\begin{aligned}
L &= \sum p_i \dot{q}_i - H(q, p, t) \\
\therefore \delta \int_{t_1}^{t_2} \left[ \sum_i p_i \dot{q}_i - H(q, p, t) \right] dt &= 0 \\
\text{or } \delta \int_{t_1}^{t_2} \left[ \sum_i p_i \frac{dq_i}{dt} - H(q_i, p_i, t) \right] dt &= 0 \\
\text{or } \delta \sum_i \int_{q_1}^{q_2} p_i dq_i - \delta \int_{t_1}^{t_2} H dt &= 0. \quad \dots(1)
\end{aligned}$$

Eq. (1) is called the *modified Hamilton's principle*.

Now labelling each of the possible paths in the configuration space with a parameter  $\alpha$ , the  $\delta$ -variation can be expressed as

$$\delta = d\alpha (\partial/\partial\alpha).$$

$$\text{Now } \delta I = \frac{\partial I}{\partial \alpha} d\alpha = d\alpha \frac{\partial}{\partial \alpha} \int_{t_1}^{t_2} \left[ \sum_i p_i \dot{q}_i - H(q, p, t) \right] dt = 0$$

The times  $t_1, t_2$  are not varied and so they are not functions of  $\alpha$ . Thus, the differentiation and integration may be interchanged.

$$\begin{aligned} d\alpha \int_{t_1}^{t_2} \frac{\partial}{\partial \alpha} \left[ \sum_i p_i \dot{q}_i - H(q, p, t) \right] dt &= 0 \\ \text{or } d\alpha \int_{t_1}^{t_2} \sum_i \left[ \frac{\partial p_i}{\partial \alpha} \dot{q}_i + \frac{\partial \dot{q}_i}{\partial \alpha} p_i - \frac{\partial H}{\partial q_i} \frac{\partial q_i}{\partial \alpha} - \frac{\partial H}{\partial p_i} \frac{\partial p_i}{\partial \alpha} \right] dt &= 0 \end{aligned} \quad \dots(2)$$

Further, we have

$$\int_{t_1}^{t_2} \frac{\partial \dot{q}_i}{\partial \alpha} p_i dt = \int_{t_1}^{t_2} p_i \frac{d}{dt} \left( \frac{\partial q_i}{\partial \alpha} \right) dt = p_i \left[ \frac{\partial q_i}{\partial \alpha} \right]_{t_1}^{t_2} - \int_{t_1}^{t_2} \dot{p}_i \frac{\partial q_i}{\partial \alpha} dt.$$

But all the varied paths have the same end points. Hence  $\frac{\partial q_i}{\partial \alpha}$  vanishes for  $t_1$  and  $t_2$ .

$$\therefore \int_{t_1}^{t_2} \frac{\partial \dot{q}_i}{\partial \alpha} p_i dt = - \int_{t_1}^{t_2} \dot{p}_i \frac{\partial q_i}{\partial \alpha} dt.$$

Therefore Eq. (2) becomes,

$$\begin{aligned} d\alpha \int_{t_1}^{t_2} \sum_i \left[ \frac{\partial p_i}{\partial \alpha} \dot{q}_i - \frac{\partial H}{\partial q_i} \frac{\partial q_i}{\partial \alpha} - \frac{\partial H}{\partial p_i} \frac{\partial p_i}{\partial \alpha} - \dot{p}_i \frac{\partial q_i}{\partial \alpha} \right] dt &= 0 \\ \int_{t_1}^{t_2} \sum_i \left[ \dot{q}_i \frac{\partial p_i}{\partial \alpha} - \frac{\partial H}{\partial q_i} \frac{\partial q_i}{\partial \alpha} - \frac{\partial H}{\partial p_i} \frac{\partial p_i}{\partial \alpha} - \dot{p}_i \frac{\partial q_i}{\partial \alpha} \right] dt &= 0 \end{aligned} \quad \dots(3)$$

But  $\delta p_i = d\alpha \cdot (\partial p_i / \partial \alpha)$  and  $\delta q_i = (\partial q_i / \partial \alpha) \cdot d\alpha$

$$\therefore \int_{t_1}^{t_2} \sum_i \left[ \delta p_i \left( \dot{q}_i - \frac{\partial H}{\partial p_i} \right) + \delta q_i \left( - \frac{\partial H}{\partial q_i} - \dot{p}_i \right) \right] dt = 0 \quad \dots(4)$$

The variations  $\delta q_i$  and  $\delta p_i$  are independent of each other. Hence Eq. (4) holds good only when the coefficients of  $\delta p_i$  and  $\delta q_i$  vanish separately.

$$\therefore \dot{q}_i = \frac{\partial H}{\partial p_i}; \quad \dot{p}_i = - \frac{\partial H}{\partial q_i}$$

These are Hamilton's canonical equations of motion.

## 42.12 APPLICATIONS OF HAMILTON'S EQUATIONS OF MOTION

### 1. Linear Harmonic Oscillator

The system is conservative and constraint is independent of time. So Hamiltonian will represent the total energy of the system.

The kinetic energy of harmonic oscillator,

$$T = \frac{1}{2} m \dot{x}^2$$

The potential energy of harmonic oscillator,

$$\begin{aligned} V &= - \int F dx \\ &= \int kx dx = \frac{1}{2} kx^2 \end{aligned}$$

$\therefore$  The Lagrangian of the system

$$L = T - V = \frac{1}{2} m\dot{x}^2 - \frac{1}{2} kx^2$$

$$\frac{\partial L}{\partial \dot{x}} = \frac{\partial T}{\partial \dot{x}} = p = m\dot{x}$$

or  $\dot{x} = p/m$

The Hamiltonian  $H$  in terms of momenta is

$$H = T + V = \frac{1}{2} m(p/m)^2 + \frac{1}{2} kx^2 = \frac{p^2}{2m} + \frac{1}{2} kx^2$$

Equations of motion are

$$\dot{p} = -\frac{\partial H}{\partial x} = -kx \quad \dots(1)$$

and  $\dot{x} = \frac{\partial H}{\partial p} = p/m \quad \dots(2)$

From Eq. (2)  $\ddot{x} = \dot{p}/m = -kx/m$

$$\therefore m\ddot{x} + kx = 0 \quad \dots(3)$$

This relation shows that the motion is simple harmonic.

$$\therefore \text{The frequency } \nu = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad \dots(4)$$

## 2. Simple Pendulum

Let the mass of the pendulum bob be  $m$ . The angle  $\theta$  between rest position ( $OA$ ) and deflected position ( $OB$ ) is chosen as the generalized coordinate (Fig. 42.12). Let  $l$  be the length of the string.

$$\text{K.E. of the bob} = \frac{1}{2} mv^2$$

$$T = \frac{1}{2} ml^2 \dot{\theta}^2$$

P.E. of the bob is

$$\begin{aligned} V &= mgCA = mg(OA - OC) \\ &= mg(l - l \cos \theta) = mgl(1 - \cos \theta) \end{aligned}$$

The Lagrangian function  $L$  is given by

$$L = T - V = \frac{1}{2} ml^2 \dot{\theta}^2 - mgl(1 - \cos \theta) \quad \dots(1)$$

$$p_\theta = \frac{\partial L}{\partial \dot{\theta}} = ml^2 \dot{\theta} \quad \dots(2)$$

The Hamiltonian is

$$\begin{aligned} H &= \sum p_j \dot{q}_j - L \\ &= p_\theta \dot{\theta} - L \\ &= ml^2 \dot{\theta}^2 - \left\{ \frac{1}{2} ml^2 \dot{\theta}^2 - mgl(1 - \cos \theta) \right\} \\ &= \frac{1}{2} ml^2 \dot{\theta}^2 + mgl(1 - \cos \theta) \quad \dots(3) \\ &= T + V, \end{aligned}$$

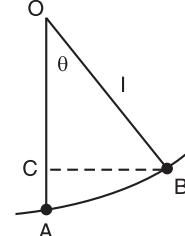


Fig. 42.12

which implies that system is conservative.

Now putting Eq. (2) into Eq. (3), we get

$$\begin{aligned} H &= \frac{1}{2} ml^2 \left( \frac{p_\theta}{ml^2} \right)^2 + mgl (1 - \cos \theta), \\ \frac{\partial H}{\partial p_\theta} &= \frac{p_\theta}{ml^2} \\ \frac{\partial H}{\partial \theta} &= mgl \sin \theta. \end{aligned} \quad \dots(4)$$

Hamilton's equations of motion for  $\dot{\theta}$  and  $\dot{p}_\theta$  are

$$\dot{\theta} = \frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{ml^2} \quad \dots(5)$$

$$\dot{p}_\theta = -\frac{\partial H}{\partial \theta} = -mgl \sin \theta. \quad \dots(6)$$

Eqs. (5) and (6) represent Hamilton's equations for a simple pendulum.

From Eq. (5), we have  $p_\theta = ml^2 \dot{\theta}$

$$\text{or } \dot{p}_\theta = ml^2 \ddot{\theta} \quad \dots(7)$$

Putting the value of  $p_\theta$  from Eq. (7) into Eq. (6), we get

$$ml^2 \ddot{\theta} = -mgl \sin \theta$$

$$\text{or } \ddot{\theta} + \frac{g}{l} \sin \theta = 0$$

If the amplitude of motion is small,  $\sin \theta \approx \theta$ . Then

$$\ddot{\theta} + \frac{g}{l} \theta = 0 \quad \dots(8)$$

This gives the equation of motion of the simple pendulum.

Eq. (8) represents a S.H.M. with period

$$T = 2\pi \sqrt{l/g} \quad \dots(9)$$

### 3. Compound Pendulum

A rigid body capable of oscillating in a vertical plane about a fixed horizontal axis is called a compound pendulum. Let vertical plane of oscillation be  $xy$ ,  $O$  the point through which axis of rotation passes,  $C$  the centre of mass. Let mass of pendulum be  $m$ , moment of inertia about axis of rotation  $I$  and distance  $OC = h$  (Fig. 42.13). Let  $\theta$  be the angle through which the body is displaced.

K.E. of the system is

$$T = \frac{1}{2} I \dot{\theta}^2$$

Consider the horizontal plane passing through  $O$  as reference level.

P.E. of the system is

$$V = -mg(OA) = -mgh \cos \theta$$

The Lagrangian  $L$  is written as

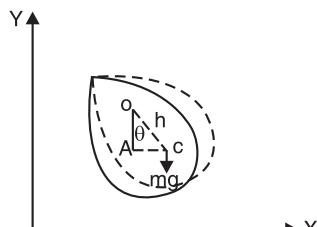


Fig. 42.13

$$L = T - V = \frac{1}{2} I \dot{\theta}^2 + mgh \cos \theta$$

Then  $p_\theta = \frac{\partial L}{\partial \dot{\theta}} = I \dot{\theta}$  ... (1)

The hamiltonian is

$$\begin{aligned} H &= \sum p_j \dot{q}_j - L = p_\theta \dot{\theta} - L \\ &= I \dot{\theta} \cdot \dot{\theta} - \frac{1}{2} I \dot{\theta}^2 - mgh \cos \theta \\ &= \frac{1}{2} I \dot{\theta}^2 - mgh \cos \theta \\ &= \frac{1}{2} I \left( \frac{p_\theta}{I} \right)^2 - mgh \cos \theta \\ &= \frac{p_\theta^2}{2I} - mgh \cos \theta \\ &= T + V \end{aligned} \quad \dots (2)$$

which implies that system is conservative.

Hamilton's equations of motion for  $\theta$  and  $p_\theta$  are

$$\begin{bmatrix} \dot{\theta} = \frac{\partial H}{\partial p_\theta} \\ \dot{p}_\theta = -\frac{\partial H}{\partial \theta} \end{bmatrix} \quad \dots (3)$$

From Eq. (2), we find

$$\frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{I}, \quad \frac{\partial H}{\partial \theta} = mgh \sin \theta.$$

Putting these in Eq. (3), we get

$$\dot{\theta} = \frac{p_\theta}{I} \quad \text{or} \quad \dot{p}_\theta = I \ddot{\theta} \quad \dots (4)$$

and

$$\dot{p}_\theta = -mgh \sin \theta. \quad \dots (5)$$

From Eqs. (4) and (5), we get

$$I \ddot{\theta} = -mgh \sin \theta$$

or  $\ddot{\theta} + \frac{mgh}{I} \sin \theta = 0$

If amplitude of oscillation is small,  $\sin \theta \approx \theta$ . Then

$$\ddot{\theta} + \frac{mgh}{I} \theta = 0 \quad \dots (6)$$

This gives the equation of motion of compound pendulum.

This is an equation for simple harmonic motion of time period

$$T = 2\pi \sqrt{\frac{I}{mgh}} \quad \dots (7)$$

Angular frequency,  $\omega = \sqrt{mgh/I}$  ... (8)

#### 4. Motion of a Particle in a Central Force Field

Let  $P$  be the point mass  $m$  moving in the  $xy$  plane (Fig. 42.14).

The point  $P$  is acted upon by a central force which, is directed towards  $O$ , along the line  $PO$ . Let  $(r, \theta)$  be the polar co-ordinates of the particle. Let  $V(r)$  be the potential energy due to central force.

$$\begin{aligned} \text{Lagrangian } L &= T - V(r) \\ &= \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\theta}^2) - V(r) \end{aligned} \quad \dots(1)$$

Momenta are given as

$$\begin{aligned} p_r &= \frac{\partial H}{\partial \dot{r}} = m \dot{r} \\ p_\theta &= \frac{\partial L}{\partial \dot{\theta}} = m r^2 \dot{\theta} \\ H &= \sum p_k \dot{q}_k - L \\ &= p_r \dot{r} + p_\theta \dot{\theta} - L \\ &= (m \dot{r}) \dot{r} + (m r^2 \dot{\theta}) \dot{\theta} - L \\ &= m \dot{r}^2 + m r^2 \dot{\theta}^2 - \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\theta}^2) + V(r) \quad \text{from (1)} \\ &= \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\theta}^2) + V(r) = T + V(r) \quad \dots(3) \\ &= \text{total energy.} \end{aligned}$$

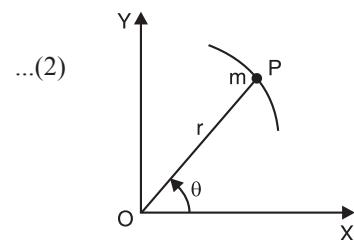


Fig. 42.14

From (2) we have

$$\dot{r} = \frac{p_r}{m}, \quad \dot{\theta} = \frac{p_\theta}{mr^2}$$

With these substitutions the Hamiltonian takes the form

$$\begin{aligned} H &= \frac{1}{2} m \left\{ \left( \frac{p_r}{m} \right)^2 + r^2 \left( \frac{p_\theta}{mr^2} \right)^2 \right\} + V(r) \\ &= \frac{1}{2m} \left( p_r^2 + \frac{p_\theta^2}{r^2} \right) + V(r) \end{aligned} \quad \dots(4)$$

The Hamiltonian equations, corresponding to generalised co-ordinates  $q_k$ , are

$$\begin{aligned} \dot{q}_k &= \frac{\partial H}{\partial p_k}, \\ -\dot{p}_k &= \frac{\partial H}{\partial q_k} \end{aligned}$$

In this case we have two co-ordinates  $r$  and  $\theta$ .

Therefore there will be four Hamilton's equations.

The two equations for  $\dot{q}_k$  are:

$$\begin{aligned} \dot{r} &= \frac{\partial H}{\partial p_r} = \frac{p_r}{m}, \\ \dot{\theta} &= \frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{mr^2} \end{aligned} \quad \dots(5)$$

The two equations for  $p_k$  are:

$$\begin{aligned}\dot{p}_r &= -\frac{\partial H}{\partial r} = \frac{p_\theta^2}{mr^3} - \frac{\partial V(r)}{\partial r} \\ \dot{p}_\theta &= -\frac{\partial H}{\partial \theta} = 0\end{aligned}$$

The last equation shows that the time rate of change of angular momentum is zero, *i.e.*, angular momentum is conserved in planetary motion.

### EXERCISE

1. Derive Lagrange's equations of motion from D'Alembert's principle. (MKU, 1999)
2. Derive Lagrange's equations of motion. Apply it in the case of (i) Atwood's machine (ii) Simple Pendulum (iii) Compound Pendulum (iv) Linear harmonic oscillator.
3. Derive Hamilton's canonical equations of motion and explain the physical significance of the Hamiltonian.
4. Explain any one of the applications of Hamilton's equations. (MKU, 2004)

## STATISTICAL MECHANICS

**After reading this chapter, you should be able to**

- ◆ Discuss microcanonical, canonical and grand canonical ensembles.
- ◆ State and prove Boltzmann's theorem connecting entropy and probability
- ◆ Explain the fundamental differences between classical, Fermi-Dirac and Bose-Einstein statistics
- ◆ Derive MB, BE, and FD distribution laws
- ◆ Discuss applications of statistical physics

### 43.1 INTRODUCTION

Consider an isolated system composed of a large number  $N$  of particles, in which each particle has available to it several states with energies  $E_1, E_2, E_3\dots$ . At a particular time the particles are distributed among the different states, so that  $n_1$  particles have energy  $E_1$ ;  $n_2$  particles have energy  $E_2$ ; and so on. The total number of particles is

$$N = n_1 + n_2 + n_3 + \dots = \sum_i n_i \quad \dots(1)$$

We assume that the total number of particles remains constant for all processes occurring in the system.

The total energy of the system is

$$E = n_1 E_1 + n_2 E_2 + \dots = \sum_i n_i E_i \quad \dots(2)$$

If the system is isolated, the total energy  $E$  must be constant. According to the basic postulate of statistical mechanics, the greater the number  $W$  of different ways in which the particles can be arranged among the available states to yield a particular distribution of energies, the more probable is the distribution. It is assumed that each state of a certain energy is equally likely to be occupied.

Therefore, the procedure for determining the most probable statistical distribution involves two steps: (i) obtaining the number of distinguishable arrangements ( $W$ ) which give rise to the same distribution, and (ii) maximizing this number of arrangements ( $W$ ) with respect to different distributions.

We shall consider systems of three different kinds of particles.

**1. Maxwell-Boltzmann statistics.** This is applicable to the identical, distinguishable particles of any spin. The molecules of a gas are the particles of this kind.

**2. Bose-Einstein statistics.** This is applicable to the identical, indistinguishable particles of zero or integral spin. These particles are called *bosons*. The examples of bosons are helium atoms at low temperature and the photons.

**3. Fermi-Dirac statistics.** This is applicable to the identical indistinguishable particles of half-integral spin. These particles obey Pauli exclusion principle and are called *fermions*. The examples of fermions are electrons, protons, neutrons etc.

Statistical mechanics can be divided into two main classes.

1. Classical Statistics or Maxwell-Boltzmann Statistics.
2. Quantum Statistics.

Classical statistics interpreted successfully many ordinarily observed phenomena such as temperature, pressure, energy etc. But it failed to account for several other experimentally observed phenomena such as black body radiation, photoelectric effect, specific heat capacity at low temperatures, etc. This failure of classical statistics forced the issue in favour of the new quantum idea of discrete exchange of energy between systems. Thus, a new quantum statistics was investigated. There are two types of quantum statistics :

1. Bose-Einstein statistics.
2. Fermi-Dirac statistics.

**1. Bose-Einstein statistics.** This is applicable to the identical, indistinguishable particles of zero or integral spin. These particles are called *bosons*. The examples of bosons are helium atoms at low temperature and the photons.

**2. Fermi-Dirac statistics.** This is applicable to the identical, indistinguishable particles of half-spin. These particles obey Pauli exclusion principle and are called *fermions*. The examples of

fermions are electrons, protons, neutrons etc. In such systems of particles, no more than one particle can be in one quantum state.

The essential difference between the three statistics is illustrated in the following manner. Let there be only two particles of a collection and only two cells to be occupied. Then Fig. 43.1 illustrates the essential difference between the three statistics.

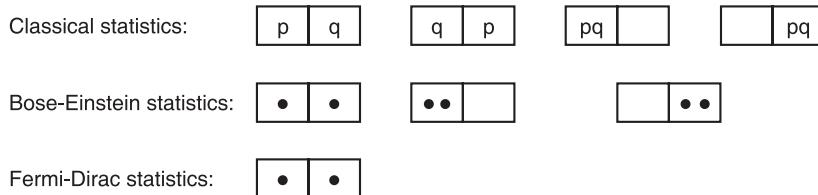


Fig. 43.1

## 43.2 PHASE SPACE

To specify the position as well as energy of a molecule inside a gas, we must specify three space coordinates  $x, y, z$  and three momentum coordinates  $p_x, p_y, p_z$ . As a purely mathematical concept, we may imagine six dimensional space in which coordinates are  $x, y, z, p_x, p_y, p_z$ . This six dimensional space for a single molecule is called *phase space* or  $\mu$ -space. The instantaneous state of a particle in the phase space is represented by a point known as phase point or representative point. In phase space we may consider an element of volume  $dx, dy, dz, dp_x, dp_y, dp_z$ . Any such element of volume in six dimensional space is called a *cell*. Thus the phase space may be divided into a large number of cells. A cell may contain a large number of phase points. Each cell in phase space has the volume  $h^3$ .

**EXAMPLE 1.** Calculate the volume of six dimensional phase space of a state. What is its unit?

(MKU, 2002)

**SOL. Division of phase space into cells:**

The phase space is divided into six dimensional cells whose sides are  $dx, dy, dz, dp_x, dp_y, dp_z$ . Such cells are called phase cells. Further, we approach close to the limit of a point in phase space reducing the size of a cell.

Volume of each cell is

$$d\tau = dx dy dz dp_x dp_y dp_z$$

But according to uncertainty principle,

$$dx dp_x \geq h, \quad dy dp_y \geq h, \quad dz dp_z \geq h,$$

Thus

$$d\tau \geq h^3.$$

A point in the phase space is actually considered to be a cell whose minimum volume is of the order of  $h^3$ .

$$\therefore \text{Volume of phase space cell} = h^3.$$

$$\text{The unit of phase-space volume} = (J - s)^3$$

So the particle in the phase space can not be considered exactly located at point  $x, y, z, p_x, p_y, p_z$  but can only be found somewhere within a phase cell centred at that point.

### 43.2.1 Volume in Phase Space

Consider the volume of the momentum space lying between the values of total momentum  $p$  and  $p + dp$ .

Fig. 43.2 shows a small element of momentum space with coordinates in the range  $p$  to  $p + dp$ ,  $\theta$  to  $\theta + d\theta$ , and  $\phi$  to  $\phi + d\phi$ .

The element is considered in polar coordinates  $p$ ,  $\theta$  and  $\phi$ .

Volume of the element is

$$d\tau_p = dp \cdot pd\theta \cdot p \sin \theta d\phi \\ d\tau_p = p^2 dp \sin \theta d\theta d\phi \quad \dots(1)$$

Volume of momentum space lying between  $p$  and  $p + dp$ , and independent of direction is

$$\Delta\tau_p = p^2 dp \cdot \int_0^\pi \sin \theta d\theta \cdot \int_0^{2\pi} d\phi \\ \Delta\tau_p = \pi p^2 dp. \quad \dots(2)$$

This volume is same as the volume of a spherical shell of thickness  $dp$  lying between  $p$  and  $p + dp$ .

Now we shall express the volume of phase space  $\Delta\tau$  in terms of a volume  $\Delta\tau_p$  of momentum space and the volume  $V$  of ordinary space. The relationship is given by

$$\Delta\tau = 4\pi p^2 dp \int_V dx dy dz$$

or

$$\Delta\tau = 4\pi p^2 dp \cdot V. \quad \dots(3)$$

$m$  is the mass of the particle.

The kinetic energy of the particle  $\varepsilon = \frac{p^2}{2m}$

$$\therefore p = \sqrt{(2m\varepsilon)} \quad \text{and} \quad dp = \left(\frac{m}{2\varepsilon}\right)^{1/2} d\varepsilon.$$

Substituting these values in Eq. (3), we get

$$d\tau = 4\pi (2m\varepsilon) \left(\frac{m}{2\varepsilon}\right)^{1/2} d\varepsilon \cdot V$$

or

$$d\tau = 2\pi (2m)^{3/2} \varepsilon^{1/2} d\varepsilon \cdot V \quad \dots(4)$$

Eq. (4) expresses the element of phase space in terms of kinetic energy of the particle.

### 43.2.2 Number of Phase Cells in the given energy Range for three Dimensional Free Particle

The microstate of a particle is specified by three position coordinates  $x, y, z$  and three momentum coordinates  $p_x, p_y, p_z$ .

The volume of phase cell in phase space is

$$\delta x \delta y \delta z \delta p_x \delta p_y \delta p_z = h^3 \quad \dots(1)$$

$$\begin{aligned} \text{Total volume of phase space} &= \iiint \iiint dx dy dz dp_x dp_y dp_z \\ &= V \iiint dp_x dp_y dp_z \end{aligned} \quad \dots(2)$$

because  $\iiint dx dy dz = V$ .

Consider that in the energy range 0 to  $E$ , the possible values of momentum are from 0 to  $p$ . Now

$$\iiint dp_x dp_y dp_z = \frac{4}{3} \pi p_{\max}^3$$

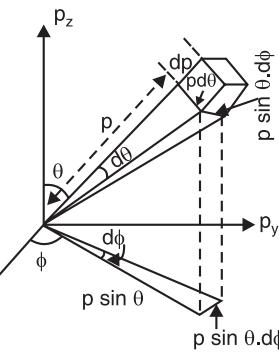


Fig. 43.2

as the volume of momentum space is a sphere of radius  $p$ .

$m$  is the mass of the particle. The energy  $E$  of a free particle is only kinetic.

$$\begin{aligned} E &= \frac{p^2}{2m} \quad \text{or} \quad p_{\max}^2 = p_x^2 + p_y^2 + p_z^2 = 2mE \\ \therefore \iiint dp_x dp_y dp_z &= \frac{4}{3}\pi(2mE)^{3/2} \end{aligned} \quad \dots(3)$$

From Eqs. (2) and (3),

$$\text{total volume of phase space} = V \left\{ \frac{4}{3}\pi(2mE)^{3/2} \right\}$$

$$\text{Volume of a phase cell} = h^3.$$

$$\begin{aligned} \therefore \text{number of cells in phase space in the energy range } 0 \text{ to } E &= V \left\{ \frac{4}{3}\pi(2mE)^{3/2} \right\} \times \frac{1}{h^3} \\ &= \frac{4\pi V}{3h^3} \times (2mE)^{3/2} \end{aligned} \quad \dots(4)$$

### 43.3 ENSEMBLES

A *system* is defined as a collection of a number of particles. An *ensemble* is defined as a collection of a large number of macroscopically identical, but essentially independent systems. By the term *macroscopically identical* we mean that each of the systems constituting an ensemble satisfies the same macroscopic conditions, e.g., volume, energy, pressure, total number of particles etc. By the term *independent systems* we mean that the systems constituting an ensemble are mutually non-interacting. In an ensemble, systems play the same role as the non-interacting molecules do in a gas.

In the formulation of statistical mechanics, Gibbs introduced three standard ensembles to which real experiments could be approximated. These are: (i) the microcanonical ensemble, (ii) the canonical ensemble and (iii) the grand canonical ensemble. Their classification depends on the manner in which their systems interact. Each of them has its own characteristics distribution. Physical systems can interact in a variety of ways; in particular, they may exchange energy, or matter or both with each other.

- (i) The ensemble in which both energy and matter are exchanged between the systems is called a *grand canonical ensemble*.
- (ii) The ensemble in which the systems exchange energy but not matter, is called a *canonical ensemble*.
- (iii) The ensemble in which neither energy nor matter is exchanged, is called a *microcanonical ensemble*.

Statistical mechanics provides a number of methods for calculating equilibrium thermodynamic properties of macroscopic systems. Explicit calculations of thermodynamic functions can be carried out using microcanonical, canonical or grand canonical ensembles. A specific choice of ensembles may be thought of as corresponding to a particular physical situation.

#### 43.3.1 Microcanonical Ensemble

- The *Microcanonical ensemble* describes an *isolated system*.

This ensemble consists of systems which are isolated from the rest of the world. Such a system is also known as a ‘closed isolated system’ and has a fixed volume, fixed total energy and a fixed total number of particles.

**Definition.** The microcanonical ensemble is a collection of essentially independent assemblies having the same energy  $E$ , volume  $V$ , and number  $N$  of systems; all the systems are of the same type.

The individual assemblies are separated by rigid, impermeable and well insulated walls (Fig. 43.3) such that the values of  $E$ ,  $V$  and  $N$  are not affected by the presence of other systems.

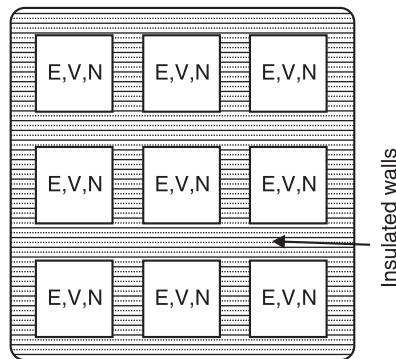


Fig. 43.3

Consider a closed system for which the total energy  $H(q, p) = E$  remains constant, according to the equation

$$E(q_1, \dots, q_f, p_1 \dots p_f) = \text{constant}.$$

Fig. 43.4 shows energy shell in phase space. The locus of all the phase points having equal energies in phase space is called *energy surface* or *ergodic surface*.

Imagine a family of such energy surfaces constructed in phase space.

- Consider two neighbouring surfaces with energies  $E$  and  $E + \delta E$ . Each surface divides the phase space into two parts, one of higher and the other of lower energy.

Hence they will never intersect each other. As they include some phase volume in between them, they will contain certain number of phase points. The number of phase points between them will be constant.

- A very useful ensemble is obtained by taking the density as equal to zero for all values of the energy except in a selected narrow range  $E$  and  $E + \delta E$ . This ensemble, specified by

$$\begin{aligned} \rho &= \text{constant} && (\text{in the range } E \text{ and } E + \delta E) \\ \rho &= 0, && (\text{outside this range}) \end{aligned}$$

is called a **microcanonical ensemble**.

#### Properties of microcanonical ensemble.

- As  $\rho$  is a function of energy, this ensemble is in statistical equilibrium.
- The average properties predicted by such ensemble will not vary in time being in statistical equilibrium.
- As  $\rho$  is constant within the energy shell, the distribution of phase points is uniform (by Liouville's theorem).

#### 43.3.2 Canonical Ensemble

**Definition.** The canonical ensemble is a collection of essentially independent assemblies having the same temperature  $T$ , volume  $V$ , and number of identical particles  $N$ .

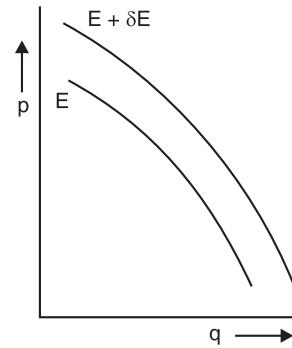


Fig. 43.4

Fig. 43.5 shows a canonical ensemble. Each assembly is in thermal contact with a large heat reservoir at temperature  $T$ . So all the assemblies have the same temperature. The individual assemblies are separated by rigid, impermeable, but diathermic walls. The *diathermic* walls allow heat to flow through them but do not allow the exchange of mass. Since energy can be exchanged between the assemblies, they will reach a common temperature. Thus in *canonical ensemble*, system can exchange energy but not particles.

- Microcanonical ensembles are best suited for isolated systems. We seldom deal with a completely isolated system. There is always some energy exchange with surroundings. It would be more realistic, therefore, to use an ensemble in which statistical equilibrium is attained not by isolation but by free interaction. Canonical ensembles are more appropriate for the description of such systems.

### 43.3.3 Grand Canonical Ensemble

- An ensemble in which *exchange of energy as well as of particles takes place with the heat reservoir* is called *grand canonical ensemble*.

The grand canonical ensemble is a collection of essentially independent assemblies having same temperature  $T$ , volume  $V$  and chemical potential  $\mu$ .

Fig. 43.6 shows a grand canonical ensemble.

- The individual assembly systems are separated by rigid, permeable, diathermic walls.

We effectively have collection of assemblies, each occupying a separate volume  $V$ , but which can exchange energy as well as molecules with each other.

The grand canonical ensemble will thus correspond to the situation when we know both the average energy and the average number of particles in an assembly, but are otherwise completely ignorant about the state of the system.

- In grand canonical ensemble, we abandon (i) the condition of constant energy and (ii) the condition of total number of particles. Actually in chemical process this number varies and in various physical problems, e.g., radioactive decay, it is difficult to keep the number of particles constant.

## 43.4 LIOUVILLE'S THEOREM

**$\Gamma$ -space.** Consider a gas (system) containing  $n$ -molecules. For a system with  $n$  molecules, a set of  $6n$  members  $(q_1, q_2, \dots, q_{3n}, p_1, p_2, \dots, p_{3n})$  is required to represent a phase point. Such a  $6n$ -dimensional space is called  $\Gamma$ -space.

Fig. 43.7 shows  $\Gamma$ -space and phase point for the whole system.

A volume element in  $6n$ -dimensional phase space is expressed as

$$d\Gamma = dq_1 dq_2 \dots dq_{3n} dp_1 dp_2 \dots dp_{3n}$$

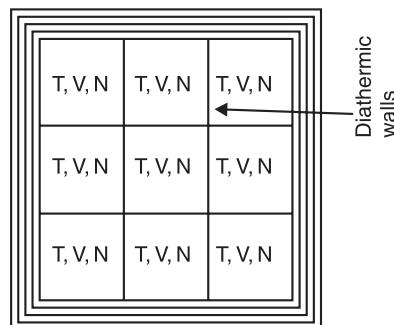


Fig. 43.5

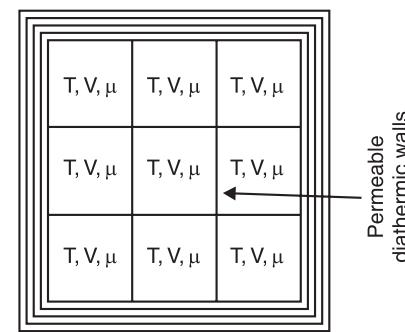


Fig. 43.6

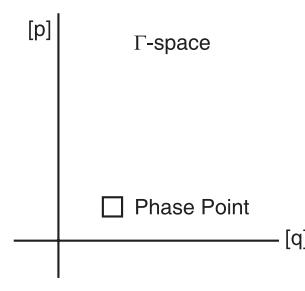


Fig. 43.7

Every possible dynamical state of a system is represented by a unique point in its phase space. The state of the system at any *given* time determines uniquely its state at any other time. The motion of every system is governed by Hamilton's equations,

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \quad \text{and} \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}.$$

Here,  $H = H(q_1 \dots q_f, \dots p_1 \dots p_f)$  is the Hamiltonian of the system. The point in phase space moves along a path, called trajectory. The trajectory is determined from Hamilton's equations of motion. One and only one trajectory passes through each point in phase space. As a result of this motion, the density  $\rho$  of the system in phase space changes with time. Let us now inquire into the rate of change of phase point density.

#### 43.4.1 Liouville's Theorem

##### Statement and Explanation.

Liouville's theorem consists of two parts:

###### (1) The principle of conservation of density in phase space:

The rate of change of density of phase points in phase space is zero or  $\frac{d\rho}{dt} = 0$ .

Gibbs referred to this fact as the *principle of conservation of density in phase space. Therefore, the density of a group of points remains constant along their trajectories in the phase space.*

Hence, the distribution of representative points moves in the  $\Gamma$ -space like an incompressible fluid.

###### (2) The principle of conservation of extension in phase space:

The volume at the disposal of a particular number of phase points is conserved throughout the phase space

$$\text{or} \quad \frac{d}{dt}(\delta\Gamma) = 0$$

It is the *principle of phase volume maintenance* (Fig. 43.8).

Thus, if a given number of representative points occupy an element of volume at a certain time, they will occupy another volume of equal size at a later time, although the volume element may change its shape.

##### Proof.

###### Part I: Principle of conservation of density in phase space:

Consider a small volume element in the hyperspace, bounded by  $q_1$  and  $q_1 + \delta q_1$ ,  $p_1$  and  $p_1 + \delta p_1$ , etc. Consider any arbitrary hypervolume

$$d\Gamma = \delta q_1 \delta q_2 \dots \delta q_f, \delta p_1 \delta p_2 \dots \delta p_f$$

in the phase space located between  $q_1$  and  $(q_1 + \delta q_1)$ , ...,  $q_f$  and  $(q_f + \delta q_f)$ ,  $p_1$  and  $(p_1 + \delta p_1)$ , ...,  $p_f$  and  $(p_f + \delta p_f)$ .

The number of systems (phase points) in this volume element  $(\delta q_1 \dots \delta q_f, \delta p_1 \dots \delta p_f)$  changes as the coordinates and momenta of the system vary with time.

$\rho$  is the density of phase points.

Number of phase points in this volume element, at any instant,  $t$ , is

$$\delta N = \rho \delta\Gamma = \rho \delta q_1 \dots \delta q_f, \delta p_1 \dots \delta p_f.$$

Change in number of phase points in the volume element per unit time is

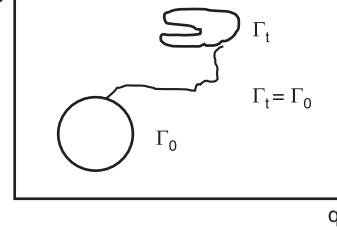
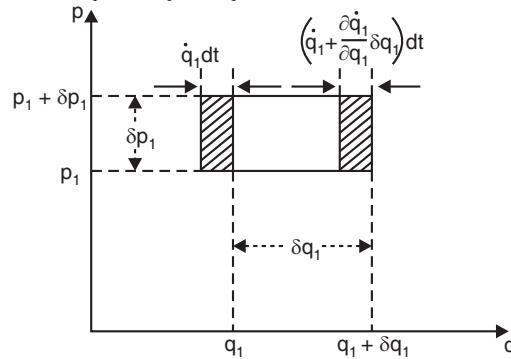


Fig. 43.8

$$\frac{d}{dt}(\delta N) = \frac{d}{dt}(\rho \delta \Gamma) = \dot{\rho} \delta \Gamma = \dot{\rho} \delta q_1 \dots \delta q_f \delta p_1 \dots \delta p_f.$$

In a time  $dt$ , the change in number of phase points within the volume of phase space  
 $= \dot{\rho} dt \delta q_1 \dots \delta q_f \delta p_1 \dots \delta p_f.$

This change is due to the number of systems entering and leaving this volume in time  $dt$ . We can construct a conceptual Euclidean space of  $2f$  dimensions. Consider two faces perpendicular to the  $q_1$ -axis which are located at  $q_1$  and  $q_1 + \delta q_1$  (Fig. 43.9).



**Fig. 43.9**

Number of phase points entering the first face in time  $dt$  is

$$\rho \dot{q}_1 dt \delta q_2 \dots \delta q_f \delta p_1 \dots \delta p_f. \quad \dots(1)$$

Here,  $\dot{q}_1$  is the velocity in the direction  $q_1$ .

Number of phase points leaving the face  $(q_1 + \delta q_1)$  in time  $dt$  is

$$\left( \rho + \frac{\partial \rho}{\partial q_1} \delta q_1 \right) \left( \dot{q}_1 + \frac{\partial \dot{q}_1}{\partial q_1} \delta q_1 \right) dt \delta q_2 \dots \delta q_f \delta p_1 \dots \delta p_f, \quad \dots(2)$$

Number of phase points leaving the face  $(q_1 + \delta q_1)$  in time  $dt$  is

$$\left[ \rho \dot{q}_1 + \left( \rho \frac{\partial \dot{q}_1}{\partial q_1} + \dot{q}_1 \frac{\partial \rho}{\partial q_1} \right) \delta q_1 \right] dt \delta q_2 \dots \delta q_f \delta p_1 \dots \delta p_f \quad \dots(3)$$

(Neglecting higher order terms)

Hence, the effective number of phase points entering the volume element in the direction  $q_1$  is

$$-\left( \rho \frac{\partial \dot{q}_1}{\partial q_1} + \dot{q}_1 \frac{\partial \rho}{\partial q_1} \right) dt \delta q_1 \dots \delta q_f \delta p_1 \dots \delta p_f \quad \dots(4)$$

[Eq. (3) – Eq. (1) gives Eq. (4)]

Similarly, for  $p_1$  coordinate, we have

$$-\left( \rho \frac{\partial \dot{p}_1}{\partial p_1} + \dot{p}_1 \frac{\partial \rho}{\partial p_1} \right) dt \delta q_1 \dots \delta q_f \delta p_1 \dots \delta p_f \quad \dots(5)$$

The net increase of number of systems in time  $dt$  in this volume of phase space is then obtained by summing the net number of systems entering the volume through all the faces labelled by  $q_1 \dots q_f$  and  $p_1 \dots p_f$ .

$$\therefore \frac{d}{dt}(\delta N) = -\sum_{i=1}^f \left\{ \rho \left( \frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) + \left( \frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right) \right\} dt \delta q_1 \dots \delta q_f \delta p_1 \dots \delta p_f \quad \dots(6)$$

We have,

$$\frac{d}{dt}(\delta N) = \frac{\partial \rho}{\partial t} dt \delta q_1 \dots \delta q_f \delta p_1 \dots \delta p_f.$$

$$\therefore \frac{\partial \rho}{\partial t} dt \delta q_1 \dots \delta q_f \delta p_1 \dots \delta p_f = - \sum_{i=1}^f \left\{ \rho \left( \frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) + \left( \frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right) \right\} dt \delta q_1 \dots \delta q_f \delta p_1 \dots \delta p_f$$

or

$$\frac{\partial \rho}{\partial t} = - \sum_{i=1}^f \left\{ \rho \left( \frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) + \left( \frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right) \right\} \quad \dots(7)$$

Hamilton's canonical equations of motion are:

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \quad \text{and} \quad \dot{p}_i = - \frac{\partial H}{\partial q_i}$$

$$\frac{\partial \dot{q}_i}{\partial q_i} = \frac{\partial}{\partial q_i} \frac{\partial H}{\partial p_i} = \frac{\partial^2 H}{\partial q_i \partial p_i}$$

$$\frac{\partial \dot{p}_i}{\partial p_i} = - \frac{\partial}{\partial p_i} \frac{\partial H}{\partial q_i} = - \frac{\partial^2 H}{\partial q_i \partial p_i}$$

$$\therefore \sum_{i=1}^f \left( \frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) = 0. \quad \dots(8)$$

Now, Eq. (7) can be written as

$$\left( \frac{\partial \rho}{\partial t} \right)_{q,p} = - \sum_{i=1}^f \left( \frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right) \quad \dots(9)$$

This result is called *Liouville's equation*.

Equation (9) can be written as

$$\left( \frac{\partial \rho}{\partial t} \right)_{q,p} + \sum_i \frac{\partial \rho}{\partial q_i} \frac{dq_i}{dt} + \sum_i \frac{\partial \rho}{\partial p_i} \frac{dp_i}{dt} = 0 \quad \dots(10)$$

If  $\rho$  is a function of  $q, p$  and  $t$ , and  $q, p$  are functions of  $t$ , then total differential coefficient of  $\rho$  with respect to  $t$  is

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial q} \cdot \frac{dq}{dt} + \frac{\partial \rho}{\partial p} \cdot \frac{dp}{dt}$$

Generalising this for all  $f$ 's and  $f$ 's, we get

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \sum_i \frac{\partial \rho}{\partial q_i} \frac{dq_i}{dt} + \sum_i \frac{\partial \rho}{\partial p_i} \frac{dp_i}{dt} \quad \dots(11)$$

Comparing Eqs. (10) and (11), we get

$$\frac{d\rho}{dt} = 0.$$

The total time derivative of  $\rho (q, p, t)$  is zero.

In other words, the density of phase points is an integral of motion.

### Part II: Principle of conservation of extension in phase space.

We know that  $\delta N = \rho \delta \Gamma$

or

$$\frac{d}{dt}(\delta N) = \frac{d\rho}{dt} \delta \Gamma + \rho \frac{d}{dt}(\delta \Gamma). \quad \dots(12)$$

Since the number of phase points  $\delta N$  in a given region of the phase space must remain fixed,

$$\begin{aligned} \frac{d}{dt}(\delta N) &= 0 \\ \therefore \frac{dp}{dt}(\delta\Gamma) + p \frac{d}{dt}(\delta\Gamma) &= 0 \quad \dots(13) \\ \text{or} \quad p \frac{d}{dt}(\delta\Gamma) &= 0. \quad \left( \text{because } \frac{dp}{dt} = 0 \right) \end{aligned}$$

Since  $p \neq 0$ , we have

$$\frac{d}{dt}(\delta\Gamma) = 0. \quad \dots(14)$$

## 43.5 BASIC CONCEPTS

### 1. Macroscopic and Microscopic Descriptions

We may mentally isolate any finite portion of matter from its local surroundings. We call such a portion the *system*. Everything outside the system which has a direct bearing on its behaviour we call the *environment*. The system and its environment taken together are usually referred to as the *universe*. Let us consider an example of a system and its surroundings. Consider some dilute gas of fixed mass enclosed in a cylinder provided with a frictionless piston. The gas is the system. The walls of the cylinder and the piston form the system boundary. The open atmosphere plays the role of the surroundings.

We want to determine the behaviour of the system by finding how it interacts with its environment. A system such as a dilute gas can be studied from (i) microscopic point of view or (ii) macroscopic point of view. Here the term ‘*micro*’ stands for small scale and “*macro*” stands for large scale. Molecules form the ‘*micro*’ systems, while any actual mass of gas is a ‘*macro*’ system. A macroscopic system is thus made up of an assembly of a large number of microscopic ones. In the microscopic description, we consider quantities that describe the atoms and molecules that make up the system, their speeds, energies, masses, behaviour during collision etc. These quantities form the basis for the science of *statistical mechanics*. The variables used in microscopic description are (i) very large in number (ii) not perceptible to our organic senses and (iii) not easily measurable by simple experimental procedures. On the other hand, in the macroscopic description of a dilute gas, we focus our attention on the gas as a whole. The variables used in the macroscopic description are: (i) few in number (ii) perceptible to our senses and (iii) easily measurable by simple experimental procedures. *Macroscopic quantities* are gross properties of the system measured by laboratory operations. The collective behaviour of a macroscopic system may be represented by some parameters like pressure ( $P$ ), volume ( $V$ ), temperature ( $T$ ), internal energy ( $U$ ) and entropy ( $S$ ). These parameters are always measurable through direct experiments. These macroscopic quantities form the basis for the science of *thermodynamics*. The microscopic interactions will interpret the properties of macroscopic system.

For any system, the macroscopic and the microscopic quantities must be related because they are simply different ways of describing the same situation. In particular, we should be able to express the macroscopic quantities in terms of the microscopic quantities. For example, a very large number of collisions of molecules with the walls give an average pressure  $P$ . Similarly, the temperature  $T$  of a gas may be related to the average kinetic energy of translation of the molecules. Thus, a very large number of position and momentum coordinates of all the molecules in microscopic description give only two macroscopic variables  $P$  and  $T$ . The macroscopic description is therefore, much simpler which is used in the analysis of thermodynamic properties of the systems. We can express the laws of thermodynamics quantitatively in the language of statistical mechanics.

## 2. Probability

The probability of an event may be defined as the ratio of the number of cases in which the event occurs to the total number of cases. That is,

$$\text{the probability of the event} = \frac{\text{Number of cases in which the event occurs}}{\text{Total number of cases}}$$

Suppose we toss a coin a large number of times and count the number of times the ‘head’ or ‘tail’ is uppermost. Experience shows that the ‘heads’ and ‘tails’ will be uppermost equal number of times. That is,

$$P(\text{head}) = 1/2 \text{ and } P(\text{tail}) = 1/2.$$

## 3. Thermodynamic Probability

*The thermodynamic probability of a particular macrostate is defined as the number of microstates corresponding to that macrostate.*

It is represented by  $W$ .

Consider two cells in phase space represented by  $i$  and  $j$  and four molecules  $a, b, c$  and  $d$ . Let  $N_i$  and  $N_j$  be the number of molecules in the cells  $i$  and  $j$  respectively. Then the possible macrostates are five in number (Fig. 43.10a).

$N_i$	4	3	2	1	0
$N_j$	0	1	2	3	4

**Fig. 43.10a**

In general, to each of these five macrostates there corresponds a different number of microstates. Let us consider the microstates corresponding to the macrostate  $N_i = 3, N_j = 1$ . The number of microstates corresponding to the macrostate  $N_i = 3, N_j = 1$  is four, as shown in Fig. 43.10b.

Cell i	abc	abd	acd	bcd
Cell j	d	c	b	a

**Fig. 43.10b**

Therefore, the thermodynamic probability for the macrostate,  $N_i = 3, N_j = 1$  is 4. That is  $W = 4$ .

## 4. Boltzmann's Theorem on Entropy and Probability

Boltzmann discovered a relation between entropy (a thermodynamical quantity) and probability (a statistical quantity). Boltzmann started from a simple idea that the equilibrium state of the system is the state of maximum probability. That is, the probability of the system in equilibrium state is maximum. But from the thermodynamic point of view, the equilibrium state of the system is the state of maximum entropy. If the system is not in equilibrium, then changes take place within the system until the equilibrium state or the state of maximum entropy is reached. Thus, in equilibrium state both the entropy and thermodynamical probability have their maximum values. As equilibrium state is the state of maximum entropy and maximum probability, Boltzmann concluded that entropy is a function of probability. That is,

$$S = f(W) \quad \dots(1)$$

Here,  $S$  is entropy and  $W$  is the thermodynamical probability of the state. Let us consider two separate systems having entropies  $S_1$  and  $S_2$  and thermodynamic probabilities  $W_1$  and  $W_2$ .

$$\text{Then} \quad S_1 = f(W_1) \text{ and } S_2 = f(W_2) \quad \dots(2)$$

The total entropy of the two systems is

$$S_1 + S_2 = f(W_1) + f(W_2) \quad \dots(3)$$

But the thermodynamic probability of the two systems taken together is  $W_1 W_2$ .

$$f(W_1 W_2) = f(W_1) + f(W_2) = S_1 + S_2 \quad \dots(4)$$

If this relation is to be satisfied,  $f(W)$  must be a logarithmic function of  $W$ .

$$\therefore f(W) = k \log W$$

$$\text{or } S = k \log W. \quad \dots(5)$$

### 5. Fundamental Postulates of Statistical Mechanics

1. Any gas may be considered to be composed of molecules that are in motion and behave like very small elastic spheres.
2. All the cells in the phase space are of equal size.
3. All accessible microstates corresponding to possible macrostates are equally probable.  
This is called the *postulate of equal a priori probability*.
4. The equilibrium state of a gas corresponds to the macrostate of maximum probability.
5. *The total number of molecules is constant.*

### 6. Statistical Equilibrium

Boltzmann canonical principle is applied to determine the equilibrium state of the system. According to this principle, *Equilibrium state of a system is that which is most probable*.

Consider an isolated system composed of a large number  $N$  of particles, in which each particle has available to it several states with energies  $E_1, E_2, E_3, \dots$ . At a particular time the particles are distributed among the different states, so that  $n_1$  particles have energy  $E_1$ ;  $n_2$  particles have energy  $E_2$ ; and so on. The total number of particles is

$$N = n_1 + n_2 + n_3 + \dots = \sum_i n_i \quad \dots(1)$$

We assume that the total number of particles remains constant for all processes occurring in the system.

The total energy of the system is

$$E = n_1 E_1 + n_2 E_2 + \dots = \sum_i n_i E_i \quad \dots(2)$$

If the system is isolated, the total energy  $E$  must be constant. However, as a result of their mutual interactions and collisions, the distribution of the particles among the available energy states may be changing. For example in a gas, a fast molecule may collide with a slow one; after the collision, the fast molecule may have slowed down and the slow one may have sped up. Or an excited atom may collide inelastically with another atom, with a transfer of its excitation energy into kinetic energy of both atoms. Hence, in both examples, the particles after the collision are in different states. In other words, the numbers  $n_1, n_2, n_3, \dots$  which give the distribution of the  $N$  particles among the available energy states, may be changing. It is reasonable to assume that for each macroscopic state of a system of particles there is a distribution which is more favoured than any other. In other words, we may say that *given the physical conditions of the system of particles, there is a most probable distribution*. When this distribution is achieved, the system is said to be in *statistical equilibrium*.

## MAXWELL-BOLTZMANN STATISTICS

### 43.6 MAXWELL-BOLTZMANN DISTRIBUTION LAW

Consider a system of  $N$  distinguishable molecules of a gas. Suppose  $N_1$  of them have energy  $E_1$ ,  $n_2$  have energy  $E_2$ , ...  $n_i$  have energy  $E_i$  and so on. Thus the entire assembly of molecules can be divided into different energy states with energies  $E_1, E_2, E_3, \dots E_i$  and having  $n_1, n_2, n_3, \dots n_i$  molecules.

- (1) The total number of molecules  $N$  is constant. Hence

$$\begin{aligned} N &= n_1 + n_2 + n_3 + \dots + n_i + \dots = \text{constant} \\ \text{or } \delta N &= \delta n_1 + \delta n_2 + \delta n_3 + \dots + \delta n_i + \dots = 0 \\ \text{i.e., } \sum_i \delta n_i &= 0 \end{aligned} \quad \dots(1)$$

- (2) The total energy  $E$  of the gas molecules is constant. Hence

$$\begin{aligned} E &= E_1 n_1 + E_2 n_2 + E_3 n_3 + \dots + E_i n_i + \dots = \text{constant} \\ \text{or } \delta E &= E_1 \delta n_1 + E_2 \delta n_2 + E_3 \delta n_3 + \dots + E_i \delta n_i + \dots = 0 \\ \text{i.e., } \sum_i E_i \delta n_i &= 0 \end{aligned} \quad \dots(2)$$

(3) Suppose there are  $g_i$  cells with the energy  $E_i$ . The total number of ways in which  $n_i$  molecules can have the energy  $E_i$  is  $(g_i)^{n_i}$ . Hence the total number of ways in which  $N$  molecules can be distributed among the various energies is

$$W_1 = (g_1)^{n_1} (g_2)^{n_2} (g_3)^{n_3} \dots (g_i)^{n_i} \dots$$

The number of ways in which the groups of  $n_1, n_2, n_3 \dots n_i \dots$  particles can be chosen from  $N$  particles is given by

$$W_2 = \frac{N!}{n_1! n_2! n_3! \dots}$$

The number of distinguishable ways in which  $N$  molecules can be distributed among the possible energy levels is

$$W = W_1 W_2 = \frac{N!}{n_1! n_2! n_3! \dots} (g_1)^{n_1} (g_2)^{n_2} (g_3)^{n_3} \dots \quad \dots(3)$$

The quantity  $W$  is called the *thermodynamic probability* for the system. For the most probable distribution,  $W$  is a maximum subject to the restriction that the total number of particles  $N$  and the total energy  $E$  are constants.

The natural logarithm of Eq. (3) is

$$\ln W = \ln N! - \sum n_i \ln g_i$$

By Stirling's theorem,  $\ln n! = n \ln n - n$

$$\ln W = N \ln N - N - \sum n_i \ln n_i + \sum n_i + \sum n_i \ln g_i$$

$$\ln W = N \ln N - \sum n_i \ln n_i + \sum n_i \ln g_i \quad \dots(4)$$

From Eq. (4), we have for maximum  $W$

$$\delta \ln W_{\max} = -\sum n_i \frac{1}{n_i} \delta n_i - \sum (\ln n_i) \delta n_i + \sum (\ln g_i) \delta n_i = 0$$

$$\text{or } -\sum (\ln n_i) \delta n_i + \sum (\ln g_i) \delta n_i = 0 \quad [\because \sum \delta n_i = 0] \quad \dots(5)$$

Eqs. (1) and (2) can be incorporated into Eq. (5) by making use of Lagrange's method of undetermined multipliers. Multiplying Eq. (1) by  $-\alpha$  and Eq. (2) by  $-\beta$  and adding to Eq. (5), we get

$$\sum (-\ln n_i + \ln g_i - \alpha - \beta E_i) \delta n_i = 0 \quad \dots(6)$$

$$\text{or } -\ln n_i + \ln g_i - \alpha - \beta E_i = 0$$

$$\text{or } n_i = g_i e^{-\alpha} e^{-\beta E_i} \quad \dots(7)$$

Eq. (7) is called Maxwell-Boltzmann *distribution law*.

**M-B distribution in terms of temperature.** It can be shown that  $\beta = 1/kT$  where  $k$  = Boltzmann's constant and  $T$  = absolute temperature of the gas.

$$\therefore n_i = g_i e^{-\alpha} e^{-(E_i/kT)}. \quad \dots(8)$$

### 43.6.1 Molecular Energies in an Ideal Gas

The M.B. distribution law is

$$n_i = g_i e^{-\alpha} e^{-(E_i/kT)} \quad \dots(1)$$

Consider an ideal gas that contains  $N$  molecules. Consider a continuous distribution of molecular energies instead of the discrete set  $E_1, E_2, E_3, \dots$ . Then Eq. (1) becomes

$$n(E) dE = g(E) e^{-\alpha} e^{-(E/kT)} dE \quad \dots(2)$$

Here  $n(E) dE$  represents the number of molecules having energies between  $E$  and  $E + dE$ .  $g(E) dE$  represents the number of states that have energies between  $E$  and  $E + dE$ .

Let us first find  $g(E) dE$ . A molecule of energy  $E$  has a momentum  $p$  whose magnitude is

$$p = \sqrt{2mE} = \sqrt{p_x^2 + p_y^2 + p_z^2} \quad \dots(3)$$

Each set of momentum components  $p_x, p_y, p_z$  specifies a different state of motion. Let us imagine a *momentum space* whose coordinate axes are  $p_x, p_y, p_z$  (Fig. 43.11). The number of momentum states available to a particle with a momentum whose magnitude is between  $p$  and  $p + dp$  is proportional to the volume of a spherical shell in momentum space of radius  $p$  and thickness  $dp$ . Volume of this spherical shell is  $4\pi p^2 dp$ . Hence

$$g(p) dp = B p^2 dp \quad \dots(4)$$

where  $B$  is some constant. Since each momentum magnitude  $p$  corresponds to a single energy  $E$ , the number of energy states  $g(E) dE$  between  $E$  and  $E + dE$  is the same as the number of momentum states  $g(p) dp$  between  $p$  and  $p + dp$ .

$$\therefore g(E) dE = B p^2 dp \quad \dots(5)$$

$$\text{But } p^2 = 2mE \text{ and } dp = \frac{m dE}{\sqrt{2mE}}$$

$$\therefore g(E) dE = \sqrt{2} m^{3/2} B \sqrt{E} dE \quad \dots(6)$$

Substituting for  $g(E) dE$  in Eq. (2),

$$n(E) dE = C \sqrt{E} e^{-(E/kT)} dE \quad \dots(7)$$

where

$$C = \sqrt{2} m^{3/2} B e^{-\alpha} = \text{a constant.}$$

To find  $C$  we make use of the normalization condition that the total number of molecules is  $N$ . Therefore,

$$N = \int_0^\infty n(E) dE = C \int_0^\infty \sqrt{E} e^{-(E/kT)} dE \quad \dots(8)$$

From a table of definite integrals we find that

$$\int_0^\infty \sqrt{x} e^{-ax} dx = \frac{1}{2a} \sqrt{\frac{\pi}{a}}$$

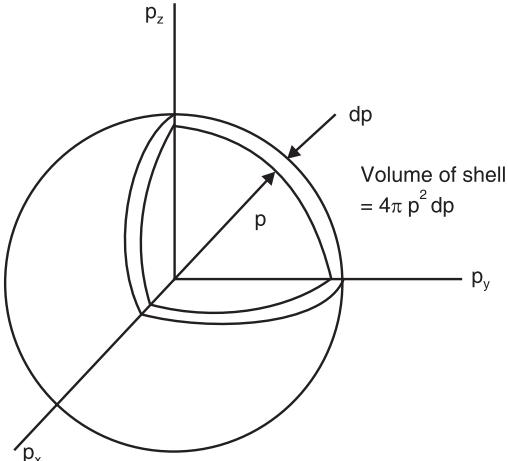


Fig. 43.11

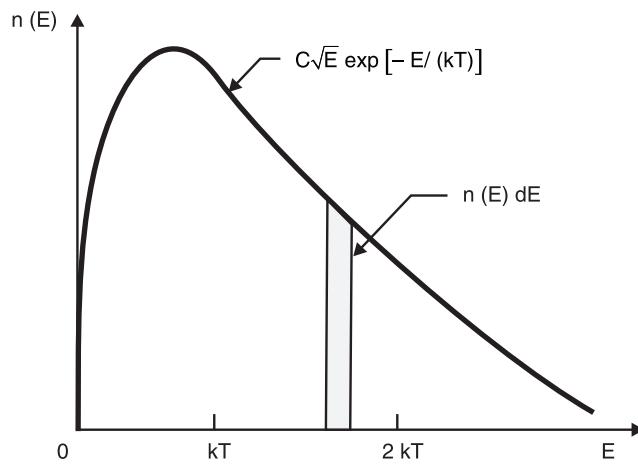
$$\therefore N = C \frac{1}{2(1/kT)} \sqrt{\frac{\pi}{(1/kT)}} = C \times \frac{\sqrt{\pi}}{2} \times (kT)^{\frac{3}{2}}$$

or

$$C = \frac{2\pi N}{(\pi kT)^{\frac{3}{2}}} \quad \dots(9)$$

$$\therefore n(E) dE = \frac{2\pi N}{(\pi kT)^{\frac{3}{2}}} \sqrt{E} e^{-E/kT} dE \quad \dots(10)$$

Eq. (10) is plotted in Fig. 43.12 in terms of  $kT$ .



**Fig. 43.12**

The total internal energy of the system is

$$E = \int_0^{\infty} E n(E) dE = \frac{2\pi N}{(\pi kT)^{\frac{3}{2}}} \int_0^{\infty} E^{\frac{3}{2}} e^{-E/kT} dE$$

The value of the definite integral is  $\frac{3}{4}(kT)^2 \sqrt{\pi kT}$ .

$$\therefore E = \frac{2\pi N}{(\pi kT)^{\frac{3}{2}}} \times \frac{3}{4}(kT)^2 \sqrt{\pi kT} = \frac{3}{2} N k T \quad \dots(11)$$

The average energy of an ideal-gas molecule is  $E/N$ .

$$\therefore \bar{E} = \frac{3}{2} k T \quad \dots(12)$$

#### Maxwell-Boltzmann Velocity Distribution Law

Substitute  $E = 1/2 mv^2$  and  $dE = mv dv$  in Eq. (10). Then we get

$$n(v) dv = \frac{\sqrt{2} \pi N m^{\frac{3}{2}}}{(\pi kT)^{\frac{3}{2}}} v^2 e^{-mv^2/2kT} dv \quad \dots(13)$$

This equation represents the number of molecules with speeds between  $v$  and  $v + dv$  in an assembly of ideal gas containing  $N$  molecules at absolute temperature  $T$ . This formula is plotted in Fig. 43.13.

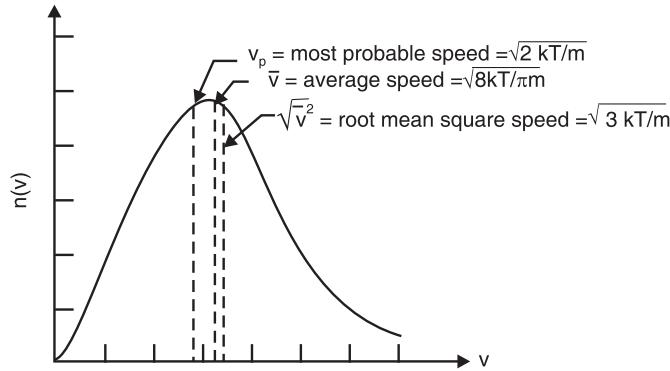


Fig. 43.13

## BOSE-EINSTEIN STATISTICS

### 43.7 BOSE-EINSTEIN DISTRIBUTION LAW

Consider an assembly of  $N$  bosons. They are identical and indistinguishable. No restriction is imposed as to the number of particles that may occupy a given cell. Let us now consider a box divided into  $g_i$  sections by  $(g_i - 1)$  partitions and  $n_i$  indistinguishable particles to be distributed among these sections. The permutations of  $n_i$  particles and  $(g_i - 1)$  partitions simultaneously is given by  $(n_i + g_i - 1)!$ . But this includes also the permutations of  $n_i$  particles among themselves and also  $(g_i - 1)$  partitions among themselves, as both these groups are internally indistinguishable. Hence the actual number of ways in which  $n_i$  particles are to be distributed in  $g_i$  sublevels is

$$\frac{(n_i + g_i - 1)!}{n_i! (g_i - 1)!}.$$

Therefore, the total number of distinguishable and distinct ways of arranging  $N$  particles in all the variable energy states is given by

$$W = \prod \frac{(n_i + g_i - 1)!}{n_i! (g_i - 1)!} \quad \dots(1)$$

$n_i$  and  $g_i$  are large numbers. Hence we may neglect 1 in the above expression.

$$\therefore W = \prod \frac{(n_i + g_i)!}{n_i! g_i!} \quad \dots(2)$$

$$\ln W = \sum [\ln(n_i + g_i)! - \ln n_i! - \ln g_i!]$$

As  $n_i$  and  $g_i$  are large numbers, we can use Stirling approximation.

$$\ln W = \sum (n_i + g_i) \ln(n_i + g_i) - n_i \ln n_i - g_i \ln g_i \quad \dots(3)$$

Here,  $g_i$  is not subject to variation and  $n_i$  varies continuously.

For most probable distribution,  $\delta \ln W_{\max} = 0$ .

Hence, if the  $W$  of Eq. (3) represents a maximum,

$$\delta \ln W_{\max} = \sum [\ln(n_i + g_i) - \ln n_i] \delta n_i = 0 \quad \dots(4)$$

The total number of particles and total energy are constants.

$$\therefore \sum \delta n_i = 0 \quad \dots(5)$$

$$\sum E_i \delta n_i = 0 \quad \dots(6)$$

Multiplying Eq. (5) by  $-\alpha$ , Eq. (6) by  $-\beta$  and adding to Eq. (4),

$$\sum [\ln(n_i + g_i) - \ln n_i - \alpha - \beta E_i] \delta n_i = 0$$

The variations  $\delta n_i$  are independent of each other. Hence we get

$$\ln\left(\frac{n_i + g_i}{n_i}\right) - \alpha - \beta E_i = 0$$

or

$$n_i = \frac{g_i}{(e^{\alpha + \beta E_i}) - 1} \quad \dots(7)$$

or

$$n_i = \frac{g_i}{(e^\alpha e^{E_i/kT}) - 1} \quad \dots(8)$$

This is called Bose-Einstein distribution law.

#### 43.7.1 Application of Bose Einstein Statistics – Planck Radiation Formula

**Photon Gas.** The radiation from a black body at a certain temperature  $T$ , in thermal equilibrium, consists of light quanta or photons of energy  $h\nu$ .  $\nu$  is different for different photons of different energies. These photons move in all possible directions, without collision among themselves, with the speed of light and, therefore, possess momentum  $h\nu/c$  ( $\equiv h/\lambda$ ) also.

*The aggregate of photons in a cavity (called the black body) thus behaves as a gas, the photon gas.*

##### Properties of photon gas

- (i) Photons are particles of zero rest mass.
- (ii) Photons are bosons (*i.e.*, the particles with spin 1) and have two modes of propagation (due to a clockwise and a counterclockwise polarisation).
- (iii) Photons are mutually indistinguishable.
- (iv) The number of photons is not conserved because at the instant of emission a new quantum is formed. If a photon of frequency  $\nu$  is absorbed by the walls of the enclosure, several photons may be emitted provided the total energy of the system is conserved.
- (v) The photons do not interact among themselves.

We shall study the application of the B-E distribution to photons and obtain Planck's radiation formula. Consider a box containing electromagnetic radiation (*i.e.*, photons) at a temperature  $T$ . These photons would be absorbed and re-emitted by the container walls. Under thermal equilibrium, the energy distribution can be determined by the Bose-Einstein statistics. However, there is an important difference between the photons and the other particles, namely, the photon number is not conserved. This is because the walls can absorb a photon of energy  $E$  and emit two photons of energy  $E_1$  and  $E_2$  such that  $E_1 + E_2 = E$ . Thus, we have only the energy restriction and no restriction on the photon number.

The B. E. distribution function is

$$f(E) = \frac{1}{(e^{\alpha} e^{E/kT}) - 1} \quad \dots(1)$$

The non-conservation of photons means that  $\alpha = 0$ .

Therefore, the B-E distribution function for photons is

$$f(\nu) = \frac{1}{e^{h\nu/kT} - 1} \quad (\because E = h\nu) \quad \dots(2)$$

The number of independent standing waves per unit volume in the frequency interval from  $\nu$  to  $\nu + d\nu$  is

$$G(v) dv = \frac{8\pi v^2}{c^3} dv \quad \dots(3)$$

The energy-density of radiation in the frequency range  $v$  to  $v + dv$  is

$$\begin{aligned} u(v) dv &= hv G(v) f(v) dv \\ \therefore u(v) dv &= \frac{8\pi h}{c^3} \left( \frac{v^3 dv}{e^{hv/kT} - 1} \right) \end{aligned} \quad \dots(4)$$

This is Planck radiation formula in terms of frequency.

It can be given in terms of wavelength also.

We have,  $c = v\lambda$  and  $dv = -\frac{c}{\lambda^2} d\lambda$ .

$$\therefore u(\lambda) d\lambda = \frac{8\pi hc}{\lambda^5 [e^{hc/\lambda kT} - 1]} d\lambda \quad \dots(5)$$

Eq. (5) is the Planck's formula for black body radiation in terms of wavelength.

#### 43.7.2 Derivation of Wien's Law from Planck's Law

Planck's radiation law is

$$E_\lambda d\lambda = \frac{8\pi hc \lambda^{-5}}{(e^{hc/\lambda kT} - 1)} d\lambda \quad \dots(1)$$

When  $\lambda$  is small,  $e^{hc/\lambda kT}$  is large when compared to 1.

So Eq. (1) reduce to

$$E_\lambda d\lambda = 8\pi hc \lambda^{-5} e^{-hc/\lambda kT} d\lambda \quad \dots(2)$$

This is Wien's law.

Thus Planck's law reduces to Wien's law for shorter wavelengths.

#### 43.7.3 Derivation of Rayleigh - Jean's Law from Planck's Law

Planck's radiation law is

$$E_\lambda d\lambda = \frac{8\pi hc \lambda^{-5}}{(e^{hc/\lambda kT} - 1)} d\lambda \quad \dots(1)$$

When  $\lambda$  is large,  $e^{hc/\lambda kT} \approx \left(1 + \frac{hc}{\lambda kT}\right)$

Hence Planck's law reduces to

$$\begin{aligned} E_\lambda d\lambda &= \frac{8\pi hc \lambda^{-5}}{\left(1 + \frac{hc}{\lambda kT}\right) - 1} = 8\pi \left(\frac{hc \lambda^{-5}}{\lambda kT}\right) d\lambda \\ \therefore E_\lambda d\lambda &= 8\pi kT \lambda^{-4} d\lambda \end{aligned} \quad \dots(2)$$

This is *Rayleigh - Jean's formula*.

Thus Planck's formula reduces to Rayleigh Jean's formula for longer wavelengths.

## FERMI-DIRAC STATISTICS

### 43.8 FERMI-DIRAC DISTRIBUTION LAW

F-D statistics is obeyed by indistinguishable particles of half-integral spin that have antisymmetric wave functions and obey Pauli exclusion principle. Consider  $N$  fermions with the total energy  $E$ . Suppose that  $n_i$  particles occupy the first energy level with energy  $E_1$ ,  $n_2$  particles occupy the second energy level with energy  $E_2$  and so on. Let us now find out the total number of ways in which  $n_i$  particles can be distributed in  $g_i$  cells having the same energy  $E_i$ .

The number of distinguishable arrangements of  $n_i$  particles in  $g_i$  cells is

$$\frac{g_i!}{n_i!(g_i-n_i)!}$$

The total number of Eigen states for the whole system is given by

$$W = \prod g_i! \quad \dots(1)$$

Taking the natural logarithm of both sides,

$$\ln W = \sum [\ln g_i! - \ln n_i! - \ln (g_i - n_i)!]$$

Applying Stirling's approximation,

$$\begin{aligned} \ln W &= \sum [g_i \ln g_i - g_i - n_i \ln n_i + n_i - (g_i - n_i) \ln (g_i - n_i) + (g_i - n_i)] \\ \ln W &= \sum [g_i \ln g_i - n_i \ln n_i - (g_i - n_i) \ln (g_i - n_i)] \end{aligned} \quad \dots(2)$$

Here  $g_i$  is not subject to variation and  $n_i$  varies continuously.

For most probable distribution,  $\delta \ln W_{max} = 0$ .

$$\delta \ln W_{max} = \sum [-\ln n_i + \ln (g_i - n_i)] \delta n_i = 0 \quad \dots(3)$$

$$\text{But } \sum \delta n_i = 0 \quad \dots(4)$$

$$\text{and } \sum E_i \delta n_i = 0 \quad \dots(5)$$

Multiplying Eq. (4) by  $-\alpha$ , Eq. (5) by  $-\beta$  and adding to Eq. (3),

$$\sum [-\ln n_i + \ln (g_i - n_i) - \alpha - \beta E_i] \delta n_i = 0$$

As the variations  $\delta n_i$  are independent of each other, we get

$$\ln \left( \frac{g_i - n_i}{n_i} \right) - \alpha - \beta E_i = 0$$

$$\text{or } n_i = \frac{g_i}{(e^\alpha e^{\beta E_i}) + 1} \quad \dots(6)$$

This is known as Fermi-Dirac Distribution law.

$\beta = 1/kT$  and  $\alpha = -E_F/kT$  where  $E_F$  is called Fermi energy.  $E_F = -\alpha kT$ .

$$n_i = \frac{g_i}{(e^\alpha e^{E_i/kT}) + 1} \quad \dots(7)$$

**Definition of Fermi energy :** At the absolute zero of temperature the maximum kinetic energy that a free electron can have is called the Fermi energy,  $E_F$ .

$$\therefore n_i = \frac{g_i}{[e^{(E_i - E_F)/kT}] + 1}$$

### 43.8.1 Application of Fermi Dirac Statistics - Electron Gas

**Electron Gas.** The valence electrons in a metal are free to roam in the ionic array of positive fixed nuclei. These mobile electrons in a metal behave, more or less as particles of a perfect gas and form what is known as *electron gas* (or Fermi-gas). These electrons, however, cannot come out of the metal surface because of surface potential barrier (work function).

#### Expression for Fermi Energy

The “free electron gas” in a solid obeys Fermi-Dirac statistics. Suppose in an assemblage of fermions, there are  $M(E)$  allowed quantum states in an energy range between  $E$  and  $E + dE$  and  $N(E)$  is the number of particles in the same range. Then  $N(E)$  quantum states are filled and  $M(E) - N(E)$  are vacant. The F-D distribution function  $f(E)$  is defined as :

$$f(E) = \frac{N(E)}{M(E)} = \frac{1}{1+e^{(E-E_F)/kT}} \quad \dots(1)$$

$N(E) / M(E)$  represents the fraction of the possible quantum states which is occupied. The distribution of electrons among the levels is usually described by the *distribution function*  $f(E)$ . It is defined as the probability that the level  $E$  is occupied by an electron. Thus, if the level is certainly empty, then  $f(E) = 0$ . If the level is certainly full, then  $f(E) = 1$ . In general  $f(E)$  has a value between zero and unity.

The distribution function for electrons at  $T = 0\text{ K}$  has the form

$$\begin{aligned} f(E) &= 1 \text{ when } E < E_F \\ \text{and} \quad f(E) &= 0 \text{ when } E > E_F \end{aligned} \quad \dots(2)$$

That is, all levels below  $E_F$  are completely filled and all levels above  $E_F$  are completely empty. This function is plotted in Fig. 43.14 which shows the discontinuity at the Fermi-energy. As the temperature rises,  $f(E)$  changes from 1 to 0 more and more gradually as shown in Fig. 43.14.

$$\text{For } E = E_F, f(E) = \frac{1}{1+e^0} = \frac{1}{2}, \text{ at all temperatures.}$$

Thus, the probability of finding an electron with energy equal to the Fermi-energy in a metal is  $1/2$  at any temperature.

Since the electrons are confined inside the crystal, their wave properties will limit the energy values which they may have. Let  $g(E) dE$  be the number of quantum states available to electrons with energies between  $E$  and  $E + dE$ . It can be shown that

$$g(E) dE = \frac{8\sqrt{2}\pi V m^{3/2}}{h^3} \sqrt{3} dE \quad \dots(3)$$

where  $m$  is the mass of the electron and  $V$  is the volume of the electron gas.

We can calculate the Fermi-energy  $E_F$  by filling up the energy states in the metal sample with the  $N$  free electrons it contains in order of increasing energy, starting from  $E = 0$ . The highest state to be filled will then have the energy  $E = E_F$  by definition. The number of electrons that can have the same energy  $E$  is equal to the number of states that have this energy, since each state is limited to one electron. Hence,

$$N = \int_0^{E_F} g(E) dE = \frac{8\sqrt{2}\pi V m^{3/2}}{h^3} \int_0^{E_F} E^{1/2} dE$$

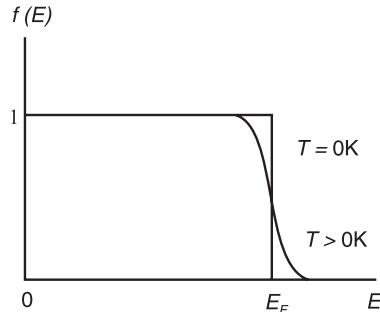


Fig. 43.14

$$\begin{aligned}
 &= \frac{16\sqrt{2}\pi V m^{3/2}}{3h^3} E_F^{3/2} \\
 \therefore E_F &= \left(\frac{h^2}{2m}\right) \left(\frac{3N}{8\pi V}\right)^{2/3} \quad \dots(4)
 \end{aligned}$$

The quantity  $N/V$  is the density of free electrons.

$N/V$  represents the number of free electrons per unit volume of the metal.

An effective temperature of the electron-gas, known as the *Fermi-temperature*, is defined by

$$T_F = E_F/k.$$

### 43.9 COMPARISON OF THE THREE DISTRIBUTION LAWS

Let us define a quantity  $f(E_i) = \frac{n_i}{g_i}$ .

$f(E_i)$  is called the ‘occupation index’ of a state of energy  $E_i$ .

$$f_{MB}(E_i) = \frac{1}{e^\alpha e^{E_i/kT}}$$

$$f_{BE}(E_i) = \frac{1}{(e^\alpha e^{E_i/kT}) - 1}$$

$$f_{FD}(E_i) = \frac{1}{(e^\alpha e^{E_i/kT}) + 1}$$

When  $E \gg kT$ ,  $f_{BE}$  and  $f_{FD}$  approach  $f_{MB}$ . In general,

$$f(E) = \frac{1}{\exp[\alpha + E/(kT)] + \delta}$$

is the probability of occupation of a single state at energy  $E$  and  $\delta$  is +1 for F-D, 0 for M-B, and -1 for B-E statistics. We will refer to  $f(E)$  as the *occupation probability* or the *distribution function*.

#### Comparison of MB, BE, and FD Distribution Functions

MB	BE	FD
(1) Holds for distinguishable particles; approximations of BE and FD distributions at $E \gg kT$ .	Holds for indistinguishable particles not obeying Pauli's exclusion principle	Holds for indistinguishable particles obeying the exclusion principle.
(2) $f(E) = A e^{-E/kT}$	$f(E) = \frac{1}{e^\alpha e^{E/kT} - 1}$	$f(E) = \frac{1}{e^{(E-E_F)/kT} + 1}$
(3) Applies to common gases at normal temperatures.	Applies to photon gas, phonon gas.	Applies to electron gas in metals.

### EXERCISE

1. What is phase space? (MKU, 2002)
2. What are ensembles? (MKU, 2003)
3. State and prove Liouville's theorem.

4. Derive the Maxwell-Boltzmann law of distribution of energy.
5. Applying Maxwell-Boltzmann distribution law, show that the internal energy of an ideal monoatomic gas depends only on its temperature.
6. Derive Bose-Einstein distribution law. Apply it to derive Planck's law of radiation.
7. (a) Derive the Fermi Dirac Statistical distribution law.  
(b) Compare the three statistics. (MKU, 2004)
8. Define Fermi energy of a metal and obtain an expression for it. Prove that Fermi energy is independent of the size of the metal. (Calicut University, 2012)

## MATHEMATICAL PHYSICS

**After reading this chapter, you should be able to**

- ◆ Define beta and gamma functions and list a select few of their properties which are of immediate utility in the study of quantum mechanics
- ◆ Solve Bessel, Legendre, Hermite, and Laguerre differential equations which play an important role in quantum mechanics
- ◆ Prove Gauss, Green, and Stokes theorems which serve as powerful tools in theoretical physics
- ◆ List the important properties of matrices and their uses in the matrix formulation of quantum mechanics.

## THE BETA, GAMMA FUNCTIONS

### 44.1 INTRODUCTION

#### Definitions

(a) **The Beta Function.** The Beta function  $\beta(m, n)$  is defined by the definite integral

$$\beta(m, n) = \int_0^1 x^{m-1} (1-x)^{n-1} dx \quad \dots(1)$$

where  $m$  and  $n$  are positive.

(b) **The Gamma Function.** The Gamma function  $\Gamma n$  is defined by the definite integral,

$$\Gamma n = \int_0^\infty e^{-x} x^{n-1} dx, \quad \dots(2)$$

where  $n$  is positive. The restriction on  $n$  is necessary to avoid divergence of the integral.

#### Symmetry Property of Beta Function

**To show that  $\beta(m, n) = \beta(n, m)$**

$$\text{By definition, } \beta(m, n) = \int_0^1 x^{m-1} (1-x)^{n-1} dx \quad \dots(1)$$

Put

$x = 1-y$ .  $\therefore dx = -dy$ . Then,

$$\begin{aligned} \beta(m, n) &= \int_0^1 (1-y)^{m-1} y^{n-1} dy \\ &= \int_0^1 x^{n-1} (1-x)^{m-1} dx = \beta(n, m) \end{aligned} \quad \dots(2)$$

i.e., Beta function  $\beta(m, n)$  is symmetric with respect to  $m$  and  $n$ .

### 44.2 EVALUATION OF BETA FUNCTION

$$\text{By definition, } \beta(m, n) = \int_0^1 x^{m-1} (1-x)^{n-1} dx \begin{cases} m > 0 \\ n > 0 \end{cases}$$

Integrating by parts keeping  $(1-x)^{n-1}$  as first function, we have

$$\begin{aligned} \beta(m, n) &= \left[ (1-x)^{n-1} \frac{x^m}{m} \right]_0^1 + \int_0^1 (n-1)(1-x)^{n-2} \frac{x^m}{m} dx \\ &= \frac{n-1}{m} \int_0^1 (1-x)^{n-2} x^m dx \end{aligned}$$

Integrating again by parts, we get

$$\beta(m, n) = \frac{(n-1)(n-2)}{m(m+1)} \int_0^1 (1-x)^{n-3} x^{m+1} dx$$

Continuing the process of integration by parts and assuming that  $n$  is a positive integer, we get

$$\begin{aligned} \beta(m, n) &= \frac{(n-1)(n-2) \dots 2 \cdot 1}{m(m+1) \dots (m+n-2)} \int_0^1 x^{m+n-2} dx \\ &= \frac{(n-1)(n-2) \dots 2 \cdot 1}{m(m+1) \dots (m+n-2)} \left[ \frac{x^{m+n-1}}{m+n-1} \right]_0^1 \\ &= \frac{(n-1)!}{m(m+1) \dots (m+n-2)(m+n-1)} \end{aligned} \quad \dots(1)$$

Again, if  $m$  is also a positive integer, then

$$\beta(m, n) = \frac{(n-1)!(m-1)!}{(m+n-1)!} \quad \dots(2)$$

In case  $m$  alone is a positive integer, then in view of the symmetry property  $\beta(m, n) = \beta(n, m)$ , we have

$$\beta(m, n) = \frac{(m-1)!}{n(n+1)\dots(n+m-1)} \quad \dots(3)$$

### Other Forms of the Beta Function

$$\text{By definition, } \beta(m, n) = \int_0^1 x^{m-1} (1-x)^{n-1} dx \quad \dots(1)$$

$$(a) \text{ Put } x = \frac{y}{1+y} \therefore dx = \frac{dy}{(1+y)^2} \text{ and } 1-x = \frac{1}{1+y}$$

Therefore Eq. (1) becomes

$$\begin{aligned} \beta(m, n) &= \int_0^\infty \frac{y^{m-1}}{(1+y)^{m+n}} \cdot \frac{1}{(1+y)^{n-1}} \cdot \frac{dy}{(1+y)^2} \\ &= \int_0^\infty \frac{y^{m-1}}{(1+y)^{m+n}} dy \end{aligned} \quad \dots(2)$$

(b) We have,  $\beta(m, n) = \beta(n, m)$

$$\therefore \beta(m, n) = \int_0^\infty \frac{y^{n-1}}{(1+y)^{m+n}} dy. \quad \dots(3)$$

Eq. (3) can be obtained directly from Eq. (1) by substituting  $x = \frac{1}{1+y}$ .

### 44.3 EVALUATION OF GAMMA FUNCTION

$$\text{By definition, } \Gamma n = \int_0^\infty e^{-x} x^{n-1} dx$$

Integrating by parts taking  $e^{-x}$  as second function, we get

$$\begin{aligned} \Gamma n &= \left[ -x^{n-1} e^{-x} \right]_0^\infty + \int_0^\infty (n-1) x^{n-2} e^{-x} dx \\ &= (n-1) \int_0^\infty x^{n-2} e^{-x} dx = (n-1) \Gamma(n-1) \\ &\left[ \because \lim_{x \rightarrow 0} x^{n-1} e^{-x} = \lim_{x \rightarrow 0} \frac{x^{n-1}}{e^x} = 0 \text{ and } \lim_{x \rightarrow \infty} x^{n-1} e^{-x} = \lim_{x \rightarrow \infty} \frac{x^{n-1}}{e^x} = 0 \right] \\ \therefore \Gamma n &= (n-1) \Gamma(n-1) \end{aligned} \quad \dots(2)$$

This is the fundamental recursion relation satisfied by the gamma function.

$$\text{Now, } \Gamma(n-1) = (n-2) \Gamma(n-2)$$

$$\therefore \Gamma n = (n-1)(n-2) \Gamma(n-2)$$

If  $n$  is positive integer, then proceeding as above repeatedly, we get

$$\Gamma n = (n-1)(n-2)(n-3)\dots3.2.1 \Gamma 1$$

$$\text{But } \Gamma 1 = \int_0^\infty e^{-x} dx = \left[ -e^{-x} \right]_0^\infty = 1 \quad \dots(3)$$

Hence, when  $n$  is a positive integer

$$\Gamma n = (n-1)(n-2)\dots3.2.1 = (n-1)! \quad \dots(4)$$

This is the *fundamental property* of Gamma functions. From this property we may write

$$\Gamma(n+1) = n\Gamma n \quad i.e., \quad \Gamma n = \frac{\Gamma(n+1)}{n} \quad \dots(5)$$

Putting  $n=0$  in Eq. (5), we get

$$\Gamma_0 = \infty \quad \dots(6)$$

It can be further shown that

$$\Gamma(-n) = \infty \quad \dots(7)$$

#### 44.4 THE VALUE OF $\Gamma\left(\frac{1}{2}\right)$ AND GRAPH OF THE GAMMA FUNCTION

We have by definition,

$$\Gamma n = \int_0^\infty e^{-x} x^{n-1} dx \quad \dots(1)$$

Putting  $x=y^2$ ;  $dx=2y dy$ , we get

$$\Gamma(n) = 2 \int_0^\infty y^{2n-1} e^{-y^2} dy \quad \dots(2)$$

If we now substitute,  $n=\frac{1}{2}$ , then

$$\Gamma\left(\frac{1}{2}\right) = 2 \int_0^\infty e^{-y^2} dy \quad \dots(3)$$

Now,  $\int_0^\infty e^{-y^2} dy = \frac{\sqrt{\pi}}{2} \quad \dots(4)$

$$\therefore \Gamma\left(\frac{1}{2}\right) = 2 \frac{\sqrt{\pi}}{2} = \sqrt{\pi} \quad \dots(5)$$

We can use the relation  $\Gamma n = \frac{\Gamma(n+1)}{n}$  to find  $\Gamma\left(-\frac{1}{2}\right)$ ,  $\Gamma\left(-\frac{3}{2}\right)$  etc.

$$\Gamma\left(-\frac{1}{2}\right) = \frac{\Gamma\left(\frac{1}{2}\right)}{-\frac{1}{2}} = -2\sqrt{\pi} \quad \dots(6)$$

$$\Gamma\left(-\frac{3}{2}\right) = \frac{\Gamma\left(-\frac{1}{2}\right)}{-\frac{3}{2}} = -\frac{2}{3}(-2\sqrt{\pi}) = \frac{4}{3}\sqrt{\pi} \text{ etc.} \quad \dots(7)$$

$$\Gamma(0) = \frac{\Gamma(1)}{0} = \frac{1}{0} = \infty$$

By repeated application of this formula it is seen that the gamma function becomes infinite when  $n$  is zero or a negative integer. Fig. 44.1 represents the graph of  $\Gamma n$ .

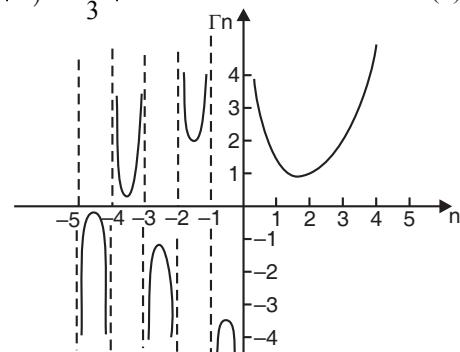


Fig. 44.1

### 44.5 OTHER FORMS OF GAMMA FUNCTION

By definition

$$\Gamma n = \int_0^\infty e^{-x} x^{n-1} dx \quad \dots(1)$$

(a) Putting  $x = \lambda y, dx = \lambda dy$  in Eq. (1), we get

$$\begin{aligned} \Gamma n &= \int_0^\infty e^{-\lambda y} \lambda^{n-1} y^{n-1} \cdot \lambda dy = \lambda^n \int_0^\infty e^{-\lambda y} y^{n-1} dy \\ \therefore \int_0^\infty e^{-\lambda y} y^{n-1} dy &= \frac{\Gamma n}{\lambda^n} \end{aligned} \quad \dots(2)$$

(b) Put

$$e^{-x} = y \text{ in Eq. (1)} \quad \therefore x = \log_e \frac{1}{y} \quad \text{and} \quad dx = -\frac{dy}{y}$$

$$\Gamma n = \int_0^\infty e^{-x} x^{n-1} dx = - \int_1^0 \left( \log \frac{1}{y} \right)^{n-1} dy = \int_0^1 \left( \log \frac{1}{y} \right)^{n-1} dy \quad \dots(3)$$

(c) Put  $x^n = y \quad \therefore x = y^{1/n}$  and  $dx = \frac{1}{n} y^{(1-n)/n} dy$  in Eq. (1)

$$\begin{aligned} \Gamma n &= \int_0^\infty e^{-x} x^{n-1} dx = \int_0^\infty e^{-y^{1/n}} y^{(n-1)/n} \cdot \frac{1}{n} y^{(1-n)/n} dy \\ &= \frac{1}{n} \int_0^\infty e^{-y^{1/n}} dy \end{aligned} \quad \dots(4)$$

### 44.6 RELATION BETWEEN BETA AND GAMMA FUNCTIONS

$$\beta(m, n) = \frac{\Gamma m \Gamma n}{\Gamma(m+n)}$$

By definition,

$$\Gamma n = \int_0^\infty e^{-x} x^{n-1} dx \quad \dots(1)$$

Put  $x = \lambda y, dx = \lambda dy$  in Eq. (1). Then we get

$$\Gamma n = \int_0^\infty e^{-\lambda y} \lambda^{n-1} y^{n-1} \cdot \lambda dy = \lambda^n \int_0^\infty e^{-\lambda y} y^{n-1} dy \quad \dots(2)$$

Now,

$$\Gamma m = \int_0^\infty \lambda^m e^{-\lambda x} x^{m-1} dx \quad \dots(3)$$

Multiplying both sides by  $e^{-\lambda} \lambda^{n-1}$  and integrating with respect to  $\lambda$  between the limits 0 and  $\infty$ , we get

$$\Gamma m \int_0^\infty e^{-\lambda} \lambda^{n-1} d\lambda = \int_0^\infty \left[ \int_0^\infty e^{-\lambda(1+x)} \lambda^{m+n-1} d\lambda \right] x^{m-1} dx \quad \dots(4)$$

But

$$\int_0^\infty e^{-\lambda} \lambda^{n-1} d\lambda = \Gamma n$$

and  $\int_0^\infty e^{-\lambda(1+x)} \cdot \lambda^{m+n-1} d\lambda = \frac{\Gamma(m+n)}{(1+x)^{m+n}}$  by (2) of section 44.5

Substituting these values in Eq. (4), we get

$$\begin{aligned} \Gamma m \Gamma n &= \int_0^\infty \frac{\Gamma(m+n)}{(1+x)^{m+n}} x^{m-1} dx \\ &= \Gamma(m+n) \int_0^\infty \frac{x^{m-1}}{(1+x)^{m+n}} dx \end{aligned} \quad \dots(5)$$

$$\begin{aligned} &= \Gamma(m+n) \beta(m, n) && \text{by Eq. (2) of section 9.4} \\ \therefore \quad \beta(m, n) &= \frac{\Gamma m \Gamma n}{\Gamma(m+n)} && \dots(6) \end{aligned}$$

**EXAMPLE 1.** Show that  $2^n \Gamma\left(n + \frac{1}{2}\right) = 1.3.5 \dots (2n-1) \sqrt{\pi}$ .

**SOL.** We know that  $\Gamma(n+1) = n\Gamma n$ .

$$\begin{aligned} \therefore \quad \Gamma\left(n + \frac{1}{2}\right) &= \left(n - \frac{1}{2}\right) \Gamma\left(n - \frac{1}{2}\right) \\ &= \left(n - \frac{1}{2}\right) \left(n - \frac{3}{2}\right) \Gamma\left(n - \frac{3}{2}\right) \\ &= \left(n - \frac{1}{2}\right) \left(n - \frac{3}{2}\right) \left(n - \frac{5}{2}\right) \left(n - \frac{7}{2}\right) \dots \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \\ &= \left(n - \frac{1}{2}\right) \left(n - \frac{3}{2}\right) \left(n - \frac{5}{2}\right) \dots \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \sqrt{\pi} \quad \left(\because \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\right) \\ &= \frac{\sqrt{\pi}}{2^n} (2n-1)(2n-3)\dots5.3.1 \end{aligned}$$

or  $2^n \Gamma\left(n + \frac{1}{2}\right) = 1.3.5 \dots (2n-1) \sqrt{\pi}$ .

**EXAMPLE 2.** Show that

$$\int_0^\infty \frac{x^8 (1-x^6)}{(1+x)^{24}} dx = 0.$$

**SOL.** Let

$$\begin{aligned} I &= \int_0^\infty \frac{x^8 (1-x^6)}{(1+x)^{24}} dx \\ &= \int_0^\infty \frac{x^8}{(1+x)^{24}} dx - \int_0^\infty \frac{x^{14}}{(1+x)^{24}} dx \\ &= \int_0^\infty \frac{x^{9-1}}{(1+x)^{9+15}} dx - \int_0^\infty \frac{x^{15-1}}{(1+x)^{15+9}} dx \\ &= \beta(9, 15) - \beta(15, 9) \\ &= 0. \quad [\because \beta(m, n) = \beta(n, m)] \end{aligned}$$

**EXAMPLE 3.** Show that  $\int_0^{\pi/2} \sin^p \theta \cos^q \theta d\theta = \frac{\Gamma\left(\frac{p+1}{2}\right) \Gamma\left(\frac{q+1}{2}\right)}{2\Gamma\left(\frac{p+q+2}{2}\right)}$

Hence evaluate  $\int_0^{\pi/2} \sin^p \theta d\theta$  and  $\int_0^{\pi/2} \cos^q \theta d\theta$

**SOL.** From definition of Beta function

$$\beta(m, n) = \int_0^1 x^{m-1} (1-x)^{n-1} dx$$

Put  $x = \sin^2 \theta$ .  $\therefore dx = 2 \sin \theta \cos \theta d\theta$  and  $1-x = \cos^2 \theta$ .

$$\begin{aligned}\beta(m, n) &= \int_0^{\pi/2} \sin^{2m-2} \theta \cos^{2n-2} \theta \cdot 2 \sin \theta \cos \theta d\theta \\ &= 2 \int_0^{\pi/2} \sin^{2m-1} \theta \cos^{2n-1} \theta d\theta\end{aligned}\dots(1)$$

But  $\beta(m, n) = \frac{\Gamma m \Gamma n}{\Gamma(m+n)}$  ... (2)

Comparing Eqs. (1) and (2), we get

$$\int_0^{\pi/2} \sin^{2m-1} \theta \cos^{2n-1} \theta d\theta = \frac{\Gamma m \Gamma n}{2\Gamma(m+n)} \dots(3)$$

Let  $2m - 1 = p$  and  $2n - 1 = q$ . Then Eq. (3) becomes

$$\int_0^{\pi/2} \sin^p \theta \cos^q \theta d\theta = \frac{\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right)}{2\Gamma\left(\frac{p+q+2}{2}\right)} \dots(4)$$

Putting  $q = 0$  and  $p = 0$  successively, we get

$$\int_0^{\pi/2} \sin^p \theta d\theta = \frac{\Gamma\left(\frac{p+1}{2}\right)\Gamma\frac{1}{2}}{2\Gamma\left(\frac{p+2}{2}\right)} = \frac{\Gamma\left(\frac{p+1}{2}\right)\sqrt{\pi}}{2\Gamma\left(\frac{p}{2}+1\right)} \dots(5)$$

and  $\int_0^{\pi/2} \cos^q \theta d\theta = \frac{\Gamma\frac{1}{2}\Gamma\left(\frac{q+1}{2}\right)}{2\Gamma\left(\frac{q+2}{2}\right)} = \frac{\Gamma\left(\frac{q+1}{2}\right)\sqrt{\pi}}{2\Gamma\left(\frac{q}{2}+1\right)}$  ... (6)

## DIFFERENTIAL EQUATIONS

### 44.7 BESSEL'S DIFFERENTIAL EQUATION

The differential equation

$$x^2 \frac{d^2y}{dx^2} + x \frac{dy}{dx} + (x^2 - n^2)y = 0 \dots(1a)$$

is called *Bessel's differential equation*. Here,  $n$  is a constant.

This equation can also be put in the form

$$\frac{d^2y}{dx^2} + \frac{1}{x} \frac{dy}{dx} + \left(1 - \frac{n^2}{x^2}\right)y = 0 \dots(1b)$$

There is singularity at  $x = 0$ , and this is non-essential or removable singularity. Hence the given equation may be solved by the method of series integration as allowed by Fusch theorem.

The solutions of this equation are called *Bessel's functions* of order  $n$ .

The *series solution* of Eq. (1) in ascending powers of  $x$  may be written as

$$y = x^k (a_0 + a_1 x + a_2 x^2 + \dots + a_r x^r + \dots) = \sum_{r=0}^{\infty} a_r x^{k+r} \dots(2)$$

$$\therefore \frac{dy}{dx} = \sum_{r=0}^{\infty} a_r (k+r) x^{k+r-1} \dots(3)$$

$$\frac{d^2y}{dx^2} = \sum_{r=0}^{\infty} a_r (k+r)(k+r-1) x^{k+r-2} \quad \dots(4)$$

Substituting these values in Eq. (1b), we get

$$\begin{aligned} & \sum_{r=0}^{\infty} a_r (k+r)(k+r-1) x^{k+r-2} + \frac{1}{x} \sum_{r=0}^{\infty} a_r (k+r) x^{k+r-1} + \left(1 - \frac{n^2}{x^2}\right) \sum_{r=0}^{\infty} a_r x^{k+r} = 0 \\ \text{or } & \sum_{r=0}^{\infty} a_r [(k+r)(k+r-1) + (k+r) - n^2] x^{k+r-2} + \sum_{r=0}^{\infty} a_r x^{k+r} = 0 \\ \text{or } & \sum_{r=0}^{\infty} a_r [(k+r)(k+r-1+1) - n^2] x^{k+r-2} + \sum_{r=0}^{\infty} a_r x^{k+r} = 0 \\ \text{or } & \sum_{r=0}^{\infty} a_r \{(k+r)^2 - n^2\} x^{k+r-2} + \sum_{r=0}^{\infty} a_r x^{k+r} = 0 \end{aligned} \quad \dots(5)$$

This equation is an identity. Therefore, the coefficients of various powers of  $x$  must be equal to zero.

Equating to zero the coefficient of lowest power of  $x$  i.e.,  $x^{k-2}$  by putting  $r=0$  in Eq. (5), we get

$$a_0 (k^2 - n^2) = 0$$

But  $a_0 \neq 0$  ( $\because$  the first term of the series is non-vanishing)

$$\therefore k^2 - n^2 = 0 \quad \text{or} \quad k = \pm n \quad \dots(6)$$

Now equating to zero the coefficient of  $x^{k-1}$  by putting  $r=1$  in Eq. (5),

$$a_1 [(k+1)^2 - n^2] = 0$$

But from Eq. (6),  $(k+1)^2 - n^2 \neq 0$ .

$$\therefore a_1 = 0.$$

Equating to zero coefficient of general term i.e.,  $x^{k+r}$  in Eq. (5),

$$a_{r+2} [(k+r+2)^2 - n^2] + a_r = 0$$

$$\text{or } a_{r+2} [(k+r+2+n)(k+r+2-n)] + a_r = 0$$

$$\therefore a_{r+2} = -\frac{a_r}{(k+r+n+2)(k+r-n+2)} \quad \dots(7)$$

Since  $a_1 = 0$ , Eq. (7) gives  $a_3 = a_5 = a_7 = \dots = 0$ .

Now, for  $k = \pm n$ , there are two cases:

**Case (i)** When  $k = n$ ; we have

$$a_{r+2} = -\frac{a_r}{(2n+r+2)(r+2)} \quad \dots(8)$$

By putting  $r = 0, 2, 4, \dots$  etc., we get

$$\begin{aligned} a_2 &= -\frac{a_0}{(2n+2)\cdot 2} = \frac{-a_0}{2\cdot 2(n+1)} = -\frac{1}{2^2 \cdot 1!(n+1)} a_0 \\ a_4 &= -\frac{a_2}{(2n+4)(4)} = -\frac{a_2}{2^2 2!(n+2)} \\ &= +\frac{1}{2^2 2!(n+2)} \cdot \frac{1}{2^2 (n+1)} a_0 \\ &= (-1)^2 \frac{a_0}{2^4 2!(n+1)(n+2)} \end{aligned}$$

Similarly,  $a_6 = \frac{a_0}{2^6 3!(n+1)(n+2)(n+3)}$

Substituting  $k = n$  and the values of  $a_1, a_2, a_3, \dots$  etc., in Eq. (2), the series solution is

$$\begin{aligned} y &= a_0 x^n \left[ 1 + (-1) \frac{x^2}{2^2 1!(n+1)} + (-1)^2 \frac{x^4}{2^4 2!(n+1)(n+2)} \right. \\ &\quad \left. + \dots + (-1)^r \frac{x^{2r}}{2^{2r} r!(n+1)(n+2)\dots(n+r)} \right] \end{aligned} \quad \dots(9)$$

Let us put  $a_0 = \frac{1}{2^n \Gamma(n+1)}$ . Then the solution of Bessel's equation represented by Eq. (9) is

called Bessel's function of first kind. It is denoted by  $J_n(x)$ .

$$\begin{aligned} J_n(x) &= \frac{x^n}{2^n \Gamma(n+1)} \left[ 1 + (-1) \frac{x^2}{2^2 1!(n+1)} + \right. \\ &\quad \left. (-1)^2 \frac{x^4}{2^4 2!(n+1)(n+2)} + \dots + (-1)^r \frac{x^{2r}}{2^{2r} r!(n+1)(n+2)\dots(n+r)} + \dots \right] \\ &= \frac{x^n}{2^n \Gamma(n+1)} \times \sum_{r=0}^{\infty} (-1)^r \frac{x^{2r}}{2^{2r} r!(n+1)(n+2)\dots(n+r)} \\ &= \sum_{r=0}^{\infty} (-1)^r \cdot \frac{1}{r! \Gamma(n+r+1)} \left(\frac{x}{2}\right)^{n+2r} \end{aligned} \quad \dots(10)$$

**Case (ii).** When  $k = -n$ ; then the series solution is obtained by replacing  $n$  by  $-n$  in Eq. (10).

$$J_{-n}(x) = \sum_{r=0}^{\infty} (-1)^r \frac{1}{r! \Gamma(-n+r+1)} \left(\frac{x}{2}\right)^{-n+2r} \quad \dots(11)$$

#### General Solution.

When  $n$  is not an integer,  $J_{-n}(x)$  is distinct from  $J_n(x)$ . Hence the **most general solution of Bessel's equation is**

$$y = A J_n(x) + B J_{-n}(x) \quad \dots(12)$$

Here  $A$  and  $B$  are two arbitrary constants.

**Corollary.** Bessel's equation for  $n = 0$  is

$$\frac{d^2y}{dx^2} + \frac{1}{x} \frac{dy}{dx} + y = 0.$$

Its series solution by the same substitution  $y = \sum_{r=0}^{\infty} a_r x^{k+r}$  (as above) is obtained to be

$$y = a_0 \left( 1 - \frac{x^2}{2^2} + \frac{x^4}{2^2 \cdot 4^2} - \frac{x^6}{2^2 \cdot 4^2 \cdot 6^2} + \dots \right)$$

If  $a_0 = 1$ , this solution is denoted by  $J_0(x)$ , i.e.,

$$J_0(x) = 1 - \frac{x^2}{2^2} + \frac{x^4}{2^2 \cdot 4^2} - \frac{x^6}{2^2 \cdot 4^2 \cdot 6^2} + \dots$$

where  $J_0(x)$  is called *Bessel function of zeroeth order*.

**Show that  $J_{-n}(x) = (-1)^n J_n(x)$ , where  $n$  is a positive integer.**

We know that

$$J_{-n}(x) = \sum_{r=0}^{\infty} \frac{(-1)^r}{r! \Gamma(r-n+1)} \left(\frac{x}{2}\right)^{-n+2r}$$

In the above series for values of  $r$  from 0 to  $n-1$ , the terms obtained will have  $\Gamma$  **Function of negative numbers in the denominator**. As the  $\Gamma$  function of negative numbers is always infinite, all these terms will vanish. Hence we will have terms left for  $r=n$  and onwards.

$$\therefore J_{-n}(x) = \sum_{r=n}^{\infty} \frac{(-1)^r}{r! \Gamma(r-n+1)} \left(\frac{x}{2}\right)^{-n+2r}$$

Now putting  $r=n+s$ , we get

$$J_{-n}(x) = \sum_{s=0}^{\infty} \frac{(-1)^{n+s}}{(n+s)! \Gamma(s+1)} \left(\frac{x}{2}\right)^{n+2s}$$

Again  $(n+s)! = \Gamma(n+s+1)$  and  $\Gamma(s+1) = s!$

$$\begin{aligned} \therefore J_{-n}(x) &= (-1)^n \sum_{s=0}^{\infty} \frac{(-1)^s}{s! \Gamma(n+s+1)} \left(\frac{x}{2}\right)^{n+2s} \\ \Rightarrow J_{-n}(x) &= (-1)^n J_n(x) \end{aligned}$$

## 44.8 LEGENDRE'S DIFFERENTIAL EQUATION

The differential equation

$$(1-x^2) \frac{d^2y}{dx^2} - 2x \frac{dy}{dx} + n(n+1)y = 0 \quad \dots(1)$$

is called *Legendre's equation*.

The Legendre equation can be solved in series of ascending or descending powers of  $x$ . But the solution in descending powers of  $x$  is of more physical importance.

Let us consider the solution of Legendre's equation in a series of descending powers of  $x$  as

$$\begin{aligned} y &= x^k [a_0 + a_1 x^{-1} + a_2 x^{-2} + a_3 x^{-3} + \dots + a_r x^{-r} + \dots] \\ &= \sum_{r=0}^{\infty} a_r x^{k-r} \end{aligned} \quad \dots(2)$$

$$\therefore \frac{dy}{dx} = \sum_{r=0}^{\infty} a_r (k-r) x^{k-r-1} \quad \dots(3)$$

$$\text{and } \frac{d^2y}{dx^2} = \sum_{r=0}^{\infty} a_r (k-r)(k-r-1) x^{k-r-2} \quad \dots(4)$$

Now, putting the values of  $y$ ,  $\frac{dy}{dx}$  and  $\frac{d^2y}{dx^2}$  in Eq. (1), we get

$$\begin{aligned} (1-x^2) \sum_{r=0}^{\infty} a_r (k-r)(k-r-1) x^{k-r-2} \\ - 2x \sum_{r=0}^{\infty} a_r (k-r) x^{k-r-1} + n(n+1) \sum_{r=0}^{\infty} a_r x^{k-r} = 0 \end{aligned}$$

$$\begin{aligned}
 \text{or} \quad & \sum_{r=0}^{\infty} [(k-r)(k-r-1)x^{k-r-2} \\
 & - \{(k-r)(k-r-1) + 2(k-r) - n(n+1)\} x^{k-r}] a_r = 0 \\
 \text{or} \quad & \sum_{r=0}^{\infty} [(k-r)(k-r-1)x^{k-r-2} \\
 & + \{n(n+1) - (k-r)(k-r+1)\} x^{k-r}] a_r = 0 \quad \dots(5)
 \end{aligned}$$

Eq. (5) is an identity. Therefore, the coefficients of various powers of  $x$  can be equated to zero. First we equate to zero the coefficient of  $x^k$  (the highest power of  $x$ ) by putting  $r=0$  in (5), i.e.,

$$\{n(n+1) - k(k+1)\} a_0 = 0$$

Here  $a_0$  being the coefficient of the first term of the series, cannot be zero.

$$\text{Hence } n(n+1) - k(k+1) = 0$$

$$\text{or } n^2 + n - k^2 - k = 0$$

$$\text{or } (n^2 - k^2) + (n - k) = 0$$

$$\text{or } (n - k)(n + k + 1) = 0$$

This gives either  $k = n$  or  $k = -(n+1)$

... (6)

Now equating to zero the coefficient of  $x^{k-1}$  by putting  $r=1$  in (5), we get

$$\{n(n+1) - (k-1)k\} a_1 = 0$$

$$\text{or } \{n^2 + n - k^2 + k\} a_1 = 0$$

$$\text{or } \{(n^2 - k^2) + (n + k)\} a_1 = 0$$

$$\text{or } (n + k)(n - k + 1) a_1 = 0$$

$$\text{As } (n + k)(n - k + 1) \neq 0, \quad a_1 = 0 \quad \dots(7)$$

To find a relation between successive coefficients, we equate to zero the coefficients of  $x^{k-r-2}$  in Eq. (5)

$$(k-r)(k-r-1)a_r + \{n(n+1) - (k-r-2)(k-r-1)\} a_{r+2} = 0$$

This equation gives

$$\begin{aligned}
 a_{r+2} &= \frac{(k-r)(k-r-1)}{n(n+1) - (k-r-2)(k-r-1)} a_r \\
 &= \frac{(k-r)(k-r-1)}{\{(k-r)+(n-1)\} \{(k-r)-(n+2)\}} a_r \quad \dots(8)
 \end{aligned}$$

We know that  $a_1 = 0$ . Hence from Eq. (8) we get

$$a_1 = a_3 = a_5 = a_7 = \dots = 0 \quad \dots(9)$$

**Case (i)** When  $k = n$ , we get from Eq. (8)

$$a_{r+2} = \frac{(n-r)(n-r-1)}{(2n-r-1)(-r-2)} a_r = -\frac{(n-r)(n-r-1)}{(2n-r-1)(r+2)} a_r$$

Putting  $r = 0, 2, 4, \dots$  we get

$$a_2 = -\frac{n(n-1)}{(2n-1) \cdot 2} a_0$$

$$a_4 = -\frac{(n-2)(n-3)}{(2n-3) \cdot 4} a_2 = \frac{n(n-1)(n-2)(n-3)}{(2n-1)(2n-3) \cdot 2 \cdot 4} a_0$$

[by putting value of  $a_2$ ]

Similarly  $a_6 = \frac{n(n-1)(n-2)(n-3)(n-4)(n-5)}{(2n-1)(2n-3)(2n-5)2 \cdot 4 \cdot 6} a_0$  and so on.

Also we have  $a_1 = a_3 = a_5 = \dots = 0$  (from 9)

Substituting values of  $a_0, a_1, a_2, a_3$ , etc., in Eq. (2), we get the series solution for  $k = n$  as

$$y = a_0 \left[ x^n - \frac{n(n-1)}{2 \cdot (2n-1)} x^{n-2} + \frac{n(n-1)(n-2)(n-3)}{2 \cdot 4 \cdot (2n-1)(2n-3)} x^{n-4} - \dots \right] \dots(10)$$

where  $a_0$  is arbitrary constant and  $n$  is positive integer. If

$$a_0 = \frac{1 \cdot 3 \cdot 5 \dots (2n-1)}{n!},$$

then above solution is called the *Legendre polynomial* or *Legendre function of first kind* and is represented by  $P_n(x)$ .

Thus,

$$P_n(x) = \frac{1 \cdot 3 \cdot 5 \dots (2n-1)}{n!} \left[ x^n - \frac{n(n-1)}{2 \cdot (2n-1)} x^{n-2} + \frac{n(n-1)(n-2)(n-3)}{2 \cdot 4 \cdot (2n-1)(2n-3)} x^{n-4} - \dots \right] \dots(11)$$

$P_n(x)$  is a terminating series and for different values of  $n$  we get *Legendre polynomials*.

**Case (ii).** When  $k = -(n+1)$ .

From Eq. (8) we get,

$$a_{r+2} = \frac{(n+r+1)(n+r+2)}{(r+2)(2n+r+3)} a_r \dots(12)$$

Putting

$r = 0, 2, 4, \dots$ , we get

$$a_2 = \frac{(n+1)(n+2)}{2 \cdot (2n+3)} a_0$$

$$a_4 = \frac{(n+3)(n+4)}{4 \cdot (2n+5)} a_2 = \frac{(n+1)(n+2)(n+3)(n+4)}{2 \cdot 4 \cdot (2n+3)(2n+5)} a_0$$

(by putting the value of  $a_2$ )

$$a_6 = \frac{(n+5)(n+6)}{6 \cdot (2n+7)} a_4$$

$$= \frac{(n+1)(n+2)(n+3)(n+4)(n+5)(n+6)}{2 \cdot 4 \cdot 6 \cdot (2n+3)(2n+5)(2n+7)} a_0 \quad \text{and so on.}$$

From Eq. (9) we have  $a_1 = a_3 = a_5 = \dots = 0$

Putting the values of  $a_0, a_1, a_2, a_3, \dots$  etc., in Eq. (2), we get the series solution for  $k = -(n+1)$  as

$$y = a_0 \left[ x^{-n-1} + \frac{(n+1)(n+2)}{2 \cdot (2n+3)} x^{-n-3} + \frac{(n+1)(n+2)(n+3)(n+4)}{2 \cdot 4 \cdot (2n+3)(2n+5)} x^{-n-5} + \dots \right]$$

If  $a_0 = \frac{n!}{1.3.5 \dots (2n+1)}$ , the above solution is called Legendre polynomial or Legendre function of second kind and is denoted by  $Q_n(x)$ . Thus

$$Q_n(x) = \frac{n!}{1.3.5 \dots (2n+1)} \left[ x^{-n-1} + \frac{(n+1)(n+2)}{2.(2n+3)} x^{-n-3} + \frac{(n+1)(n+2)(n+3)(n+4)}{2.4.(2n+3)(2n+5)} x^{-n-5} + \dots \right] \dots(13)$$

This is an infinite or non-terminating series, since  $n$  is a positive integer.

$P_n(x)$  and  $Q_n(x)$  are two independent solutions of Legendre's equation. Therefore, the most general solution of *Legendre's equation* is

$$y = A P_n(x) + B Q_n(x) \dots(14)$$

Here  $A$  and  $B$  are arbitrary constants.

## 44.9 HERMITE'S DIFFERENTIAL EQUATION

The differential equation

$$\frac{d^2y}{dx^2} - 2x \frac{dy}{dx} + 2ny = 0 \dots(1)$$

where  $n$  is a constant, is called *Hermite differential equation*.

Suppose its series solution is

$$y = \sum_{r=0}^{\infty} a_r x^{k+r} \dots(2)$$

$$\frac{dy}{dx} = \sum_{r=0}^{\infty} a_r (k+r) x^{k+r-1} \quad \text{and} \quad \frac{d^2y}{dx^2} = \sum_{r=0}^{\infty} a_r (k+r)(k+r-1) x^{k+r-2}$$

Now, substituting the values of  $y$ ,  $\frac{dy}{dx}$  and  $\frac{d^2y}{dx^2}$  in Eq. (1), we get

$$\sum_{r=0}^{\infty} a_r [(k+r)(k+r-1) x^{k+r-2} - 2(k+r) x^{k+r} + 2n x^{k+r}] = 0$$

$$\Rightarrow \sum_{r=0}^{\infty} a_r [(k+r)(k+r-1) x^{k+r-2} - 2(k+r-n) x^{k+r}] = 0 \dots(3)$$

Equating the coefficient of the first term (*i.e.*,  $x^{k-2}$  by putting  $r=0$ ), to zero, we get

$$a_0 k (k-1) = 0 \dots(4)$$

But  $a_0 \neq 0$ , as it is the coefficient of the first term.

$$\therefore \text{either } k = 0 \quad \text{or} \quad k = 1 \dots(5)$$

Now equating to zero the coefficient of  $x^{k-1}$  in (3), we get

$$a_1 (k+1) k = 0 \dots(6)$$

$k+1 \neq 0$  for any value of  $k$  given by (5).

$$\Rightarrow a_1 = 0 \quad \text{or} \quad k = 0 \quad \text{or both are zero.} \dots(7)$$

Now equating to zero the coefficient of general term  $x^{k+r}$ , we get

$$a_{r+2} (k+r+2) (k+r+1) - 2a_r (k+r-n) = 0$$

$$\begin{aligned} \text{or } a_{r+2} &= \frac{2(k+r-n)}{(k+r+2)(k+r+1)} a_r \\ \text{or } a_{r+2} &= \frac{2(k+r)-2n}{(k+r+2)(k+r+1)} a_r \end{aligned} \quad \dots(8)$$

Now we shall discuss two cases:

**Case I.** When  $k = 0$ , we have from Eq. (8),

$$a_{r+2} = \frac{2r-2n}{(r+2)(r+1)} a_r \quad \dots(9)$$

Substituting  $r = 0, 2, 4 \dots$  etc., we get

$$\begin{aligned} a_2 &= \frac{-2n}{2.1} a_0 = \frac{-2n}{2!} a_0 \\ a_4 &= \frac{4-2n}{4.3} a_2 = \frac{(4-2n)}{4.3} \cdot \left( \frac{-2n}{2!} a_0 \right) = \frac{2^2 n(n-2)}{4!} a_0 \end{aligned}$$

and so on. In general

$$a_{2m} = \frac{(-2)^m n(n-2) \dots (n-2m+2)}{(2m)!} a_0 \quad \dots(10)$$

Again, substituting  $r = 1, 3, 5 \dots$  etc., in (9), we get

$$\begin{aligned} a_3 &= \frac{2-2n}{3.2} a_1 = -\frac{2(n-1)}{3!} a_1 \\ a_5 &= \frac{6-2n}{5.4} a_3 = \frac{6-2n}{5.4} \cdot \left( -\frac{2(n-1)}{3!} \right) a_1 \\ &= (-2)^2 \frac{(n-1)(n-3)}{5!} a_1 \quad \text{etc.} \end{aligned}$$

$$\text{In general } a_{2m+1} = (-2)^m \frac{(n-1)(n-3) \dots (n-2m+1)}{(2m+1)!} a_1$$

Now, if  $a_1 \neq 0$ , then we have

$$\begin{aligned} y &= \sum_{r=0}^{\infty} a_r x^r = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots \\ &= a_0 \left[ 1 - \frac{2n}{2!} x^2 + \frac{2^2 n(n-2)}{4!} x^4 + \dots \right. \\ &\quad \left. + (-2)^m \frac{n(n-2) \dots (n-2m+2)}{(2m)!} x^{2m} + \dots \right] \\ &\quad + a_1 \left[ x - \frac{2(n-1)}{3!} x^3 + \frac{2^2 (n-1)(n-3)}{5!} x^5 + \dots \right. \\ &\quad \left. + (-2)^m \frac{(n-1)(n-3) \dots (n-2m+1)}{(2m+1)!} x^{2m+1} + \dots \right] \dots(11) \end{aligned}$$

If  $a_1 = 0$ , Eq. (11) reduces to

$$\begin{aligned} y &= a_0 \left[ 1 - \frac{2n}{2!} x^2 + 2^2 \frac{n(n-2)}{4!} x^4 + \dots \right. \\ &\quad \left. + (-2)^m \frac{n(n-2)\dots(n-2m+2)}{(2m)!} x^{2m} + \dots \right] = y_1 \text{ (say)} \dots(12) \end{aligned}$$

**Case II.** When  $k = 1$ , we have from Eq. (8),

$$a_{r+2} = \frac{2(r+1)-2n}{(r+3)(r+2)} a_r \dots(13)$$

Putting  $r = 1, 3, \dots$  etc., in the above relation, we obtain

$$a_3 = a_5 = a_7 = \dots = 0 \quad [\because a_1 = 0 \text{ from Eq. (6)}]$$

Substituting  $r = 0, 2, 4, \dots$  etc., in Eq. (13), we get

$$\begin{aligned} a_2 &= \frac{(2-2n)}{3.2} a_0 = \frac{-2(n-1)}{3!} a_0 \\ a_4 &= \frac{6-2n}{5.4} a_2 = \frac{(6-2n)}{5.4} \left\{ -\frac{2(n-1)}{3!} a_0 \right\} = \frac{2^2(n-1)(n-3)}{5!} a_0 \end{aligned}$$

In general,

$$a_{2m} = (-2)^m \frac{(n-1)(n-3)\dots(n-2m+1)}{(2m+1)!} a_0$$

Hence,

$$\begin{aligned} y &= \sum_r a_r x^{r+1} = a_0 x + a_2 x^3 + a_4 x^5 \dots + \dots \\ &\quad (\because a_1 = a_3 = a_5 = \dots = 0) \\ &= a_0 \left[ x - \frac{2(n-1)}{3!} x^3 + \frac{2^2(n-1)(n-3)}{5!} x^5 - \dots \right. \\ &\quad \left. + (-2)^m \frac{(n-1)(n-3)\dots(n-2m+1)}{(2m+1)!} x^{2m+1} \dots \right] = y_2 \text{ (say)} \dots(14) \end{aligned}$$

Inspection of equations (11) and (14) shows that (14) is a part of solution as given by (11). The two are the solutions of the same equation; so we can say that (14) is not a part of (11); but it is a separate solution. Hence  $a_1 = 0$  and so the solution in case  $k = 0$  must be given by (12). In view of this, the general solution of Hermite equation is given by

$$y = A y_1 + B y_2 \dots(15)$$

where  $A$  and  $B$  are arbitrary constants and  $y_1, y_2$  are given by equations (12) and (14).

#### 44.10 LAGUERRE'S DIFFERENTIAL EQUATION

Laguerre differential equation is

$$x \frac{d^2y}{dx^2} + (1-x) \frac{dy}{dx} + ny = 0 \dots(1)$$

Let its series solution be

$$y = \sum_{r=0}^{\infty} a_r x^{k+r} \dots(2)$$

$$\therefore \frac{dy}{dx} = \sum_{r=0}^{\infty} a_r (k+r) x^{k+r-1}$$

$$\text{and } \frac{d^2y}{dx^2} = \sum_{r=0}^{\infty} a_r (k+r)(k+r-1) x^{k+r-2}$$

Substituting these values in (1), we get the identity

$$\begin{aligned} x \sum_{r=0}^{\infty} a_r (k+r)(k+r-1) x^{k+r-2} + (1-x) \sum_{r=0}^{\infty} a_r (k+r) x^{k+r-1} \\ + n \sum_{r=0}^{\infty} a_r x^{k+r} = 0 \\ \text{or} \\ \sum_{r=0}^{\infty} a_r [(k+r)^2 x^{k+r-1} - (k+r-n) x^{k+r}] = 0 \end{aligned} \quad \dots(3)$$

Equating to zero the coefficient of lowest power of  $x$

i.e.,  $x^{k-1}$  (putting  $r=0$ ), we get

$$a_0 k^2 = 0$$

As  $a_0 \neq 0$ , being the coefficient of first term of the series,

$$k = 0 \quad \dots(4)$$

Now equating to zero the coefficient of general term  $x^{k+r}$ , we get

$$a_{r+1} (k+r+1)^2 - a_r (k+r-n) = 0$$

$$\text{or} \quad a_{r+1} = \frac{k+r-n}{(k+r+1)^2} a_r$$

$$\text{As } k=0, \text{ we have } a_{r+1} = \frac{r-n}{(r+1)^2} a_r \quad \dots(5)$$

Substituting  $r=0, 1, 2, 3, \dots$  etc., in Eq. (5), we get

$$\begin{aligned} a_1 &= -\frac{n}{1} a_0 = (-1) n a_0 \\ a_2 &= \frac{1-n}{2^2} a_1 = -\frac{(n-1)}{2^2} \times (-1) n a_0 = (-1)^2 \frac{n(n-1)}{(2!)^2} a_0 \\ a_3 &= \frac{2-n}{3^2} a_2 = (-1)^3 \frac{n(n-1)(n-2)}{(3!)^2} a_0 \\ \dots &\dots \dots \dots \dots \\ a_r &= (-1)^r \frac{n(n-1)(n-2)\dots(n-r+1)}{(r!)^2} a_0 \end{aligned}$$

Therefore from Eq. (2), we have (for  $k=0$ )

$$\begin{aligned} y &= \sum_{r=0}^{\infty} a_r x^{k+r} = \sum_r a_r x^r \\ &= a_0 + a_1 x + a_2 x^2 + \dots + a_r x^r + \dots \\ &= a_0 \left[ 1 - nx + \frac{n(n-1)}{(2!)^2} x^2 + \dots + (-1)^r \frac{n(n-1)\dots(n-r+1)}{(r!)^2} x^r + \dots \right] \\ &= a_0 \sum_{r=0}^{\infty} (-1)^r \frac{n(n-1)\dots(n-r+1)}{(r!)^2} x^r \\ &= a_0 \sum_{r=0}^{\infty} \frac{(-1)^r n!}{(r!)^2 (n-r)!} x^r \end{aligned} \quad \dots(6)$$

In case  $n$  is a positive integer and  $a_0 = 1$  and the series terminates after  $n^{\text{th}}$  term, the solution (6) is said to be **Laguerre polynomial of degree  $n$**  and is denoted by  $L_n(x)$ .

$$L_n(x) = \sum_{r=0}^n \frac{(-1)^r \times (n!)^2}{(r!)^2 (n-r)!} x^r \quad \dots(7)$$

**Note.** Some authors define  $L_n(x)$  by taking  $a_0 = n!$ . In that case, we have

$$L_n(x) = \sum_{r=0}^n (-1)^r \frac{(n!)^2}{(n-r)!(r!)^2} x^r.$$

Then the solution of Laguerre equation for  $n$  to be a positive integer is

$$y = A L_n(x) \quad \dots(8)$$

## VECTOR ANALYSIS

### 44.11 GAUSS' DIVERGENCE THEOREM

**Statement.** The surface integral of the normal component of a vector  $\mathbf{A}$  taken over a closed surface  $S$  is equal to the volume integral of the divergence of  $\mathbf{A}$  taken over the volume  $V$  enclosed by the surface  $S$ , i.e.

$$\int_S \int \mathbf{A} \cdot d\mathbf{S} = \iint_V \int \operatorname{div} \mathbf{A} dV.$$

Thus the theorem gives us a method of reducing triple integrals to double integrals.

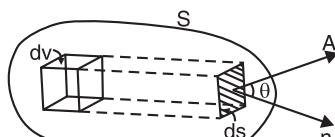


Fig. 44.2 (a)

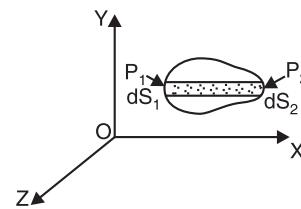


Fig. 44.2 (b)

**Proof.** Consider a ‘closed’ surface  $S$  of any arbitrary shape drawn in a vector field  $\mathbf{A}$  and enclosing a volume  $V$  (Fig. 44.2 (a)). Let  $dV = dx dy dz$  be a small cubical volume element around the point where the value of the vector field is  $\mathbf{A}$ .

$$\begin{aligned} \text{R.H.S.} &= \iint_V \int \operatorname{div} \mathbf{A} dV = \iint_V \int \left( \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \right) dx dy dz \\ &= \iint_V \int \frac{\partial A_x}{\partial x} dx dy dz + \iint_V \int \frac{\partial A_y}{\partial y} dx dy dz + \iint_V \int \frac{\partial A_z}{\partial z} dx dy dz \end{aligned}$$

Let us now consider the first integral on the right. We integrate it with respect to  $x$ , along a strip of cross-section  $dy dz$  extending from  $P_1(x_1, y, z)$  to  $P_2(x_2, y, z)$  [Fig. 44.2 (b)]

$$\iint_V \int \frac{\partial A_x}{\partial x} dx dy dz = \iint [A_x(x_2, y, z) - A_x(x_1, y, z)] dy dz$$

For the face  $P_2$ , the outward drawn normal is along positive  $X$ -axis. But for face  $P_1$  the outward drawn normal is along negative  $X$ -axis. Thus

$$\text{at } P_1, \quad dy dz = -dS_x$$

$$\text{at } P_2, \quad dy dz = +dS_x$$

$$\therefore \iint_V \int \frac{\partial A_x}{\partial x} dx dy dz = \int_{S_1} \int A_x dS_x + \int_{S_2} \int A_x dS_x = \int_S \int A_x dS_x \quad \dots(1)$$

Here the surface integration ( $S$ ) is evaluated over the entire surface. Similarly, we can show that

$$\iint_V \int \frac{\partial A_y}{\partial y} dx dy dz = \iint_S A_y dS_y \quad \dots(2)$$

$$\iint_V \int \frac{\partial A_z}{\partial z} dx dy dz = \iint_S A_z dS_z \quad \dots(3)$$

Adding Eqs. (1), (2) and (3), we obtain

$$\begin{aligned} \iint_V \int \left( \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \right) dx dy dz &= \iint_S (A_x dS_x + A_y dS_y + A_z dS_z) \\ \therefore \iint_V \int \operatorname{div} \mathbf{A} dV &= \iint_S \mathbf{A} \cdot d\mathbf{S} \end{aligned} \quad \dots(4)$$

Therefore, Gauss divergence theorem is proved.

Eq. (4) may also be expressed as

$$\iint_V \int \operatorname{div} \mathbf{A} dV = \iint_S \mathbf{A} \cdot \hat{\mathbf{n}} d\mathbf{S} \quad \dots(5)$$

where  $\hat{\mathbf{n}}$  is a unit vector along the outward drawn normal.

**EXAMPLE 1.** Evaluate  $\iint_S \mathbf{r} \cdot \hat{\mathbf{n}} d\mathbf{S}$ , where  $S$  is a closed surface.

**SOL.** By the divergence theorem, we have

$$\begin{aligned} \iint_S \int \mathbf{r} \cdot \hat{\mathbf{n}} d\mathbf{S} &= \iint_V \int \operatorname{div} \mathbf{r} dV \\ &= \iint_V \int \left( \frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z} \right) dV \quad [\because r = x\hat{i} + y\hat{j} + z\hat{k}] \\ &= 3 \iint_V \int dV \\ &= 3 V. \end{aligned}$$

Here  $V$  is the volume enclosed by  $S$ .

## 44.12 STOKE'S THEOREM

Stoke's theorem enables us to transform the surface integral of the curl of the vector field  $\mathbf{A}$  into the line integral of that vector field  $\mathbf{A}$  over the boundary  $C$  of that surface and vice-versa.

**Statement.** The line integral of a vector  $\mathbf{A}$  taken around a closed curve  $C$ , which bounds a surface  $S$ , is equal to the surface integral of the curl of  $\mathbf{A}$  taken over  $S$ , i.e.,

$$\int_C \mathbf{A} \cdot d\mathbf{r} = \iint_S \operatorname{curl} \mathbf{A} \cdot d\mathbf{S}$$

**Proof.** Consider an open surface  $S$  in a vector field  $\mathbf{A}$  (Fig. 44.3). The bounding edge of the surface is a closed curve. The line integral of  $\mathbf{A}$  round the closed curve when it is traced in the anticlockwise sense as shown in Fig. 44.3 is

$$\oint_C \mathbf{A} \cdot d\mathbf{r} \quad \dots(1)$$

Divide the surface  $S$  into a large number of small areas. Consider one such element of area  $dS$ . Let  $\hat{\mathbf{n}}$  be a unit positive (outward) normal upon  $dS$ . The vector area of the element is

$$\hat{\mathbf{n}} dS = d\mathbf{S}.$$

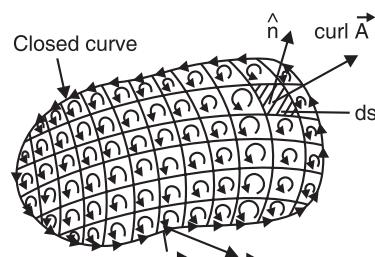


Fig. 44.3

By definition, the curl of a (non-lamellar) vector field at any point is defined as the maximum line integral of the vector computed per unit area along the boundary of an infinitesimal area at that point.

Therefore the line integral of  $\mathbf{A}$  around the boundary of the area  $dS$  is  
 $\text{curl } \mathbf{A} \cdot d\mathbf{S}$ .

This applies to all surface elements. Hence the sum of the line integrals of  $\mathbf{A}$  around the boundaries of all the elements is

$$\int_S \int \text{curl } \mathbf{A} \cdot d\mathbf{S}, \quad \dots(2)$$

which is the surface integral of  $\text{curl } \vec{\mathbf{A}}$ .

Now, it is clear from the figure that the line integrals along the common sides of the adjacent elements mutually cancel because they are traversed in opposite directions. Thus only those sides of the element which lie in the periphery of the surface contribute to the line integral. Therefore, the surface integral (2) gives the line integral of  $\mathbf{A}$  round the periphery of the surface, which is also given by (1).

$$\therefore \oint_C \mathbf{A} \cdot d\mathbf{r} = \iint_S \text{curl } \mathbf{A} \cdot d\mathbf{S}$$

This proves Stokes' theorem.

**EXAMPLE 1.** Verify Stoke's theorem for the vector

$$\mathbf{A} = (2x - y)\hat{\mathbf{i}} - yz^2\hat{\mathbf{j}} - y^2z\hat{\mathbf{k}},$$

over the upper half surface of the sphere  $x^2 + y^2 + z^2 = 1$

**SOL.** The boundary  $C$  of surface  $S$  is a circle in the  $xy$  plane ( $z = 0$ ) of radius 1 and centre at the origin (Fig. 44.4).  $x^2 + y^2 = 1$ .

Let  $x = \cos \theta, y = \sin \theta, z = 0,$

$0 \leq \theta \leq 2\pi$  be parametric equation of  $C$ .

$$\therefore \int_C \mathbf{A} \cdot d\mathbf{r} =$$

$$\begin{aligned} & \int_C [(2x - y)\hat{\mathbf{i}} - yz^2\hat{\mathbf{j}} - y^2z\hat{\mathbf{k}}]_{z=0} \cdot (\hat{\mathbf{i}} dx + \hat{\mathbf{j}} dy + \hat{\mathbf{k}} dz) \\ &= \int_C (2x - y) dx \\ &= \int_0^{2\pi} (2 \cos \theta - \sin \theta) (-\sin \theta) d\theta \\ &= \pi \end{aligned} \quad \dots(1)$$

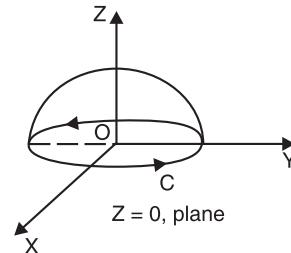


Fig. 44.4

Also

$$\nabla \times \mathbf{A} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ 2x - y & -yz^2 & -y^2z \end{vmatrix}$$

$$\begin{aligned} \int_S \int (\nabla \times \mathbf{A}) \cdot d\mathbf{S} &= \int_S \int \hat{\mathbf{k}} \hat{\mathbf{k}} dx dy = \int_S \int dx dy \\ &= \int_{x=-1}^{+1} \int_{y=-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dx dy \\ &= 2 \cdot \int_{-1}^{+1} \sqrt{(1-x^2)} dx \\ &= 4 \cdot \int_0^1 \sqrt{(1-x^2)} dx = 4 \int_{\pi/2}^0 (\sin \theta) (-\sin \theta) d\theta \\ &= \pi \end{aligned} \quad \dots(2)$$

Thus Stokes' theorem is verified.

### 44.13 GREEN'S THEOREM

**Statement.** If  $\psi_1, \psi_2$  be the scalar point functions having continuous derivatives at least of second order; then

$$\iint_V (\psi_1 \nabla^2 \psi_2 - \psi_2 \nabla^2 \psi_1) dx dy dz = \int_S (\psi_1 \operatorname{grad} \psi_2 - \psi_2 \operatorname{grad} \psi_1) \cdot d\mathbf{S}$$

**Proof.** Gauss divergence theorem is

$$\iint_V \operatorname{div} \mathbf{A} dV = \iint_S \mathbf{A} \cdot d\mathbf{S} \quad \dots(1)$$

In (1), substitute

$$\mathbf{A} = \psi_1 \nabla \psi_2 \quad \dots(2)$$

$$\begin{aligned} \operatorname{div} \mathbf{A} = \nabla \cdot \mathbf{A} &= \left( \mathbf{i} \frac{\partial}{\partial x} + \mathbf{j} \frac{\partial}{\partial y} + \mathbf{k} \frac{\partial}{\partial z} \right) \cdot \left( \psi_1 \frac{\partial \psi_2}{\partial x} \mathbf{i} + \psi_1 \frac{\partial \psi_2}{\partial y} \mathbf{j} + \psi_1 \frac{\partial \psi_2}{\partial z} \mathbf{k} \right) \\ &= \frac{\partial}{\partial x} \left( \psi_1 \frac{\partial \psi_2}{\partial x} \right) + \frac{\partial}{\partial y} \left( \psi_1 \frac{\partial \psi_2}{\partial y} \right) + \frac{\partial}{\partial z} \left( \psi_1 \frac{\partial \psi_2}{\partial z} \right) \\ &= \psi_1 \left( \frac{\partial^2 \psi_2}{\partial x^2} + \frac{\partial^2 \psi_2}{\partial y^2} + \frac{\partial^2 \psi_2}{\partial z^2} \right) + \frac{\partial \psi_1}{\partial x} \frac{\partial \psi_2}{\partial x} + \frac{\partial \psi_1}{\partial y} \frac{\partial \psi_2}{\partial y} + \frac{\partial \psi_1}{\partial z} \frac{\partial \psi_2}{\partial z} \\ &= \psi_1 (\nabla \cdot \nabla \psi_2) + \nabla \psi_1 \cdot \nabla \psi_2 \\ &= \psi_1 \nabla^2 \psi_2 + \nabla \psi_1 \cdot \nabla \psi_2 \end{aligned} \quad \dots(3)$$

Putting (2) and (3) in (1), we have

$$\iint_V (\psi_1 \nabla^2 \psi_2 + \nabla \psi_1 \cdot \nabla \psi_2) dV = \int_S (\psi_1 \nabla \psi_2) \cdot d\mathbf{S} \quad \dots(4)$$

Eq. (4) is called *first form of Green's theorem*.

Interchanging  $\psi_1$  and  $\psi_2$  in Eq. (4), we get

$$\iint_V (\psi_2 \nabla^2 \psi_1 + \nabla \psi_2 \cdot \nabla \psi_1) dV = \int_S (\psi_2 \nabla \psi_1) \cdot d\mathbf{S} \quad \dots(5)$$

Subtracting Eq. (5) from (4), we get

$$\iint_V (\psi_1 \nabla^2 \psi_2 - \psi_2 \nabla^2 \psi_1) dV = \int_S (\psi_1 \nabla \psi_2 - \psi_2 \nabla \psi_1) \cdot d\mathbf{S} \quad \dots(6)$$

Eq. (6) is called the *second form of Green's theorem* or simply the *Green's theorem*.

### MATRICES

#### 44.14 SPECIAL TYPES OF MATRICES

Here we define certain important matrices.

**1. Transpose of a Matrix.** A matrix of order  $n \times m$  obtained by interchanging the rows and columns of a  $(m \times n)$  matrix  $A$  is called the *transpose* of  $A$  and is denoted by the symbol  $A'$  or  $\tilde{A}$  or  $A^T$  ( $A$  transpose) i.e.,

$$\text{If } A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \text{ then } \tilde{A} = A^T = A' = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}$$

#### Properties

- (i)  $(A^T)^T = A$ .
- (ii)  $(\lambda A)^T = \lambda A^T$ ,  $\lambda$  being any scalar (real or complex).
- (iii)  $(A + B)^T = A^T + B^T$ ,  $A$  and  $B$  being conformable for addition.
- (iv)  $(AB)^T = B^T A^T$ ,  $A$  and  $B$  being conformable for multiplication.

**2. Complex-Conjugate Matrix.** The *complex conjugate* of an arbitrary matrix  $A$  is formed by taking the complex conjugate of each element. Hence we have

$$A^* = a_{ij}^* \text{ (for all } i \text{ and } j)$$

If  $A^* = A$ , then  $A$  is a real matrix.

#### Properties

- (i)  $(A^*)^* = A$ .
- (ii)  $(A + B)^* = A^* + B^*$ ,  $A$  and  $B$  being conformable.
- (iii)  $(\lambda A)^* = \lambda^* A^*$ ,  $\lambda$  being any complex number.
- (iv)  $(AB)^* = A^* B^*$ ,  $A$  and  $B$  being conformable for multiplication.

**3. Hermitian Conjugate (The conjugate transpose of a matrix).** The *Hermitian conjugate* of an arbitrary matrix  $A$  is obtained by taking the complex conjugate of the matrix and then the transpose of the complex conjugate matrix.

$$A^\dagger = (A^*)^T = (A^T)^*$$

#### Properties

- (i)  $(A^\dagger)^\dagger = A$ .
- (ii)  $(A + B)^\dagger = A^\dagger + B^\dagger$ ,  $A$  and  $B$  being conformable for addition.
- (iii)  $(\lambda A)^\dagger = \lambda^* A^\dagger$ ,  $\lambda$  being any complex number.
- (iv)  $(AB)^\dagger = B^\dagger A^\dagger$ ,  $A$  and  $B$  being conformable for multiplication.

#### 4. Symmetric and Antisymmetric Matrices

If  $A^T = A$ ,  $A$  is said to be a *symmetric matrix*.

**Example.** 
$$\begin{bmatrix} a & b & c \\ b & c & d \\ c & d & e \end{bmatrix}$$

If  $A^T = -A$ ,  $A$  is said to be an *antisymmetric (skew) matrix*.

**Example.** 
$$\begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}$$

#### 5. Hermitian matrix

If  $A^\dagger = A$ ,  $A$  is said to be a *Hermitian matrix*.

**Example.** 
$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

In quantum mechanics, all physical observables are represented by Hermitian operators (matrices).

#### 6. Skew-Hermitian Matrix

If  $A^\dagger = -A$ ,  $A$  is said to be a *skew-Hermitian matrix*.

**Example.** 
$$\begin{bmatrix} 0 & 1+i \\ -1+i & 0 \end{bmatrix}$$

### 7. Unitary Matrix

A square finite matrix  $A$  is said to be unitary if

$$A^\dagger A = I$$

This implies

$$AA^\dagger = I$$

where  $A^\dagger$  is conjugate transpose of  $A$  and  $I$  is a unit matrix.

We know that  $|A^\dagger| = |A^*|$  and  $|A^\dagger A| = |A^\dagger| |A|$

Hence if  $A^\dagger A = I$ , we have  $|A^\dagger A| = |A^\dagger| |A| = |I| = 1$ .

This shows that the determinant of unitary matrix is always of unit modulus and hence a unitary matrix is non-singular.

### 8. Orthogonal Matrix

A square finite matrix  $A$  is said to be orthogonal if

$$A^T A = I$$

This implies

$$AA^T = I$$

where  $A^T$  is the transpose of  $A$  and  $I$  is unit matrix.

We know that  $|A^T| = |A|$  and  $|A^T A| = |A^T| |A|$

Hence if  $A^T A = I$ , we have  $|A|^2 = 1$  i.e.,  $|A| = \pm 1$

This shows that the determinant of an orthogonal matrix can only have values +1 or -1.

**Example.**  $\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$

## 44.15 EIGEN VALUES, EIGEN-VECTORS : CHARACTERISTIC EQUATION OF A MATRIX

Let  $\mathbf{A} = [\mathbf{a}_{ij}]$  be a square matrix of order  $n$ . Suppose there is an  $n$ -dimensional nonzero column vector  $\mathbf{X}$ , such that the action of  $\mathbf{A}$  on  $\mathbf{X}$  (i.e., the matrix product  $\mathbf{AX}$ ), gives a vector which is just a multiple of  $\mathbf{X}$ , that is

$$\mathbf{AX} = \lambda \mathbf{X}$$

where  $\lambda$  is a scalar. In other words, the transformation represented by the matrix  $\mathbf{A}$  just multiplies the vector  $\mathbf{X}$  by a scalar  $\lambda$ . The vector  $\mathbf{X}$  is then called an *eigen vector* of the matrix  $\mathbf{A}$ .  $\lambda$  is called an *Eigen value* of  $\mathbf{A}$  corresponding to the eigenvector  $\mathbf{X}$ . The problem of finding the eigenvectors and the eigenvalues of a matrix is called the *eigenvalue problem*.

*Definition of Eigenvector.* A nonzero vector  $\mathbf{X}$  is called an eigenvector of a matrix  $\mathbf{A}$  if there is a number  $\lambda$  such that  $\mathbf{AX} = \lambda \mathbf{X}$ .

Here  $\lambda$  is called an eigen value of  $\mathbf{A}$  corresponding to the eigenvector  $\mathbf{X}$  and vice versa.

We have,  $\mathbf{AX} = \lambda \mathbf{X} = \lambda \mathbf{I} \mathbf{X}$ ,  $\mathbf{I}$  being a unit matrix.

or  $(\mathbf{A} - \lambda \mathbf{I}) \mathbf{X} = 0$ .

This is the matrix form of an eigenvalue problem.

Since  $\mathbf{X} \neq 0$ , the matrix  $(\mathbf{A} - \lambda \mathbf{I})$  is singular, so that

$$|\mathbf{A} - \lambda \mathbf{I}| = 0 \quad \dots(1)$$

Equation (1) is called the *characteristic equation* of  $\mathbf{A}$ . The eigenvalues are just the roots of the equation obtained by expanding the determinant in Eq. (1). Then  $n$ -roots  $\lambda_1, \lambda_2, \dots, \lambda_n$  of the characteristics equation are not necessarily all different.

**EXAMPLE 1.** Determine the eigen values and eigen vectors of the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 4 \\ 0 & 2 & 6 \\ 0 & 0 & 5 \end{bmatrix}$$

**SOL.** The characteristic equation of  $\mathbf{A}$  is  $|\mathbf{A} - \lambda\mathbf{I}| = 0$  i.e.,

$$\begin{vmatrix} 3-\lambda & 1 & 4 \\ 0 & 2-\lambda & 6 \\ 0 & 0 & 5-\lambda \end{vmatrix} = 0 \text{ i.e., } (\lambda-2)(\lambda-3)(\lambda-5) = 0$$

$$\therefore \lambda_1 = 2, \lambda_2 = 3, \lambda_3 = 5$$

These are the eigen values of  $\mathbf{A}$ .

To determine eigen vectors let us consider the eigen values one by one.

(i) When  $\lambda_1 = 2$  the eigen vector  $\mathbf{X}_1$  is given by  $(\mathbf{A} - 2\mathbf{I}) \mathbf{X}_1 = 0$

$$\text{i.e., } \begin{bmatrix} 1 & 1 & 4 \\ 0 & 0 & 6 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The rank of coefficient matrix being 2, the equation will have only  $3 - 2$  i.e., 1 linearly independent solution.

These are equivalent to  $x_1 + x_2 + 4x_3 = 0$

$$6x_3 = 0$$

$$3x_3 = 0$$

The last two give  $x_3 = 0$ . Then first one gives  $x_1 + x_2 = 0$ .

Take  $x_1 = 1$ , then  $x_2 = -1$  and  $x_3 = 0$ .

$$\text{Hence } \mathbf{X}_1 = C_1 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad C_1 \text{ being a scalar.}$$

(ii) When  $\lambda_2 = 3$ , the eigen vector  $\mathbf{X}_2$  is given by  $(\mathbf{A} - 3\mathbf{I}) \mathbf{X}_2 = 0$

$$\text{i.e., } \begin{bmatrix} 0 & 1 & 4 \\ 0 & -1 & 6 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

These are equivalent to  $x_2 + 4x_3 = 0$

$$-x_2 + 6x_3 = 0$$

$$2x_3 = 0$$

giving  $x_3 = 0, x_2 = 0$  and  $x_1$  is arbitrary, say  $x_1 = 1$ .

$$\text{Then } \mathbf{X}_2 = C_2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad C_2 \text{ being a scalar.}$$

(iii) When  $\lambda_3 = 5$ , the eigen vector  $\mathbf{X}_3$  is given by  $(\mathbf{A} - 5\mathbf{I}) \mathbf{X}_3 = 0$

$$\text{i.e., } \begin{bmatrix} -2 & 1 & 4 \\ 0 & -3 & 6 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

These are equivalent to  $-2x_1 + x_2 + 4x_3 = 0$

$$-3x_2 + 6x_3 = 0$$

giving

$$x_2 = 2x_3 = \frac{2}{3}x_1, \quad i.e., \quad 2x_1 = 3x_2 = 6x_3$$

Take  $x_3 = 1$ , so that  $x_2 = 2$  and  $x_1 = 3$

Hence

$$\mathbf{X}_3 = C_3 \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

$C_3$  being a scalar.

#### 44.16 CAYLEY-HAMILTON THEOREM

**Statement.** Every square matrix satisfies its own characteristic equation

OR

If  $|\mathbf{A} - \lambda \mathbf{I}| = a_0 + a_1 \lambda + a_2 \lambda^2 + \dots + a_n \lambda^n = 0$

be the characteristic equation of a square matrix  $\mathbf{A}$ , then

$$a_0 \mathbf{I} + a_1 \mathbf{A} + a_2 \mathbf{A}^2 + \dots + a_n \mathbf{A}^n = 0$$

**Proof.** The characteristic polynomial is

$$|\mathbf{A} - \lambda \mathbf{I}| = a_0 + a_1 \lambda + a_2 \lambda^2 + \dots + a_n \lambda^n \quad \dots(1)$$

Each element of characteristic matrix  $(\mathbf{A} - \lambda \mathbf{I})$  is an ordinary polynomial of degree  $n$  (at most). Therefore the co-factor of every element of  $|\mathbf{A} - \lambda \mathbf{I}|$  is an ordinary polynomial of degree  $n - 1$  (at most). Consequently each element of

$$\mathbf{B} = \text{adj}(\mathbf{A} - \lambda \mathbf{I}) \quad \dots(2)$$

is an ordinary polynomial of degree  $(n - 1)$  (at most).

$$\therefore \mathbf{B} = \text{adj}(\mathbf{A} - \lambda \mathbf{I}) = \mathbf{B}_0 + \mathbf{B}_1 \lambda + \mathbf{B}_2 \lambda^2 + \dots + \mathbf{B}_{n-1} \lambda^{n-1} \quad \dots(3)$$

Here  $\mathbf{B}_0, \mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_{n-1}$  are all square matrices of the same order  $n$  whose elements are polynomials in the elements of  $\mathbf{A}$ .

$$\text{Now, } (\mathbf{A} - \lambda \mathbf{I}) \text{ adj}(\mathbf{A} - \lambda \mathbf{I}) = |\mathbf{A} - \lambda \mathbf{I}| \mathbf{I}$$

Using Eqs. (3) and (1), we get

$$\begin{aligned} & (\mathbf{A} - \lambda \mathbf{I}) [\mathbf{B}_0 + \mathbf{B}_1 \lambda + \mathbf{B}_2 \lambda^2 + \dots + \mathbf{B}_{n-1} \lambda^{n-1}] \\ &= (a_0 + a_1 \lambda + a_2 \lambda^2 + \dots + a_n \lambda^n) \mathbf{I} \end{aligned} \quad \dots(4)$$

Comparing the coefficients of like powers of  $\lambda$  on both the sides, we get

$$\mathbf{A}\mathbf{B}_0 = a_0 \mathbf{I}$$

$$\mathbf{A}\mathbf{B}_1 - \mathbf{B}_0 = a_1 \mathbf{I}$$

$$\mathbf{A}\mathbf{B}_2 - \mathbf{B}_1 = a_2 \mathbf{I}$$

.....

.....

$$\mathbf{A}\mathbf{B}_{n-1} - \mathbf{B}_{n-2} = a_{n-1} \mathbf{I}$$

$$-\mathbf{B}_{n-1} = a_n \mathbf{I}$$

Premultiplying these by  $\mathbf{I}, \mathbf{A}, \mathbf{A}^2, \mathbf{A}^3, \dots, \mathbf{A}^n$  in order and adding,

$$a_0 \mathbf{I} + a_1 \mathbf{A} + a_2 \mathbf{A}^2 + \dots + a_n \mathbf{A}^n = 0$$

This is **Cayley-Hamilton theorem**.

**Corollary.** To determine  $\mathbf{A}^{-1}$  by using Cayley-Hamilton theorem.

Let  $\mathbf{A}$  be a non-singular matrix of order  $n$  so that  $|\mathbf{A}| \neq 0$ .

According to Cayley-Hamilton theorem

$$a_0 \mathbf{I} + a_1 \mathbf{A} + a_2 \mathbf{A}^2 + \dots + a_n \mathbf{A}^n = 0 \quad \dots(1)$$

The characteristic polynomial is

$$|\mathbf{A} - \lambda \mathbf{I}| = a_0 + a_1 \lambda + a_2 \lambda^2 + \dots + a_n \lambda^n \quad \dots(2)$$

For  $\lambda = 0$ , Eq. (2) gives  $|\mathbf{A}| = a_0$ .  $\therefore a_0 \neq 0$

Now dividing Eq. (1) by  $a_0$ , we get

$$\mathbf{I} = -\left(\frac{a_1}{a_0} \mathbf{A} + \frac{a_2}{a_0} \mathbf{A}^2 + \dots + \frac{a_n}{a_0} \mathbf{A}^n\right) \quad \dots(3)$$

Pre-multiplying Eq. (3) by  $\mathbf{A}^{-1}$ , we get

$$\mathbf{A}^{-1} = -\left(\frac{a_1}{a_0} \mathbf{I} + \frac{a_2}{a_0} \mathbf{A} + \dots + \frac{a_n}{a_0} \mathbf{A}^{n-1}\right) \quad \dots(4)$$

**EXAMPLE 1.** Find the characteristic equation of the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & -1 & 4 \\ 3 & 1 & 1 \end{pmatrix}$$

and verify the Cayley-Hamilton theorem for it. Hence or otherwise find  $\mathbf{A}^{-1}$ .

$$\begin{aligned} \text{SOL.} \quad \mathbf{A} - \lambda \mathbf{I} &= \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & 4 \\ 3 & 1 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1-\lambda & 2 & 3 \\ 2 & -1-\lambda & 4 \\ 3 & 1 & 1-\lambda \end{bmatrix} \\ \therefore |\mathbf{A} - \lambda \mathbf{I}| &= \begin{vmatrix} 1-\lambda & 2 & 3 \\ 2 & -1-\lambda & 4 \\ 3 & 1 & 1-\lambda \end{vmatrix} = -\lambda^3 + \lambda^2 + 18\lambda + 30 \end{aligned}$$

Hence the characteristic equation is

$$-\lambda^3 + \lambda^2 + 18\lambda + 30 = 0$$

Now, in order to verify Cayley-Hamilton theorem, we have to show that

$$-\mathbf{A}^3 + \mathbf{A}^2 + 18\mathbf{A} + 30\mathbf{I} = 0$$

$$\text{Here } \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & 4 \\ 3 & 1 & 1 \end{bmatrix}$$

$$\mathbf{A}^2 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & 4 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & 4 \\ 3 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 3 & 14 \\ 12 & 9 & 6 \\ 8 & 6 & 14 \end{bmatrix}$$

$$\mathbf{A}^3 = \mathbf{A}^2 \mathbf{A} = \begin{bmatrix} 14 & 3 & 14 \\ 12 & 9 & 6 \\ 8 & 6 & 14 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & 4 \\ 3 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 62 & 39 & 68 \\ 48 & 21 & 78 \\ 62 & 24 & 62 \end{bmatrix}$$

$$\therefore -\mathbf{A}^3 + \mathbf{A}^2 + 18\mathbf{A} + 30\mathbf{I}$$

$$= -\begin{bmatrix} 62 & 39 & 68 \\ 48 & 21 & 78 \\ 62 & 24 & 62 \end{bmatrix} + \begin{bmatrix} 14 & 3 & 14 \\ 12 & 9 & 6 \\ 8 & 6 & 14 \end{bmatrix} + 18 \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & 4 \\ 3 & 1 & 1 \end{bmatrix} + 30 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0$$

Hence Cayley-Hamilton theorem is verified.

To find  $\mathbf{A}^{-1}$ :  $30\mathbf{I} - \mathbf{A}^3 + \mathbf{A}^2 + 18\mathbf{A} = 0$

$$\mathbf{I} = \frac{1}{30} (\mathbf{A}^3 - \mathbf{A}^2 - 18\mathbf{A})$$

Premultiplying the above equation by  $\mathbf{A}^{-1}$ , we get

$$\begin{aligned} \mathbf{A}^{-1} &= \frac{1}{30} (\mathbf{A}^2 - \mathbf{A} - 18\mathbf{I}) \\ \mathbf{A}^{-1} &= \frac{1}{30} \begin{bmatrix} 14 & 3 & 14 \\ 12 & 9 & 6 \\ 8 & 6 & 14 \end{bmatrix} - \frac{1}{30} \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & 4 \\ 3 & 1 & 1 \end{bmatrix} - \frac{18}{30} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ \mathbf{A}^{-1} &= \begin{bmatrix} -\frac{5}{30} & \frac{1}{30} & \frac{11}{30} \\ \frac{10}{30} & -\frac{8}{30} & \frac{2}{30} \\ \frac{5}{30} & \frac{5}{30} & -\frac{5}{30} \end{bmatrix} \end{aligned}$$

## EXERCISE

1. Prove that  $\Gamma(n+1) = n\Gamma(n)$ . (B.U. 2014)
2. Find the value of  $\Gamma\left(\frac{1}{2}\right)$ . (B.U. 2014)
3. Prove that  $\beta(m, n) = 2 \int_0^{\frac{\pi}{2}} \sin^{2m-1} \theta \cdot \cos^{2n-1} \theta \cdot d\theta$  (B.U. 2014)
4. State and prove Stoke's theorem. (B.U. 2014)
5. State and prove Cayley Hamilton theorem. (B.U. 2014)